

JACET INTERNATIONAL CONVENTION  
SELECTED PAPERS  
VOL.4



ISSN 2188-8612



**JACET International Convention**  
**Selected Papers**  
**Volume 4**



**JACET 55<sup>th</sup> (2016) International Convention**

Theme: Designing English Education in a Borderless Era

September 1 – 3, 2016

Hokusei Gakuen University

**JACET International Convention Selected Papers, Volume 4**  
Published by the Japan Association of College English Teachers (JACET)

---

**JACET Selected Papers Committee**

**Director-in-Charge**

KAWANO, Madoka (Meiji University)

**Editorial Board**

**Chair**

KAWAI, Yasushi (Hokkaido University)

**Deputy Chair**

KANAOKA, Masao (Kagoshima University)

OKUGIRI, Megumi (University of the Sacred Heart)

DALSKY, David (Kyoto University)

MATSUMOTO, Hiroyuki (Hokkai-Gakuen University)

HAENOUCHE, Hiroko (Nihon University)

PENG, Virginia (Ritsumeikan University)

HIRAMOTO, Satoshi (Yasuda Women's University)

TAKAHASHI, Sachi (Kyoto University)

IIDA, Atsushi (Gunma University)

TOMITA, Kaoru (Yamagata University)

ITO, Mika (Tokai University)

TOYA, Mitsuyo (Ryukyu University)

KANAMARU, Toshiyuki (Kyoto University)

WISTNER, Brian (Hosei University)

KANEKO, Emiko (University of Aizu)

YOKOYAMA, Yoshiki (Hokkaido University of

KURAHASHI, Yoko (Tokai Gakuen University)

Education)

**Advisory Board**

CHON, Kaye (The Hong Kong Polytechnic University)

SPADA, Nina (University of Toronto)

TONO, Yukio (Tokyo University of Foreign Studies)

---

**JACET Prize & Academic Publication Selection Committee**

**Director-in-Charge**

OZEKI, Naoko (Meiji University)

**Head**

IWAI, Chiaki (Hiroshima City University)

**Vice-Head**

TAJINO, Akira (Kyoto University)

## Reviewers

ANTHONY, Laurence (Waseda University)	ONO, Naomi (Seikei University)
ASAOA, Chitose (Dokkyo University)	OSUKA, Naoko (Meiji University)
FUJIEDA, Miho (Osaka Medical College)	SAKAI, Hideki (Shinshu University)
HAYASHI, Hideo (Kumamoto Gakuen University)	SASAKI, Masako (Akita University)
IKENO, Osamu (Ehime University)	SASAO, Yosuke (Kyoto University)
IRIE, Kay (Gakushuin University)	SUZUKI, Chizuko (Nagasaki Junshin Catholic University)
ISO, Tatsuo (Tokyo Denki University)	TAGUCHI, Etsuo (Daito Bunka University)
IZUMI, Emiko (Kyoto University of Education)	TAKAGI, Akiko (Aoyama Gakuin University)
KAWAUCHI, Chieko (Kurume University)	TERAUCHI, Masanori (Hosei University)
KOMIYA, Tomiko (Okazaki Women's University)	TORIKAI, Shinichiro (Rikkyo University)
KONNO, Katsuyuki (Shizuoka Institute of Science and Technology)	TSUDA, Sanae (Professor Emerita, Tokaigakuen University)
MIYAHARA, Masuko (International Christian University)	YAMANE, Shigeru (Kansai University)
MURANOI, Hitoshi (Tohoku Gakuin University)	YAMATO, Ryusuke (Kyoto Sangyo University)
NAITO, Hisashi (Hokkai-Gakuen University)	YOSHIKAWA, Hiroshi (Chukyo University)
NITTA, Ryo (Nagoya Gakuin University)	
OISHI, Harumi (Gifu Shotoku Gakuen University)	

---

## Cover Design

KASUNO, Shin-ichi

## Editorial Office

The JACET Office

Address: 55 Yokoteramachi, Shinjuku-ku, Tokyo 162-0831 JAPAN

Phone: +81-3-3268-9686 / Fax: +81-3-3268-9695

E-mail: [jacet@zb3.so-net.ne.jp](mailto:jacet@zb3.so-net.ne.jp)

---

First published in August, 2017

ISSN 2188-8612

Copyright © 2017 by JACET

All rights reserved. No part of this publication may be reproduced or republished in any form without permission in writing from JACET.

The articles published herein do not reflect the opinions of JACET.



## Message from the Editorial Board Chair

We proudly present the fourth issue of Selected Papers from the Japan Association of College English Teachers [JACET] 55th International Convention held at Hokusei Gakuen University, Sapporo, Japan, from September 1 to 3, 2016. Under the conference theme: *Designing English Education in a Borderless Era*, this convention held numerous productive sessions with 880 participants from all over the world, including keynote speeches, invited lectures, research paper presentations, educational practice reports, symposiums, and workshops. This volume comprises six articles that represent a collection of selected papers from the convention.

The first two papers are contributions from the plenary speakers, Dr. Nina Spada, Professor in the language and literacies education program at the University of Toronto, and Dr. Yukio Tono, Professor of corpus linguistics at Tokyo University of Foreign Studies. They each provided us with profound insight into English education in Japan from the perspective of language and content integration as well as a reference framework for teaching, learning and assessing in English education in Japan.

The next four papers were written by conference presenters. After a highly rigorous and stringent review process, four papers out of eleven were selected from JACET members who responded to our call for papers. Throughout the six-month-long selection process, our greatest concern was to uphold academic and educational quality in research and in writing. Indeed, this volume of selected papers represents the most interesting and inspiring works from the convention. They will certainly capture the attention of all those who have access online.

Finally, as the Chair of the Editorial Board, I appreciate each of the reviewers, the Editorial Board members, and all those who assisted us in publishing the fourth volume of the *JACET International Convention Selected Papers*. I sincerely hope that this publication will illuminate the path to the future of English education in this borderless era.

Yasushi Kawai, Ph.D.

Chair, The Editorial Board of the JACET International Convention Selected Papers, Volume 4





# Contents

Message from the Editorial Board Chair .....	v
--	---

## **Invited Papers**

Focusing on Language in Meaning-based and Content-based Instruction .....	Nina Spada	3
The CEFR-J and its Impact on English Language Teaching in Japan .....	Yukio Tono	31

## **Selected Papers**

### Research Articles

Linguistic Features of Discussion Sections of English Medical Research Paper Genres and Their Pedagogical Implications .....	Motoko Asano, Judy Noguchi	55
Assessing the Effectiveness of the COLT Scheme as a Reflection Tool for High School Teachers of English ...	Aiko Sano, Noriaki Katagiri, Yuko Sakai, Akinobu Shimura	85
Exploring Japanese College Student Perceptions of Native and Nonnative English-Speaking Teachers: The Case of Repeaters .....	Takaharu Saito	115
Students' Uptake and Task Activities as Measures of Effective Learning .....	Yukiko Ohashi	133
<b>Submission Guidelines .....</b>		<b>165</b>



# **Invited Papers**



## **Focusing on Language in Meaning-based and Content-based Instruction<sup>1</sup>**

Nina Spada

University of Toronto

### **Abstract**

In this paper I discuss the importance and challenges that come with focusing on language in meaning-based and content-based instruction. I use the term meaning-based instruction (MBI) to refer to approaches to second/foreign language teaching that focus on communication and functional/purposeful language use through topics and themes that are appropriate to the age, interests, and goals of learners. Examples of MBI include communicative language teaching (CLT) and task-based language teaching (TBLT). I use the term content-based instruction (CBI) to refer to the simultaneous teaching and learning of a second/foreign language and academic content. Examples of CBI include content and language integrated learning (CLIL) and Canadian French immersion programs. Both MBI and CBI are premised on the assumption that language is learned through its use in communicative interaction and its integration within meaningful and engaging content. The main difference between the two is that in CBI learners are expected to learn academic content at the same time as they learn the language of that content. The focus of this paper is to review some of the research that has investigated the effects of language-focused

---

<sup>1</sup> This paper is a written version of a plenary presentation that was delivered at the 2016 JACET Convention hosted by Hokusei Gakuen University in Sapporo, Hokkaido. It has been slightly revised and expanded.

instruction in MBI and CBI with adult and university-level learners of a second or foreign language.

**Keywords:** content and meaning-based instruction, integrated and isolated form-focused instruction, explicit and implicit teaching and learning, functional grammar

## **Introduction**

In this paper I discuss the importance and challenges that come with focusing on language in meaning and content-based instruction. I begin with a discussion of the similarities and differences between these two types of second/foreign language instruction (L2)<sup>2</sup>. This is followed by a description of ways in which they are realized. This includes reference to strong and weak versions of communicative language teaching (CLT), implicit and explicit language-focused instruction and different degrees of content and language focus in CBI programs. I devote most of my attention to a review of research on the effectiveness of focusing on language in meaning and content-based instruction targeted to adult and university-level learners. This research examines questions related to planned versus incidental language-focused instruction and the integration and isolation of attention to language within meaning/content-based teaching. While most of the research reviewed is rooted within conceptions of L2 learning and teaching that are cognitively based I also briefly discuss research that is situated within socially oriented theoretical work specifically that which is influenced by systemic functional linguistics (Halliday, 1978, 1994).

The distinction that I am making between meaning-based instruction (MBI) and content-based instruction (CBI) is as follows. In MBI, the primary focus is on meaning

---

<sup>2</sup> I am using L2 to refer to both second and foreign language learning unless otherwise indicated.

through different topics and themes that are appropriate to the age, interests, and goals of the learners. The focus in MBI is to teach language using approaches that emphasize communication and purposeful language use such as communicative language teaching (CLT) and task-based language teaching (TBLT). In CBI the primary focus is also on meaning but in this case it is through the teaching of subject matter and academic, for example, in Canadian French immersion programs.

Both MBI and CBI are premised on the assumption that language is learned through its use in meaningful communication. The main difference between the two is that with CBI learners are expected to get “two for one” – language and academic content simultaneously (Lightbown & Spada, 2013). In the marketplace this expression usually means that you “pay for one and get one free”. However, the assumption that learners who receive content-based instruction will “pick up language along the way” is problematic. While it is true that some aspects of language are learned incidentally, a good deal of language is not, particularly the kind of language needed to succeed in academic contexts. Indeed the challenge for learners in CBI classrooms to master both language and content is considerable. In what follows I discuss some of the theoretical, empirical and pedagogical issues relevant to the learning and teaching of L2s via meaning and content-based instruction. Given that the JACET membership consists mainly of teachers and researchers of university-level EFL learners, I will focus my attention on research that has been done with university and college level learners but I will also include some discussion of work with other target groups where relevant.

First, a few words about terminology. A label that is frequently used in the literature to describe programs in which the teaching and learning of both language and subject-matter content are the goals is *Content-based Language Teaching* (CBLT). This label is thought to have been coined in the 1960s with reference to French immersion programs in Canada and bilingual education programs in the United States. Another label – *Content and Language*

*Integrated Learning* (CLIL) – arrived on the scene in the 1990s within the context of widespread educational reforms in Europe encouraging schooling in more than one language. In most cases this includes the teaching of English in addition to for example Spanish, Polish, Italian or another European language. Subsequent to the introduction of CLIL in Europe, it has spread throughout the world primarily (e.g. Asia, South America) within the context of English foreign language instruction (Dalton-Puffer, 2007; Coyle, Hood & Marsh, 2010).

There are debates as to whether CLIL and CBLT represent similar or different approaches (Lasagabaster & Sierra, 2010; Cenoz, Genesee, & Gorter, 2014). Some have argued that CLIL espouses a stronger integration of language and content than has been observed in CBLT programs. While this might be true in theory and perhaps in pedagogical prescriptions, it has not been demonstrated empirically. I know of no research that has systematically compared how language and content are integrated more (or less) in CLIL versus CBLT. While other differences between CLIL and CBLT have been discussed in the literature (e.g. foreign versus second language focus; older versus younger learners; differences in proficiency objectives), in my view, CBLT and CLIL are much more similar than they are different. Nonetheless, in this paper I have decided to use the term CBI as an umbrella term for both.

It is important to note that CBI programs are not monolithic. That is, there are differences in the degree to which language and content are the focus of attention. This variation can be captured as a continuum that reflects ‘strong’ and ‘weak’ versions with regard to the role of content and language. As indicated in Table 1, at one end of the continuum are programs that are distinctly content-driven. That is, content determines the course goals, content learning outcomes are assessed and the teacher is a content expert (e.g. French immersion). At the other end of the continuum are programs that use content to teach the language. That is, language determines the content and the course goals and language outcomes are assessed (e.g. English for specific purposes (ESP) programs).



Table 1

*Degrees of focus on content and language in CBI*

← Content driven	Language driven →
Content is taught in L2	Language determines content
Content determines course goals	Language determines course goals
Content learning outcomes are assessed	Language outcomes are assessed
Teacher is a content expert	Teacher is a language expert

Note. This table is slightly revised from Valeo (2010) and is originally adapted from Zyzik and Polio (2008)

There are also differences in the degree to which a focus is placed on meaning and language in programs broadly described as meaning-based. For example, two types of communicative language teaching are described by Howatt (1984) – one is referred to as the ‘strong version’ of CLT in which there is an exclusive focus on meaning without grammar instruction or corrective feedback; the other, the ‘weak version’ includes attention to both language and meaning. The strong version of CLT is based on the hypothesis that if educators can create conditions in L2/FL classrooms that are similar to those of first language development (e.g. comprehensible input, spontaneous meaningful interaction) L2 learners will develop a grammar in much the same way that child L1 learners do (Krashen, 1982; Prabhu, 1984). In the weak version of CLT or what I prefer to call the balanced version of CLT, the focus is on meaningful input, communicative interaction and fluency but not to the exclusion of the development of accuracy through grammar instruction and corrective feedback. This version of CLT is similar to another construct in the instructed second language acquisition (SLA) literature - form-focused instruction (FFI). This is defined as “any pedagogical effort

which is used to draw the learners' attention to language form either implicitly or explicitly [within meaning-based practice]. This can include the direct teaching of language (e.g. through grammatical rules) and/or reactions to learners' errors (e.g. corrective feedback)” (Spada, 1997 p.73)

### **Explicit and implicit FFI**

Direct teaching of language or explicit FFI can include metalinguistic information using grammar rules as indicated in example #1 below providing information about adverb placement in English sentences. But explicit FFI is not restricted to grammar rules and does not need to contain metalanguage. It can also include visual and/or graphic information about how the target language works as seen in example #2 with the use of arrows and other graphics.

#### **Example 1**

##### *Explicit Metalinguistic<sup>3</sup>*

In English, adverbs can be placed:

- 1) before the subject: Loudly Shroeder plays the piano
- 2) between the subject & verb: Shroeder loudly plays the piano
- 3) after the object – Shroeder plays the piano loudly

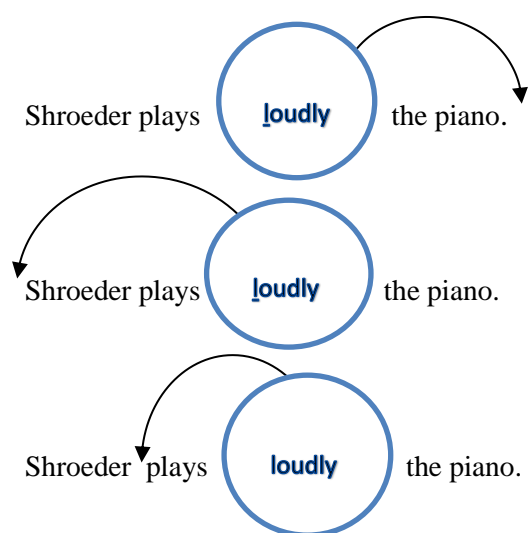
But NOT between the verb and object

“Shroeder plays ~~loudly~~ the piano”

---

<sup>3</sup> These examples come from the instructional materials used in a study reported in Trahey (1996) and Trahey & White (1993).

## Example 2

*Explicit Non-metalinguistic*

## Example 3

*Implicit FFI*

*I usually wear a white coat when I work. I often work late at night and I always work very hard to save lives. When people come to see me, I always ask them questions about their health. Usually I work in a hospital but sometimes, I work at my office. I frequently work on the weekend. Sick people usually call me. Who am I?*

Implicit FFI on the other hand is contextualized and embedded within a meaningful context. This could include exposure to exemplars through high frequency input as seen in example #3 where adverb placement is still being taught but implicitly. In this example there are several instances of adverbs in different positions in English sentences. Learners are asked to read the short passage (and others like it) and then look at photos of individuals representing different professions. Based on their understanding of the text they can choose

between 1) a fireman 2) a police woman 3) a medical doctor. The assumption is that through this exposure to many examples of adverbs and illustrations of how they are placed in different positions, learners will figure out where they can (and cannot) place adverbs in English sentences.

### **Research in MBI and CBI: Attention to language and content/meaning**

Over the past 30 years considerable research has been done to investigate the effects of different approaches to drawing learners' attention to language in meaning and content-based classrooms in relation to L2 development over time. The overall findings indicate that instruction which is primarily meaning-based and includes attention to form is more effective than instruction which focuses exclusively on meaning or exclusively on form (Lightbown & Spada, 2013). Furthermore, there is increasing evidence that explicit FFI is more beneficial than implicit FFI (Goo et al., 2015; Norris & Ortega, 2000) for a range of different language features (Spada & Tomita, 2010). While this research includes studies that have been conducted with children, adolescents and adults in MBI and CBI programs the bulk of it has been carried out with adult learners in MBI programs.

Research investigating the effects of instruction in relation to language gains over time in CBI programs has focused primarily on school-aged learners. This includes research that has examined the effects of FFI and corrective feedback on L2 learning in quasi-experimental studies in Canadian French immersion programs (e.g. Day & Shapson, 1991; Harley, 1998; Lyster, 1994, 2007). It also includes studies that have been carried out with immigrant children mainstreamed into English-medium schools in North America (e.g. Doughty & Varela, 1998; de Oliveira & Schleppegrell, 2015). A good deal of this research is descriptive in nature. That is, it examines the nature and extent of language-focused interactions that occur between teachers and learners and the impact this has on learners' performance as

lessons unfold. There appear to be fewer studies that have investigated the impact of language-focused instruction on specific linguistic gains over time in this context.

Research that has explored L2 learning with adults in CBI programs also includes classroom observation studies that have focused on the degree of attention given to language and content in interactions between teachers and learners. Several have documented a lack of sufficient attention to language. For example, in an investigation of teacher-student exchanges in three content-based language classrooms where students were learning Italian through a social geography course, Musumeci (1996) reports that while teachers regularly modified their speech in response to signals of miscomprehension, they rarely requested that students modify their own imprecise and inaccurate speech as long as the message was comprehensible. Similarly in CBI classes where university learners of ESL were taught film, culture and literature via the L2, Pica (2002) documents few opportunities for learners to focus on relationships between L2 form and meaning. She reports that while the subject-matter content was rich and engaging, there were “negligible amounts of interaction involving form-focused intervention or interaction ... [and that] the majority of student non-target utterances were not addressed in any direct way” (Pica, 2002, p. 16). Findings like these are likely related to concerns that explicit attention to language will detract from attention to content in such programs (Klee & Tedick, 1997; Toth, 2004; Zyzik & Polio, 2008).

Below I describe some studies that have been carried out in CBI programs for university and college-level learners to investigate the effects of language-focused instruction on L2 development over time. Some of these studies focus on the development of specific linguistic features and others on the development of different aspects of proficiency (e.g. reading and listening comprehension). This includes research that has compared planned versus incidental FFI and the effects of integrating and isolating language-focused instruction within content-based teaching. I discuss the results primarily with respect to learners’ progress in language development but also in terms of their academic progress in those cases

where the information is available. As indicated above, although I have made efforts to limit my focus to studies with university and college level learners, I also include one or two studies with advanced secondary level learners.

### **Planned versus Incidental FFI**

In the 1980s and 90's research inspired by Krashen's (1981) comprehensible input hypothesis was carried out at a French-English bilingual university in Canada where content-based instruction has been available in several disciplines for over three decades. This research investigated students' progress in language and academic content. The instruction took two forms: 'sheltered' courses in which L2 learners were segregated in a group for modified instruction by a content-area specialist with language support provided by a language teacher and 'adjunct' courses in which L2 learners were integrated into content courses offered for native speakers but 'sheltered' as a group in separate language courses related to the course content (see Brinton, Snow & Wesche, 1990 for descriptions of 'sheltered' and 'adjunct' and other types of CBLT programs). One of the earliest studies (Edwards et al., 1984) compared students taking a sheltered psychology course in their second language (i.e. French or English) with students taking the same course in their first language and found no differences in academic achievement. In addition, the sheltered students did as well on tests of English and French reading and listening proficiency as students in advanced language courses. Similar findings were reported in a subsequent study which took place over a longer period of time (Hauptman et al., 1988). In an investigation of 7 different adjunct courses Ready & Wesche (1992) also reported advanced level gains in listening and speaking and comparable performance on psychology tests as their L1 counterparts. Students in this study also self-reported significant gains in their speaking and writing ability and high levels of confidence using the L2.

It is important to note that in the language-focused classes for the sheltered and adjunct courses, the emphasis was on providing support for students studying academic content via a second language. Thus the instruction focused on aural and reading comprehension, note-taking skills and reading strategies (e.g. guessing from context). There was no planned grammar instruction or syllabus; any focused language teaching emerged from the instructors' responses to specific language problems raised by students. In the psychology classes the focus was exclusively on the content and not language. However, observational research revealed that the instructors used a variety of linguistic adjustments in their speech (e.g. repetitions, paraphrases, comprehension checks) in an effort to make the content comprehensible. These adjustments were used significantly more frequently by instructors teaching psychology to L2 learners than those teaching psychology to native speakers (Wesche & Ready, 1985)

Most of the research to investigate the language development of students in these Canadian 'sheltered' and 'adjunct' CBI programs measured learners' receptive ability (i.e. reading and listening) not their productive skills (i.e. speaking and writing). One exception is a study by Burger and Chrétien (2001) which focused on learners' oral productive ability and found that the English and French L2 learners significantly improved on an oral elicited imitation task and an oral discussion task over 2 semesters. These findings were considered impressive by the researchers because "although students are receiving extensive exposure to the target language through the subject-matter course, their attention is focused more on content than on language. Furthermore, the adjunct language course is allotted only half the time of regular language courses .... a great deal of attention is paid to the students understanding of content and little time is left to focus on language. Syntactic and discourse features are not taught formally... It appears that this massive exposure to reading texts and lectures plus these opportunities for [incidental, unplanned language instruction] are sufficient to ensure gains in oral proficiency" (Burger & Chrétien, 2001, p. 98).

In a study that directly compared the effects of planned and incidental FFI on L2 learning in a content-enriched French language program, Grim (2008) investigated the L2 development of beginner-level university students in a content-enriched program. Three groups of second and third semester students participated in the study. All of them received an enhanced cultural component taught in French. This included teaching materials that focused on two French-speaking countries (i.e. Belgium and Senegal). The content included geography, population, cuisine, music, literature, holidays, traditions and popular culture. The students were divided into three groups: The ‘planned FFI’ group received instruction in which lexical and grammatical forms were written in boldface type and in color and the teacher provided grammatical explanations throughout the lessons. The ‘incidental FFI’ group used the same material and the teacher explained grammar points only in response to students’ questions about grammar on an ad-hoc basis. The ‘focus on meaning’ group used the same material without any enhancement and no grammar explanations. Students received 50 minutes of the instructional intervention and they were pre-tested before the instruction, post-tested immediately after and delayed post-tested 2 weeks later. When the outcomes were compared across treatments and between second and third semester students, the results revealed that there was a significant advantage on both content and language measures for the second semester students in the planned FFI group. However, these gains were not maintained on the delayed post-tests. Significant gains were also made for the second-semester students in the planned FFI group but only on the lexical post-test. Based on these findings Grim concluded that learners benefitted from planned FFI in a content-enriched program and that FFI did not negatively affect content learning. However, the gains were not maintained in the long term which Grim suggests may have been due to the short duration of the instructional intervention.



### Integrated and isolated FFI

In a long-term study to investigate the contributions of a focus on language in a CBI program for college level students studying early childhood education in occupation-based English medium classes, Valeo (2010; 2013) followed the progress of two groups of learners over 27 weeks of a 40-week program. The students were randomly assigned to one of two groups: the form-focused (FF) group in which learners received language instruction integrated with content and the meaning-focused group (MF) in which learners received the same content-based instruction (i.e. topics in early childhood education) but without any focus on language. The FF instruction consisted of metalinguistic explanations, form-focused tasks and explicit corrective feedback. The focus of the instruction was two target features: *simple past tense* and *real conditional*. The forms were selected because of their frequency in the input and the proficiency level of the learners. Learners were tested before the instruction began, immediately after it ended and again 12 weeks later. The language measures included an error correction task, a cloze test, and oral production tasks. Content outcomes were measured via tests focused on specific topics and themes in the childhood education materials. The overall findings indicated similar language gains for learners in both groups on most of the language measures. There were also significantly greater gains on the content measures for students in the FF group. With respect to the language gains, these results would seem to offer support for the claim that content-rich meaningful input was sufficient for learners to develop their knowledge of the linguistic features. These findings appear to be consistent with those from other content-based programs where the focus on language has been characterized as unplanned or incidental (e.g. Burger et Chrétien, 2001 discussed above). However, the researcher's concerns about whether the FF instruction was implemented as intended might explain why learners in the FF group did not improve as much on the language measures as anticipated. There is evidence that the FF instruction was not sufficiently integrated with the content, for example it did not include consistent corrective feedback which is an important

strategy for integrating form and meaning. This might explain why learners in the FF group did not perform significantly better than learners in the MF group. The finding that learners in the FF group performed significantly better on the content measures is not uncommon and is likely due to the fact that the enhanced attention to language enabled learners to better understand the content.

In another long-term study to investigate the effects of a focus on language in content-based instruction, Trahey (2016) followed the ESL development of a group of immigrant senior secondary-school students mainstreamed into English-medium schools in Canada. Over a 15-week period, the students received explicit FFI that was integrated into the content of their social studies course. The instruction covered a range of activities, including metalinguistic explanation, corrective feedback and phonological input enhancement. All of the instructional strategies were fully integrated within the social studies content. For example, many of the FFI lessons began with a brief review of the content to be covered in the language-focused activities (e.g. defining geographic terms or reviewing the geographic categories to apply). Two linguistic features were targeted based on their prevalence in academic texts and their differential status in terms of complexity (i. e. *passive* and *simple present*). Using a time-series research design, the students' performance was compared on three pretests before the intervention with their progress compared at the midpoint, immediately after the intervention, and six months later. Students completed six linguistic measures at each test session: four production tasks (2 oral and 2 written) and two receptive tasks (an oral and written grammaticality judgment test). The results indicated that learners made significant gains on only 2 of the language measures. An analysis of individual and group performance based on students' pretest knowledge of the target features (low, medium high) indicated greater improvement on the *passive* than the *simple present*. The finding that little progress was made was interpreted as evidence that integrated FFI may have prevented learners from noticing the language forms in the content. That is, the competing demands of

focusing on content and language simultaneously may have been too much for the learners to cope with given that their English language proficiency was low.

What emerges from a review of the limited studies described above is that there appears to be a slight advantage for planned FFI but that both planned and incidental FFI contribute positively to L2 learning. More research is needed to investigate this question. With respect to the integration of language-focused instruction within CBI there appear to be conflicting results. On the one hand, Valeo (2010, 2013) suggests that the lack of significant instructional effects on linguistic gains may have been due to the fact that the FFI was not sufficiently integrated with the content. On the other hand Trahey (2016) suggests that the modest gains observed in her study may be because the language focus was integrated with the content and as a result, the learners may have had difficulty focusing on language. She argues that because their proficiency level was low, the students may have had difficulty making the form-meaning connections and thus may have benefited from attention to language that was separated or isolated from content-based instruction. Valeo's learners' on the other hand, were more advanced in their L2 proficiency and thus, might have been able to benefit more from fully integrated FFI. Again, more research is needed with L2 learners at different proficiency levels and in different contexts to provide more information as to the most effective ways of combining attention to language and content keeping in mind the different profiles of different populations of learners. Nonetheless, the results from both of these studies are compelling examples of just how challenging it is for L2 learners to focus on language while processing curricular content.

Questions about the potential benefits of integrating or separating attention to language within meaning-based practice have been the focus of my own work. In 2008 Patsy Lightbown and I conceptualized two types of FFI – isolated and integrated. It is important to emphasize that in both types of FFI attention is given to form and meaning. The main

difference is that in isolated FFI, form and meaning-based practice are always separated and in integrated FFI there are always combined. Although there has been discussion about whether it is best to focus on form in integrated or separated ways in the pedagogical literature (e.g. Brumfit, 1984; Celcia-Murcia, 1991), there has been little research to investigate the effects of these two types of instruction on L2 learning. To my knowledge only a handful of studies exist. One study found that both isolated and integrated FFI were equally beneficial in the acquisition of vocabulary (File & Adams, 2010); another study reports advantages for integrated over isolated FFI in the learning of vocabulary and grammar (Elgun-Gunduz, Akcan & Bayyurt, 2012).

With my research group at the University of Toronto we carried out two quasi-experimental studies to investigate the contributions of isolated and integrated FFI to L2 learning. This research took place with adult learners of English as a second language who were not in content-based programs but in meaning-based CLT programs. Four classes of students participated with two classes receiving isolated FFI and two receiving integrated FFI. In the first study, the instruction was 12 hours in duration and delivered over 4 days (see Spada et al, 2014 for a description of the instructional materials). The focus of the instruction was the *passive*. Learners' knowledge of the target feature was tested before they received instruction, immediately after the instruction and again two weeks later. The tests included an error correction task in which students were required to correct ungrammatical sentences (i.e. sentences in which the *passive* construction is used incorrectly) and a picture-cued oral production task (see Spada et al., 2014 for a description of the language measures). The research questions investigated were whether isolated and integrated FFI are equally beneficial for L2 learning and whether the two types of instruction contribute to different types of L2 knowledge. Based on transfer appropriate processing theory (Blaxton, 1989; Segalowitz & Lightbown, 1999), our hypothesis was that L2 knowledge learned in isolated FFI would be more easily retrieved on the error correction test (i.e. explicit knowledge) and

that L2 knowledge learned in integrated FFI would be more easily retrieved on the picture-cued oral production task (i.e. implicit knowledge). The results revealed that learners in both groups improved on both measures. While there was some evidence that learners in the isolated group performed better on the error correction task and that learners in the integrated group performed better on the oral production task these differences were not statistically significant. Thus, there was no support for the hypothesis that the different types of FFI lead to different types of knowledge. Concerns about whether the oral production task was an appropriate measure of implicit knowledge led to a second study which included an oral elicited imitation task – a test argued to be a more valid measure of implicit knowledge (Ellis et al, 2009; Erlam, 2006). In that study two groups of learners were compared – an experimental group that received 4 hours of integrated FFI and a control group that received 4 hours of instruction in English idiomatic expressions. Once again the target feature was the *passive* and learners were tested on their knowledge of it immediately before and after instruction and again one week later. The tests included the error correction task, the oral production task and the elicited imitation task. Similar to the first study the results indicated improvement on all measures for both groups (Spada et al., 2011)

The finding that both isolated and integrated FFI are equally beneficial is consistent with other research we have carried out investigating teacher and learner opinions and preferences about the two types of instruction. Both groups expressed strong opinions about the value of each type of FFI (Valeo & Spada, 2015). These combined results raise some interesting questions as to when it might be preferable to use one type of FFI or another in one's pedagogical practice. As discussed in Spada and Lightbown (2008), isolated FFI may be more appropriate for the teaching of less salient features – one that are not easily heard or noticed in the input (e.g. 3<sup>rd</sup> person 's' in English). Isolated FFI might also be preferable with low level proficiency learners who are not able to make the form-meaning connections because of the competing cognitive demands of focusing on language and content at the same

time. On the other hand Integrated FFI might be beneficial for complex language features with rules that are difficult to teach in an isolated manner (e.g. articles in English). Integrated FFI might be more effective with task-natural structures, that is, structures that are more likely to occur in specific communicative tasks (i.e. see Loschky & Bley-Vroman, 1993 for a description and discussion of task-natural and task essential structures).

### **Other approaches to focusing on language in CBI**

So far I have been discussing different ways in which attention to language can be provided within meaning and content-based teaching primarily within the framework of form-focused instruction, a construct that is rooted in conceptions of L2 learning and teaching that are cognitively based and associated with ‘mainstream’ SLA research. Other ways of thinking about how to draw learners’ attention to form come from socially oriented perspectives on language teaching and learning, for example systemic functional linguistics (Halliday, 1978; 1994). A functional grammar perspective considers language in authentic contexts of language use, is discourse-based and makes explicit links between social context, meaning, words, grammar and text. Such an approach to language within CBI is intended to help L2 learners recognize language forms and functions that work together to make meaning (de Oliveira and Schleppegrell, 2015). While much of the research within a functional grammar framework tends to be descriptive in nature and few studies have measured language gains over time in relation to language-focused instruction, there are some exceptions. One of them is a case study by Spycher (2007) who followed a newcomer secondary school student who received instruction and practice on strategies used in academic writing. This included adopting objective means of expression and providing cohesion through conjunctions and reference markers. The results indicated that after explicit explanation of and practice with the features, including identifying them in discourse and editing texts in groups, the student

began to incorporate the corrections into later drafts of his own text, indicating his progression along the route towards a more authoritative stance in writing. In another study Shin (2009) worked with a group of eight motivated, college-bound learners over four class periods during which time the students worked on sentence-combining exercises. They received corrective feedback and metalinguistic explanation while they revised their sentences and presented justifications for their choices. The results revealed that while students at higher proficiency levels were able to generate and justify grammatical sentence combinations, lower-level learners were limited by developing vocabulary.

Another functionally-based approach to explicit language teaching within CBI that has been used with university-level learners is genre-based pedagogy. It is based on the assumption that academic knowledge does not consist of content and behaviours learned independently of language and literacy (Hyland, 2007; Rose & Martin 2012; Swales, 1990). Within this conceptual framework, knowledge, behaviours, and language are considered to develop symbiotically. In research to investigate explicit genre-based teaching based on systemic functional linguistics, learners are guided to identify textual features of specific genres (e.g. explanation, procedure, argument) and are asked to produce them on their own. Through a cycle of discovery, modeling, joint construction of a text followed by independent construction of a text, students are encouraged to draw explicit connections between language structure and purpose.

Research exploring different ways of addressing university-level learners' language learning needs across different disciplinary fields (e.g. film study, health and social care) is documented in a special issue of *Language Learning* edited by Coffin and Donohue (2015). Drawing on systemic functional grammar and more specifically on a Language as Social Semiotic (LASS) approach to teaching and learning, the researchers describe how this approach to teaching and learning takes account of the fact that "academic knowledge does

not consist of academic content and behaviours learned independently of language and literacy. Nor are language and literacy simply carriers of academic content and behavior” (Coffin & Donahue, 2015, p. 3-4). Included in this special issue are three research-based chapters that report on the impact of a LASS approach - focused on different genres and registers - on the acquisition of subject knowledge and the development of language.

While genre-based teaching has been found to lead to benefits in learners’ rhetorical development, metalanguage and motivation, there are some concerns about the methodology – one of them being the transfer of knowledge from one genre to another (Tardy, 2008, 2013). As a result, recommendations have been made for students to be taught how to analyze genres rather than learn static features of genres. For more information and examples of materials/exercises from the University of Sydney’s learning centre on genre-based learning go to the following URLs: [http://learningcentre.usyd.edu.au/clearer\\_writing/module2/sentence\\_stages/sent\\_stages\\_info.html](http://learningcentre.usyd.edu.au/clearer_writing/module2/sentence_stages/sent_stages_info.html) and <http://writesite.elearn.usyd.edu.au/m3/m3u5/index.htm>. Subject specific issues in science are addressed at: <http://learningcentre.usyd.edu.au/wrise/>.

## Conclusions

There is no doubt that the benefits of meaning and content-based language teaching are considerable. Providing learners with content that is engaging and relevant to their interests, ages and experiences is motivating. This is particularly the case in CBI where learners need to master the academic content and thus are highly motivated to learn the language they need to succeed. However, learners need guidance and support to do so. As Lightbown (2014) states with specific reference to content-based instruction: “After decades of research on language acquisition in CBLT in a variety of educational and social contexts, it



is clear that language acquisition does not ‘take care of itself’” (p. 129). The same can be said for L2 development in communicative language teaching programs.

There is compelling evidence from research in meaning and content-based instruction that systematic language-focused teaching is an essential component and that explicit instruction is often most effective. Preliminary findings also suggest that both integrated and isolated FFI are beneficial for L2 learning. While it is true that integrating learners’ attention to language within content rich input and practice is considered to be an effective way for learners’ to establish form/meaning connections and to understand how language functions are realized formally, there are also times when it is helpful to separate language-based instruction so that learners have an opportunity to direct their attention to language. This is particularly important with learners whose language proficiency is low and thus experience difficulty focusing on form and meaning simultaneously.

Decisions about how to best help second and foreign language learners to accomplish their linguistic and academic goals should be informed by relevant theory and research. Fortunately, there is theoretical support, pedagogical guidance, and research evidence from a variety of perspectives including form-focused instruction, functional grammar and genre-based pedagogy to help us make these decisions.

### References

- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 657–668.
- Brinton, D., M. A. Snow, and M. Wesche. (1990). *Content-based second language instruction*. New York: Newbury House.

- Brumfit, C. J. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge, England: Cambridge University Press.
- Burger, S., & Chrétien, M. (2001). The development of oral production in content-based second language courses at the University of Ottawa. *Canadian Modern Language Review*, 58(1), 84-102.
- Celce-Murcia, M. (1991). Discourse analysis and grammar instruction. *Annual Review of Applied Linguistics*, 11, 135–151.
- Cenoz, J., Genesee, F. & Gorter, D, (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35 (3), 243–262.
- Coffin, C. & Donohue, J. (2014). A language as social semiotic approach to teaching and learning in higher education. *Language Learning* 64, Supplement 1, 1-10.
- Coyle, D., P. Hood, and D. Marsh (2010). *Content and language integrated learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) Classrooms*. Amsterdam: John Benjamins.
- Day, Y, E., & Shapson, S. (1991). Integrating formal and functional approaches in language teaching in French immersion: An experimental study. *Language Learning*, 41(1), 25-58.
- de Oliveira, L. & Schleppegrell, M. J. (2015). *Focus on Grammar and Meaning*. Oxford, England: Oxford University Press.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.
- Edwards, H., Wesche, M., Krashen, S., Clement, R. & Kruidenier, B. (1984). Second-language acquisition through subject-matter learning: A study of sheltered psychology classes at the University of Ottawa. *The Canadian Modern Language Review*, 41(2), 268-282.

- Elgun-Gunduz, Z., Akcan, S., & Bayyurt, Y. (2012). Isolated form-focused instruction and grammatical form: Explicit school English classrooms in Turkey. *Language, Culture and Curriculum*, 25(2), 157–171.
- Ellis, R. (2009a). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). NY: Multilingual Matters.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27(3), 464–491.
- File, K.A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, 44(2), 222–249.
- Grim, F. (2008). Integrating focus on form in L2 content-enriched instruction lessons. *Foreign Language Annals*, 41(2), 321–346.
- Goo, J., Granena, G., Yilmaz, Y., Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revised and updated. In Rebuschat, P. (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). Amsterdam: John Benjamins.
- Halliday, M.A.K. (1978). *Language as social semiotic*. London: Hodder-Arnold.
- Halliday, M.A.K. (1994). *An introduction to functional grammar*. London: Edward Arnold.
- Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 156–174). New York: Cambridge University Press.
- Hauptman, P., Wesche, M., Ready, D. (1988). Second-Language Acquisition Through Subject-Matter Learning: A Follow-up Study at the University of Ottawa. *Language Learning*, 38 (3), 433–475.
- Howatt, A. (1984). *A history of English language teaching*. Oxford, England: Oxford University Press.

- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148–164.
- Klee, C. A., & Tedick, D. J. (1997). The undergraduate foreign language immersion program in Spanish at the University of Minnesota. In S. B. Stryker & B. L. Leaver (Eds.), *Content based instruction in foreign language education: Models and methods* (pp. 173–218). Washington, DC: Georgetown University Press.
- Krashen, S. (1982). *Principles and practices in second language acquisition*. Oxford, England: Pergamon Press.
- Lasagabaster, D. and Sierra, J.M. (2010). Immersion and CLIL in English: More differences than similarities, *ELT Journal*, 64(4), 367–375.
- Lightbown, P.M., & Spada, N. (2013). *How languages are learned*. Oxford, England: Oxford University Press.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. E. Crookes & S. M. E. Gass. (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, England: Multilingual Matters.
- Lyster, R. (2007). *Learning and Teaching Languages through Content: A Counterbalanced Approach*. Amsterdam: John Benjamins.
- Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15(3), 263–287.
- Musumeci, D. (1996). Teacher-learner negotiation in content-based instruction: Communication at cross-purposes? *Applied Linguistics*, 17(3), 286–325.
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417–528.
- Pica, T. (2002). Subject-matter content: How does it assist the interactional and linguistic needs of classroom language learners? *The Modern Language Journal*, 86(1), 1–19.
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford, England: Oxford University Press.

- Ready, D., & Wesche, M. (1992). An evaluation of the University of Ottawa's sheltered program: Language teaching strategies that work. In R. J. Courchêne, J. I. Glidden, J. St. John, & C. Therien (Eds.), *Comprehension-based second language teaching* (pp. 389–404). Ottawa: University of Ottawa Press.
- Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney school*. Equinox Publishing Ltd.: Sheffield, UK.
- Segalowitz, N., & Lightbown, P. M. (1999). Psycholinguistic approaches to SLA. *The Annual Review of Applied Linguistics*, 19, 23–43.
- Shin, S. (2009). Negotiating grammatical choices: Academic language learning by secondary ESL students. *System*, 37(3), 391–402.
- Spada, N. (2010). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching*, 44(2), 225-236.
- Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30, 73-87.
- Spada, N. & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60 (2), 1-46.
- Spada, N. & Lightbown, P.M. (2008). Form-focused instruction: Isolated or Integrated? *TESOL Quarterly*, 42(2), 181-207.
- Spada, N., Jessop, L., Suzuki, W., Tomita, Y. & Valeo (2014). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research*, 18(4), 453–473.
- Spada, N., Quinn, P., Shiu, J., Yalcin, S. (2011). The effects of form-focused instruction on explicit and implicit L2 knowledge. Paper presented at AAAL 2011, Chicago, March 2011.
- Spycher, P. (2007). Academic writing of adolescent English learners: Learning to use “although.” *Journal of Second Language Writing*, 16(4), 238–254.

- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.
- Tardy, C.M. (2012). Genre-based language teaching. *The Encyclopedia of Applied Linguistics*. Oxford, England: Wiley-Blackwell.
- Tardy, C. M. (2006). Researching first and second language genre learning: A comparative review and a look ahead. *Journal of Second Language Writing*, 15(2), 79–101.
- Trahey, M. (2016). *Form-focused instruction in the mainstream classroom: Effects on the grammatical development of adolescent English language learners*. Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.
- Trahey, M. (1996). Positive evidence in second language acquisition: Some long term effects. *Second Language Research* 12(2), 111–139.
- Trahey, M. & White, L. (1993). Positive evidence and pre-emption in the second language classroom. *Studies in Second Language Acquisition* 15, 181–204.
- Toth, P. D. (2004). When grammar instruction undermines cohesion in L2 Spanish classroom discourse. *The Modern Language Journal*, 88(1), 14-30.
- Valeo, A. (2015). Focus-on-form: Addressing grammatical accuracy in an occupation-specific language program. In M-A Christison, D. Christian, P. Duff, & N. Spada (Eds.), *Teaching and learning English grammar: Research findings and future directions* (21 - 33). New York, NY: Routledge.
- Valeo, A. (2013). The integration of language and content: Form-focused instruction in a content- based language program. *The Canadian Journal of Applied Linguistics*, 16(1), 25-50.
- Valeo, A. & Spada, N. (2016). Is there a better time to focus on grammar? Teacher and learner views. *TESOL Quarterly*, 50(3)314-339. doi: 10.1002/tesq.222

- Wesche, M., B., & Ready, D. (1985). Foreigner talk in the university classroom. In S.M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 89-114). Rowley, MA: Newbury House.
- Wong, W. (2003). Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, 13(2), 17-45.
- Zyzik, E., & Polio, C. (2008). Incidental focus on form in university Spanish literature courses. *The Modern Language Journal*, 92(1), 53-70.





# **The CEFR-J and its Impact on English Language Teaching in Japan**

Yukio Tono

Tokyo University of Foreign Studies

## **Abstract**

This paper provides a progress report on the CEFR-J, an adapted version of the Common European Framework of Reference for Languages (CEFR) for English language teaching in Japan and a series of studies to prepare reference level descriptions for the CEFR-J. Following a brief overview of the CEFR and the development of the CEFR-J, various related work such as Reference Level Descriptions are described and their significance and implications for future research are discussed.

**Keywords:** CEFR, Reference Level Description

## **Introduction**

The Common European Framework of Reference for Languages (CEFR) is a framework, published by the Council of Europe in 2001, which “provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe” (CEFR, 2001, p.1). It has various practical purposes, such as describing language policies, developing syllabuses, designing courses, developing learning materials, creating exams/tests, marking exams, evaluating language learning needs, continuous/self-assessment, and teacher training programs. The self-assessment grid (CEFR, 2001, pp.26-27), which is most widely-known as the typical framework for presentation, has a vertical scale (6 levels: A1 to C2) and a horizontal scale (5 skills: listening, reading, spoken interaction, spoken production and writing<sup>1</sup>). Each level is specified by illustrative can-do

descriptors, which describe what language learners can do with language at a certain level. The B1 level is called the “Threshold level,” beyond which learners become independent users of the language and able to function in society effectively. Thus, the primary aim of foreign language teaching during primary and secondary school is to provide a learning environment where learners will develop the necessary language skills up to the B1-level, and the B2 level is considered to be covered at the tertiary level of education.

The CEFR has become increasingly influential not only in Europe but in the rest of the world, especially in the area of language testing. For instance, major language proficiency tests for 25 different languages are now aligned to the CEFR levels (cf. April 2017 edition of the Wikipedia entry of the CEFR). The impact of the CEFR on curriculum or teaching was rather limited for the first ten years after its release (North, 2007), but the last decade has seen growing interest in adopting the CEFR for preparing contents of syllabuses, actual tasks and teaching materials. Projects such as the *Core Inventory for General English* by the British Council and the EAQUALS (2010), the *English Profile* by the team at Cambridge (<http://englishprofile.org>), and the *Global Scale of English* by Pearson (<https://www.english.com/gse>) are all such attempts at going beyond the list of illustrative descriptors and proposing concrete lists of vocabulary and grammar for each CEFR level, along with specific descriptors to go with them. The CEFR-J project (<http://www.cefr-j.org>) is also one such project.

The Ministry of Education, Culture, Sports, Science and Technology (hereafter, MEXT) is planning a reform of English language teaching in Japan as a part of revision of the Course of Study to be introduced in 2020. This new reform plan has three major changes: (1) the introduction of English as a subject at primary school, (2) the use of a coherent framework for benchmarking proficiency levels throughout the primary, secondary and tertiary levels, and (3) the reform of entrance examinations. All these plans have been developed by the CEFR and its widespread influence around the world. The plan to introduce English into primary school was made on the assumption that this will make it possible to attain B1 competence by

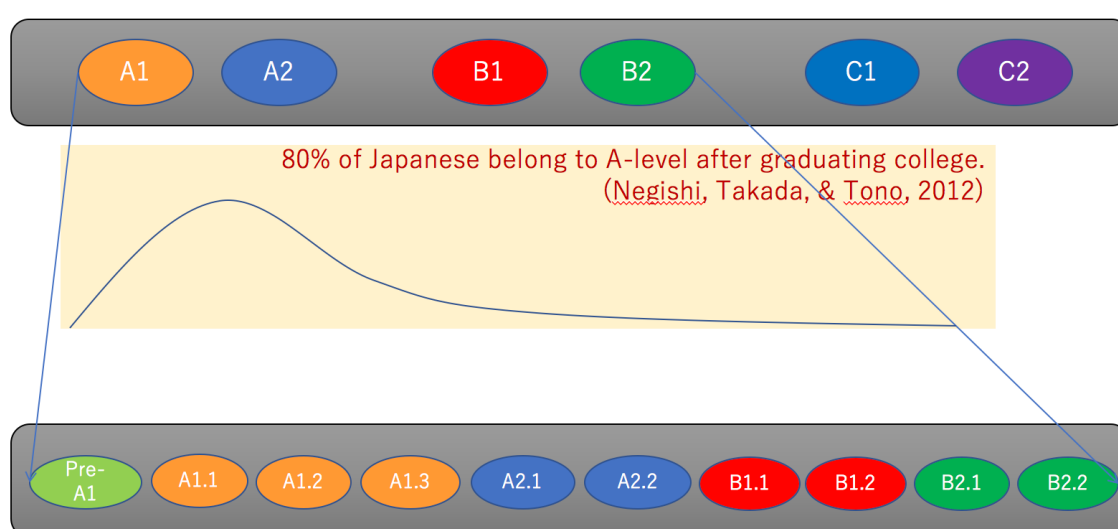
the end of upper secondary school. The transition from primary and secondary to tertiary education will also be designed based on the CEFR. Finally, the MEXT encourages the use of CEFR-based 4-skills proficiency tests in place of the existing university entrance exams focusing mainly on listening and reading. All these transitions will be carefully monitored by a nationwide survey of secondary school students' English proficiency. The MEXT hopes that if all the reform plans are implemented properly, English language teaching in Japanese schools will improve dramatically.

### **The development of the CEFR-J**

The CEFR-J is an adapted version of the CEFR for English language teaching and learning in Japan. Originally it was not our intention to develop a framework based on the CEFR. In fact, back in 2004, when we first received funding from JSPS and started investigating the learning objectives for Japanese learners of English at different proficiency levels, our initial goal was to find out how we should set our goals for English language learning at different levels of educational organizations. Through extensive surveys inside and outside Japan, we gradually came to realize the growing influence of the CEFR in Europe. The final report of the first KAKEN project in 2007 (Koike, 2007) stated that it would be unwise to develop from scratch our own proficiency scales and objectives. Rather, it claimed that we should seriously consider the possibility of applying the CEFR to the situation in Japan. That was the reason why we launched a new government funded project called the CEFR-J in 2008.

Several important modifications have been made for the CEFR-J. Approximately 80% of Japanese learners of English belong to A-levels even after they have gone through 10 years of English instruction at secondary and tertiary levels (Negishi, Takada, & Tono, 2012). Therefore, in order to properly diagnose the language abilities of this majority of users and in order to focus more on such beginner and false-beginner levels, we had to be able to specify

the A-levels in greater detail so that we could identify what exactly these false beginners can and cannot do. For pedagogical purposes, we see that finer grained scales for lower levels are desirable. Since our survey shows that most content to be taught at junior high school in Japan at the moment falls under A1 level, we need to make A1 level branch into smaller steps so that teachers find it easy to teach. This notion of branching is also recommended in the CEFR's original document. Such adaptations were made in other countries such as Finland. Figure 1 illustrates the branching made for the CEFR-J, compared with the original CEFR.



*Figure 1.* The structure of the CEFR-J, compared with the original CEFR.

The procedure of developing the CEFR-J is illustrated in Figure 2. We made three revisions before releasing version 1 in 2012. The alpha version was prepared after some planning and preparation phases from 2008 to 2009, during which we had a series of familiarisation and training sessions for writing can-do descriptors. Nick Saville at the then Cambridge ESOL and Tony Green at University of Bedfordshire were invited for these training sessions. The alpha version was revised by consulting with these experts and made into the beta version after the results of a “sorting exercise,” in which approximately 200 English teachers were asked to look at our descriptors ordered at random and put them in the

right order. This helped to decide which descriptors needed to be rewritten, due to a lack of clarity. In 2011, the beta version was tested by administering a large-scale can-do questionnaire taken by 5,468 students, based on which all the descriptors were calibrated using the two-parameter logistic model. Minor adjustments were made and the first version of the CEFR-J was released in March, 2012.

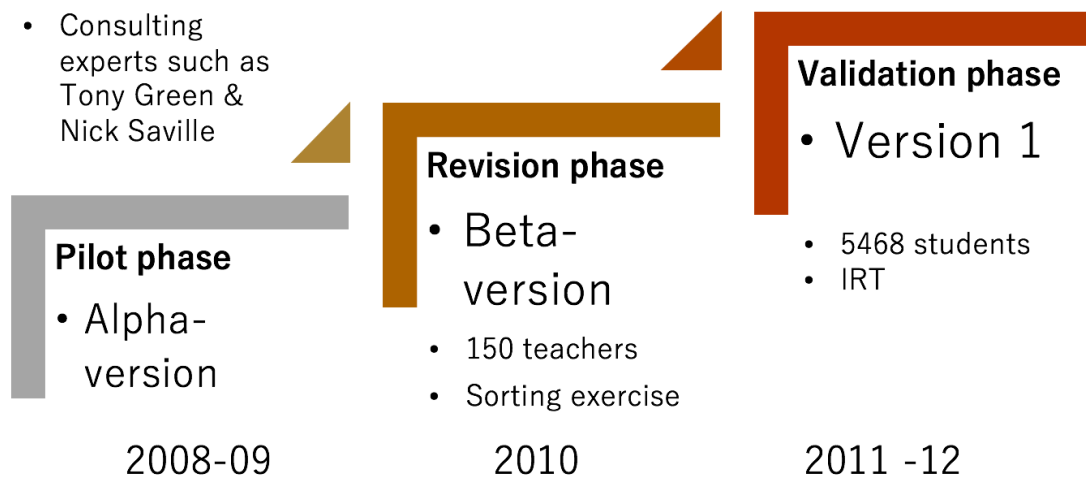


Figure 2. The process of the development of the CEFR-J.

The CEFR-J is freely available for research, teaching and commercial purposes as well. After the release, more than 5,000 organisations, schools, publishers, researchers and teachers have downloaded the framework and used it as a reference for the development of their own can-do lists, teaching materials or language tests. Some of the descriptors for lower levels have been adopted by the Council of Europe to supplement their A1-level descriptors. Pearson's Global Scale of English has adopted selected CEFR-J descriptors in their list of learning objectives. The Benesse Corporation used the CEFR-J for the basis of their can-do questionnaires for their nation-wide secondary school English skills survey funded by the MEXT.

### Developing resources for the CEFR-J

Since the public release of the CEFR-J, research has been carried out in the following two areas. One is to develop accompanying resources for the CEFR-J to facilitate its use for creating syllabuses, teaching materials or classroom tasks. For this purpose, resources such as the CEFR-J CAN-DO Descriptor Database and the Handbook of the CEFR-J were produced (Tono, 2013). The CEFR-J Wordlist is a wordlist developed based on the analysis of common vocabulary used in the EFL textbooks at primary and secondary schools in China, Korea, and Taiwan. This was a part of our analysis of ELT textbooks around the world in the first KAKEN project led by Ikuo Koike (2007). Those countries are similar to Japan in terms of school settings and EFL environment and introduce English as a school subject at Grade 3 or 4 of primary school. The analysis revealed that approximately 1,000 words were covered at Pre-A1 and A1 levels, with an additional 1,000 words at A2, plus 1700-1800 words at B1, and B2 respectively. Altogether, 5,639 words were identified. This figure, however, is relatively conservative since it only shows the words which commonly appear across the three countries. We decided to supplement the list by incorporating the words which appear uniquely in the English Vocabulary Profile, prepared by the English Profile team. As a result, our list contains 7,570 lemmas in total. See Table 1 for the breakdown of the CEFR-J Wordlist version 1:

Table 1

*The CEFR-J Wordlist Version 1*

CEFR -Level	Pre-A1	A1	A2	B1	B2	Total
Text analysis	976		1057	1884	1722	5639
Our initial target	1000		1000	2000	2000	6000
Final Version	1068		1358	2359	2785	7570

The unique feature of this wordlist is that it not only consists of the list of lemmas with part-of-speech information, but the noun entries also contain the information about semantic domains specified in *Threshold Level Series* (van Ek & Trim 1990), such as general and specific notions, which will help teachers and learners select a particular thematic domain of words. Table 2 illustrates how the wordlist can be searched for the part-of-speech information as well as the thematic domains defined by T-series. This is a useful feature, because if one wishes to use the CEFR-J can-do descriptors, he or she needs to consider the relevant expressions and vocabulary used to realize the particular function described in the descriptor.

Table 2

*Examples of the CEFR-J Wordlist (version 1)*

Entry	CEFR level	POS	Thematic domains	Thematic domains 2
activity	A1	n	Leisure activities	
actor	A1	n	Work and Jobs	Film
afternoon	A1	n		
age	A1	n	Personal information	
airplane	A1	n	Ways of travelling	
airport	A1	n	Travel and services vocab	Things in the town, shops and shopping
animal	A1	n		
answer	A1	n		
apple	A1	n	Food and drink	
apron	A1	n	Objects and rooms	
arm	A1	n	Personal information	
art	A1	n	Hobbies and pastimes	Education
aunt	A1	n	Family life	
baby	A1	n	Family life	
back	A1	n		
bag	A1	n	Shopping	Clothes
ball	A1	n	Hobbies and pastimes	
banana	A1	n	Food and drink	
bank	A1	n	Things in the town, shops and shopping	

Figure 3 illustrates the potential utility of this wordlist in creating can-do-based key expressions and a set of words to go with it. For instance, if the A1-level descriptor says “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.” (A1.2 Spoken interaction), then you can come up with key expressions such as “*I like ...*” “*I don’t like ...*” or “*Do you like ...?*” as likely phrases for A1-level users.

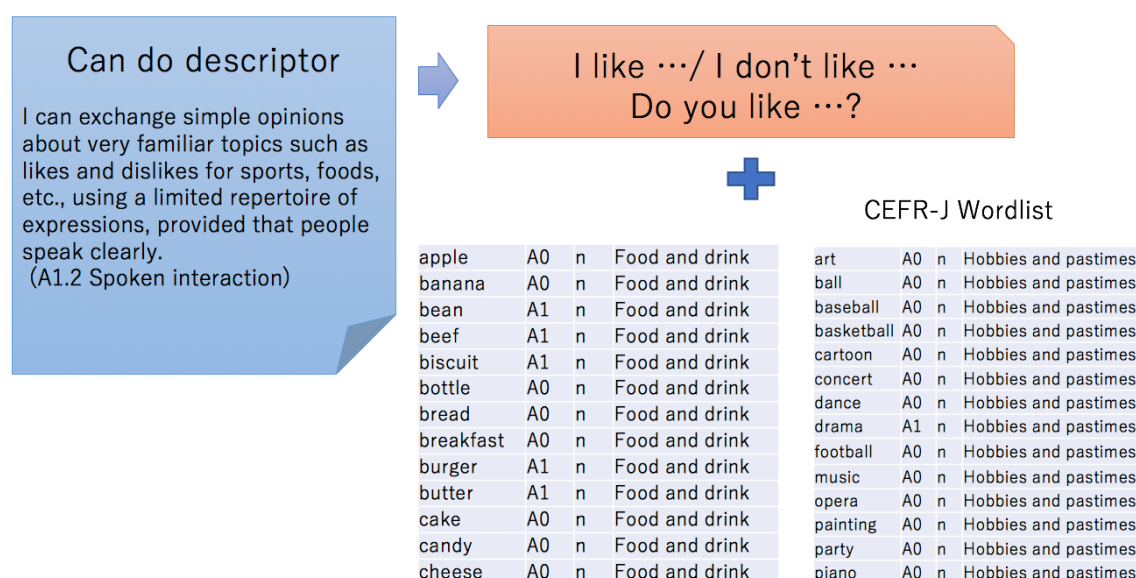


Figure 3. Using the CEFR-J Wordlist to extract key vocabulary for a particular can-do descriptor.

As you create some language activities, you might want to choose topics like sports or foods. Then the CEFR-J wordlist can provide the list of A1-level words with thematic categories “food and drink” or “hobbies and pastimes,” which will facilitate the process of integrating descriptors into specific language tasks.

The CEFR-J CAN-DO Descriptor Database was a collection of descriptors made for the European Language Portfolio (ELP, hereafter) in various countries. Altogether, 2,800 descriptors were gathered and analysed for its content, and the duplicated descriptors were eliminated and merged together. In total, 647 descriptors are in the final database (see Figure 4).



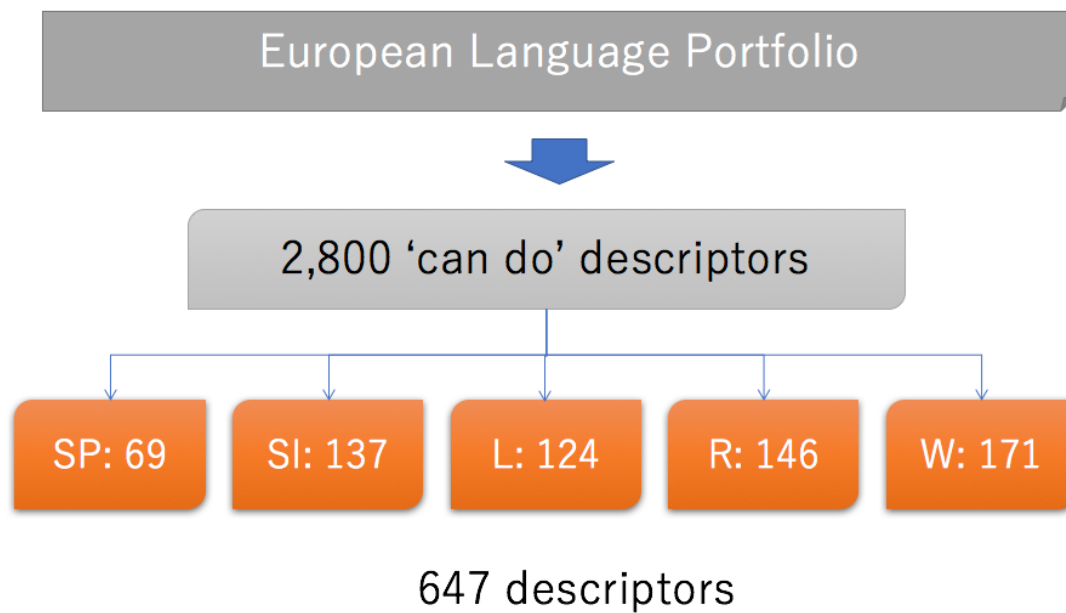


Figure 4. The design of the CEFR-J CAN-DO Descriptor Database.

For each descriptor, information such as CEFR levels, specific skill categories, and both original English descriptors and Japanese translations is available. For A-level descriptors, versions for younger children are also prepared (see Table 3 for the database format). This kind of database is useful when those who want to apply the CEFR-J to actual contexts need more concrete descriptors in addition to the 100 descriptors available for the self-assessment grid. This is sometimes needed if one wants to prepare syllabuses which cover all the available functions and situations. The descriptors used for the self-assessment grid are relatively neutral and not specific enough to readily introduce particular language (expressions and vocabulary) to realize what is illustrated in the descriptor. For this, one needs to have more concrete descriptors which are nested under a broader descriptor. The Database provides more than 600 descriptors available in various ELPs for different levels of learners, which will help syllabus and materials developers create concrete tasks and lessons for each course or programme.

Table 3.

*Sample Entries of the CEFR-J CAN-DO Descriptor Database*

Lev.	Category/ Code	ELP descriptor(s)	General descriptors (Japanese)	Descriptors for children (Japanese)
A1	IS1-A1	I can say who I am, ask someone's name and introduce someone.	自分が誰であるか言うことができ、相手の名前を尋ねたり、相手のことを紹介したりすることができる	自分の名前を言ったり、相手の名前を聞いたり、相手の紹介ができる
A1	IS1-A1-1	I can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.	簡単な質問をしたり、簡単な質問に答えたりすることができる。また必要性の高いことや身近な話題について発言し、反応することができる	簡単な質問をしたり、簡単な質問に答えたりすることができる。また必要なこと、身近なことについて話したり、質問に答えたりすることができる
A1	IS1-A1-1	I can make myself understood in a simple way but I am dependent on my partner being prepared to repeat more slowly and rephrase what I say and to help me to say what I want.	簡易な方法であれば通じるが、ゆっくり繰り返してくれたり、自分が言った事を言い直してくれたり、自分が言いたいことが言えるよう助けてくれるような相手に依存している	相手がゆっくり話してくれて、自分が言ったことを確認してくれるなど、相手が助けてくれれば簡単な英語で自分のことをわからせることができる

As we developed these resources, we also compiled the *Handbook of the CEFR-J* (Tono, 2013). The book has three parts; Part 1: a gentle introduction to the CEFR, Part 2: a report on the CEFR-J project and Part 3: how to use the CEFR-J. The handbook is a guide for researchers and teachers who are interested in using the CEFR-J for learning, teaching and assessment of English in Japan.

## Reference Level Descriptions and previous work

In 2012, we received a new funding from JSPS and started a new phase of the CEFR-J project. In this project, we aimed to prepare a set of grammar and text profile information as part of Reference Level Descriptions (RLDs, hereafter). RLDs are explained on the Council of Europe website as follows:

The descriptors specify progressive mastery of each skill, which is graded on a six-level scale (A1, A2, B1, B2, C1, C2). However, for operators, textbook authors and teachers, the specification set out in the CEFR may appear excessively broad. Work began on drafting CEFR specifications language by language. This new generation of reference level descriptions (in groups of six) is based on the CEFR level descriptors: it is a case of identifying the forms of a given language (words, grammar, etc), mastery of which corresponds to the communicational, socio-linguistic, formal and other competences defined by the CEFR. These transpositions of the CEFR into a given language are known as Reference Level Descriptions (RLDs) for national and regional languages.

(Retrieved from [http://www.coe.int/t/dg4/linguistic/DNR\\_EN.asp](http://www.coe.int/t/dg4/linguistic/DNR_EN.asp))

There are a few projects concerned with RLDs for English. Here, three projects beside our own will be briefly described. The *British Council/EAQUALS Core Inventory for General English* (North, et al. 2010) is led by Brian North and is a core curriculum inventory based around key language points for each level, including grammar, vocabulary, discourse markers and functions. The booklet contains sample syllabuses called ‘scenarios’ for different groups of learners with different learning goals, which help syllabus designers and teachers learn how to plan and implement their own ‘scenarios.’ The unique feature of the inventory is the list of text types. For instance, A1 level deals with text types such as:

signs and notices/ directions/ menus (simple)/ maps, tourist leaflets and posters/ advertisements (simple)/ timetables (simple)/ forms, invoices/ factual descriptions (visual)/ messages on postcards

Many of these text types are not always covered in English textbooks authorised by MEXT for secondary schools, and Japanese learners of English need to be more aware of the

various text types they are likely to be exposed to when they go abroad.

English Profile (EP, hereafter) is a project supported by the Council of Europe. The research was led by two departments of the University of Cambridge, UK: Cambridge University Press and Cambridge English Language Assessment. They investigated a large collection of Cambridge exam essays, called the Cambridge Learner Corpus, and identified “what aspects of English are typically learned at each CEFR level. This tells teachers, curriculum developers, course-book authors and test writers what is suitable for learning at each level.” (from the EP website). Major resources available for EP are the English Vocabulary Profile and the English Grammar Profile. The former is a list of words and phrases with the classification of CEFR levels. The EVP is quite useful because it not only assigns CEFR levels to different words, but also to different meanings and usages of a particular word. The original labelling was carried out by a team of lexicographers working for the *Cambridge Advanced Learner’s Dictionary* (3<sup>rd</sup> edition), but the information was rearranged for this profile purpose. The list is also enriched by real-life examples of Learner English from the Cambridge Learner Corpus. As is described earlier, our CEFR-J Wordlist was based on the textbook analysis, which is the input for learners, but the EVP is based on the learner writing as an output. This adds different characteristics to the two lists when compared. For example, the EVP’s CEFR levels attached to words are often one level higher than the ones in the CEFR-J Wordlist, which indicates that the CEFR level decision seems to be determined by whether you profile learning points based on L2 input or output. Another very useful resource from EP is the English Grammar Profile. It allows us to see “how learners develop competence in grammatical form and meaning, as well as pragmatic appropriateness, as they move up the CEFR levels. This provides us with typical, world-wide grammar profiles for each level.” (from the EGP website) It contains 1,239 items with information such as (a) large grammar categories, (b) subcategories, (c) CEFR levels, (d) lexical range, (e) FORM and/or USE, (f) can-do statement, and (g) examples from the

Cambridge Learner Corpus.

The Global Scale of English (GSE) Teacher Toolkit by Pearson Education is another useful RLD work. The GSE is a numerical scale developed by Pearson Education, which aligns various other language tests to one other on the GSE scale, including the CEFR itself. The strength of the GSE is its capacity to cover a broad range of language proficiency such as “below A1 level,” which is sometimes difficult for the current CEFR descriptors to properly describe. Another strength is in providing more detailed scaling for subdivisions of each CEFR level. The toolkit helps teachers access various RLD resources including learning objectives, grammar, and vocabulary by selecting appropriate target learner groups (young learners, academic learners, adult learners, professional learners). All the search results are accompanied by GSE scores, which enables teachers and learners to understand what learning objectives should be targeted and what grammar and vocabulary should be covered up to that level and onwards.

### **RLDs for the CEFR-J**

In the JSPS KAKAN project (Kiban A; No. 24242017; 2012-15), we conducted RLD research similar to previous projects such as EP or Core Inventory. There are two reasons why we had an independent RLD project. First, the CEFR-J has many sub-levels under A1 to B2, and it is desirable to specify grammar and vocabulary to go with each sub-level. For this purpose, the resources provided by EP or Core Inventory are not sufficient. Second, previous project reports on RLDs do not always specify the detail of how each item of grammar or vocabulary is assigned to a particular level. Overall methods were presented, but they did not make the actual data available. Thus, we had a genuine methodological interest in how to do RLDs properly. We tried to be as transparent as possible throughout all the stages of RLD work, and made sure that the procedure should be available as a standard for those who wish to do their own RLD research. In addition, we used corpus-based approaches similar to EP,

and our profiling technique was very different from theirs, which would be methodologically interesting to compare.

In our project, identification of the CEFR levels was considered a type of classification task defined in the field of Natural Language Processing (NLP, henceforth). Figure 5 illustrates this point. Basically, it involves supervised learning of features in the texts labelled with the CEFR levels. First, a machine will create a certain model based on a set of feature vectors from training texts with some class information, such as CEFR levels. Then the model will predict a CEFR level when a new text is given.

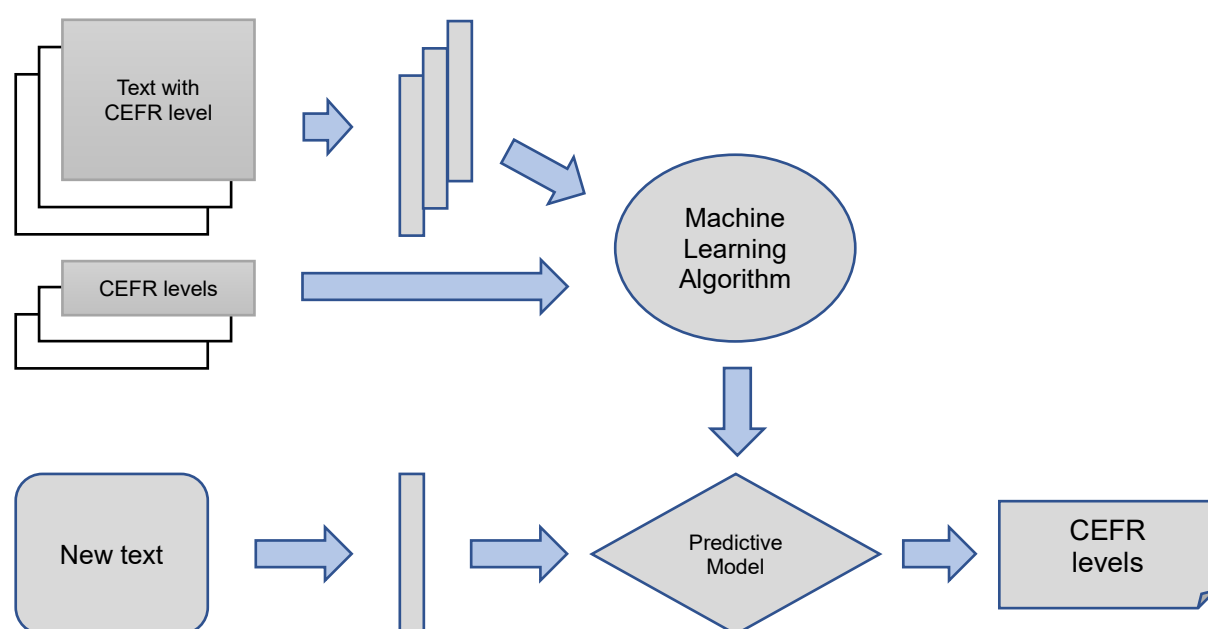


Figure 5. Supervised learning model used for the CEFR-J RLD project.

For the texts to feed in, we prepared two types of corpora, textbook corpora as ‘input’ and learner corpora as ‘output’. These two types of corpora were needed in order to do RLDs for both teaching and assessment purposes. The input corpus is a collection of CEFR course books published in the U.K. Since there is no CEFR-based English textbook published in Japan yet, course books published in the U.K. after the release of the CEFR in 2001 were collected. In total, 96 textbooks were gathered, scanned using OCR and prepared in an XML format. The corpus is tagged for CEFR levels, section tags for different skills (4 skills and

grammar), POS and lemma for each word. The data set (c. 1,640,000 tokens) is available for both normal text processing and concordancing using Sketch Engine (<http://www.sketchengine.co.uk>). The output corpus is a group of learner corpora collected by the principal researchers: JEFLL Corpus (Tono, 2012) and NICT JLE Corpus (Izumi et al. 2009). JEFLL Corpus is a collection of approximately 10,000 secondary school students' written compositions (size: 0.7 million), and NICT JLE Corpus is a collection of oral interview scripts by 1,280 test-takers (size: 2 million). Both sets of data were originally gathered without CEFR levels, but for this project all the texts were aligned to the CEFR levels.

Tono (2015) was a preliminary attempt at identifying criterial grammatical features useful for CEFR-level classification. A list of 158 grammar items used in 2 sentence patterns (declarative and negative) were selected based on the analysis of secondary school English textbooks published in Japan and a query syntax was written for retrieving these grammar items from the text corpora automatically. This part of the work was mainly done by the member of our KAKEN team, Yasutake Ishii (Ishii, 2016). Tono (2015) used Weka 3.6.12 to examine which grammatical items play an important role in classification. Since there are different machine learning algorithms, different classifiers were tested and compared against each other. Table 4 shows the results of classifier performance.

Table 4. *Classifier Performance (Weka 3.6.12)*

Classifier	PART decision list	BAGGING	Meta DECORATE	Meta nested dichotomy	Multilayer Perceptron
A1	0.720	0.846	0.846	0.800	0.720
A2	0.651	0.667	0.723	0.791	0.727
B1	0.517	0.520	0.615	0.610	0.643
B2	0.522	0.536	0.600	0.500	0.538
C1	0.353	0.167	0.154	0.000	0.167
C2	0.000	0.000	0.000	0.000	0.000
AVE.	0.555	0.562	<b>0.617</b>	0.584	0.594
CORRECTLY CLASSIFIED INSTANCES	55.79%	57.89%	<b>63.16%</b>	60.00%	61.05%

Table 4 (continued)

Classifier	Simple Logistic Regression	Support Vector Machine	Bayes Net	Bayes <sup>[L]</sup> <sub>SEP</sub> DMNBtext	Random Forests
A1	0.830	0.609	0.759	0.833	0.846
A2	0.696	0.638	0.650	0.683	0.698
B1	0.593	0.679	0.553	0.610	0.630
B2	0.533	0.667	0.517	0.577	0.519
C1	0.375	0.154	0.133	0.154	0.000
C2	0.000	0.000	0.000	0.000	0.000
AVE.	0.606	0.600	0.548	0.599	0.580
CORRECTLY CLASSIFIED INSTANCES	60.00%	61.05%	55.79%	61.05%	60.00%

Among ten different classifiers, DECORATE performed best. It is a fairly new algorithm and called a meta-learner. It builds diverse ensembles of classifiers by using specially constructed artificial training examples. Melville and Mooney (2003) shows that this technique is consistently more accurate than the base classifiers, Bagging and Random Forest. In the previous studies (Tono, 2013, 2014), Random Forest outperformed other classifiers, but DECORATE was not included at that time. Overall, meta-classifier types (DECORATE, nested dichotomy) performed better than the others. Knowing and choosing the right classifiers is also an important methodological decision for us, thus this kind of information is necessary for our evaluation.

Another important aspect of this approach is to identify which grammatical items play an important role in classification. In EP, these features are called ‘Criterial Features.’ A feature is criterial when the occurrence of this feature distinguishes one CEFR level from another. To prove this, we need information that this feature is significantly more frequent at the given CEFR level than at the other levels. To make matters more complicated, the CEFR level decision by humans is made not solely on a single feature but a bundle of features.



Therefore, we used this machine learning algorithm not only to create a model to predict the CEFR levels best, but also to select a set of grammatical features as the best predictors. Fortunately, NLP areas provide various methods of doing this. One area is called “attribute selection.” Machine learning creates a predictive model, in which some features are more salient than others in classification. There are two popular approaches to distinguish useful attributes from the rest: Filter methods and Wrapper methods. Filter methods use a proxy measure instead of the error rate to score a feature subset. This measure is chosen to be fast to compute, whilst still capturing the usefulness of the feature set. Common measures include the mutual information (MI), the pointwise mutual information, Pearson product-moment correlation coefficient, inter/intra class distance or the scores of significance tests for each class/feature combinations. Filters are usually less computationally intensive than wrappers, but they produce a feature set which is not tuned to a specific type of predictive model. Wrapper methods use a predictive model to score feature subsets. Each new subset is used to train a model, which is tested on a hold-out set. Counting the number of mistakes made on that hold-out set (the error rate of the model) gives the score for that subset. As wrapper methods train a new model for each subset, they are very computationally intensive, but usually provide the best performing feature set for that particular type of model (cf. Weka manual).

There are several different algorithms for each approach. Usually each approach combines ‘evaluator (evaluation of attributes)’ and ‘search (search for the best predictors)’. For instance, for the filter approach, “ReliefFAttributeEval” was used as the evaluator. It evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. Then for search, Ranker was used for ranking the attributes to find for the best. Table 5 shows the results of attribute selections using filter, wrapper and genetic approaches.

Table 5.

*The Results of Attribute Selections Using Different Approaches*

<b>FILTER:</b> ReliefFAttributeEval + Ranker	<b>WRAPPER:</b> WrapperSubsetEval + Best First	<b>GENETIC:</b> WrapperSubsetEval + Genetic Search
that_OBJ	ADV-attitude	PRP-You
AUX_be_able_to	PRES-PERF	PRP-POS
whether	PASSIVE-PRES-PRG	DPR_It_is
COMP_as	COMP-er	DAJ_this/that
PASSIVE-PAST	AUX_can	DAJ_these
PASSIVE-AUX	AUX_should	PHV
RC-SUB	AUX_PERF	PAST-PERF
RC-COMPX	IMP-NOT	FUTURE-PROG
SUBORD	INTRG_Will_you	FUTURE-PERF
All the features ranked	9 features selected	45 features selected

Table 6.

*The Performance of Two Classifiers With Selected Attributes Only*

Classifier	Meta DECORATE		Random Forest	
	BEFORE	<b>AFTER</b>	BEFORE	<b>AFTER</b>
A1	0.846	<b>0.880</b>	0.846	<b>0.917</b>
A2	0.723	<b>0.744</b>	0.698	<b>0.727</b>
B1	0.615	<b>0.679</b>	0.630	<b>0.566</b>
B2	0.600	<b>0.720</b>	0.519	<b>0.577</b>
C1	0.154	<b>0.533</b>	0.000	<b>0.500</b>
C2	0.000	<b>0.000</b>	0.000	<b>0.000</b>
AVE.	0.617	<b>0.710</b>	0.580	<b>0.640</b>
CORRECTLY CLASSIFIED INSTANCES	63.16%	<b>71.58%</b>	60.00%	<b>64.21%</b>

With this information, another trial for classifier performance was made using 9 features selected by the wrapper approach (WrapperSubseEval + Best First). This time, the two most promising classifiers, DECORATE and Random Forest were used. The results are shown in Table 6. As shown in Table 6, the accuracy rate of the prediction made by DECORATE increased from 63.16% to 71.58%. Random Forest also performed better (64.21% against 60.0%) with selected attributes. In this way, we can identify useful grammatical items as criterial features for better prediction of the CEFR levels.

### **The CEFR-J RLD project and future work**

In March, 2016, a symposium on the CEFR-J RLD project was held at Tokyo University of Foreign Studies, where we released a beta version of the Grammar Profile<sup>2</sup>. The profile contains tentative CEFR-J levels based on our analysis of the course book corpora, CEFR level information obtained from other RLD work such as Core Inventory, English Grammar Profile by EP, and Pearson GSE-LO level. SVM attribute weights are also provided for evaluation purposes. The CEFR-J Text Profile is also underway, but the two profiles will be complete as the second volume of the handbook will be published in early 2018.

In a new KAKEN project (2016-2019), we are planning to prepare language tasks for each CEFR-J can-do descriptor and their performance tests. This is an ambitious project, but together with all the information on grammar and vocabulary profiles, this new project will provide end-users of the CEFR-J with highly usable sets of materials to put the CEFR-J into practice. Commercially, companies such as Z-KAI have developed a new online learning programme based on the CEFR-J, called *Asteria*<sup>3</sup> and the 21<sup>st</sup> Century Learning Research Institute has also developed an assessment package based upon the CEFR-J, called *LIPHARE*<sup>4</sup>. All these activities clearly show that we are in the process of reform of English language teaching in Japan and the keyword is the CEFR. I do hope that this reform will turn out to be a great success, and to that end, we have to make a continuous effort to make the CEFR and

the CEFR-J known to teachers, learners and users of English. If all high school students can attain A2 level and beyond for the 5 skills described in the CEFR, it will surely broaden the horizon of Japan's younger generations to live their lives to the full in this globalized world.

### Notes

1. Originally, the CEFR had four kinds of language activities: reception, interaction, production, and mediation. For speaking and writing, both interaction and production are defined in the original document, but the self-assessment grid only covers spoken interaction and production, and written interaction and production were not distinguished.
2. The report on the CEFR-J RLD project was made available at the following URL:  
<http://www.cefr-j.org/sympo2016/TonoKaken2012-2015FinalReport.pdf>
3. Asteria: <http://www.zkai.co.jp/home/z-asteria/class/english.html>
4. LIPHARE: <https://www.21lri.co.jp/liphare/>

### References

- Ishii, Y. (2016). CEFR-J grammar profile kochiku no tame no eibunpoukoumoku no sentei, chushutsu, hindo-shuukei, seido-hyouka [Selection, extraction, frequency counts and evaluation of grammatical items in English for the CEFR-J grammar profile]. In Y. Tono (Ed.), *The CEFR-J RLD project: Developing grammar, text and error profiles using textbook & learner corpora. Grant-in-Aid for Scientific Research (A): No. 24242017.* (pp.31-41) An unpublished final report and collected papers. Tokyo University of Foreign Studies.
- Izumi, E., Uchimoto, K., & Isahara, H. (Eds.) (2004). *Nipponjin 1200-nin no eigo speaking corpus: The NICT JLE Corpus [A spoken corpus of 1200 Japanese EFL learners: NICT JLE Corpus]*. Tokyo: ALC Press.
- Koike, I. (Ed.) (2008). *Leading research to coordinate primary, secondary, and tertiary level*

- English education based on second language research*. Grant-in-Aid for Scientific Research (A): No.16202010. An unpublished final report and collected papers. Meikai University.
- Melville, P. & Mooney, R.J. (2003). Constructing diverse classifier ensembles using artificial training examples. *Proceedings of Eighteenth International Joint Conference on Artificial Intelligence* (the IJCAI-2003, Acapulco, Mexico), 505-510.
- Negishi, M., Takada, T., & Tono, Y. (2012). A progress report on the development of the CEFR-J. *Studies in Language Testing*, 36, 135-163.
- North, B. (2007). The CEFR common reference levels: Validated reference points and local strategies. In F. Goullier (Ed.), *Intergovernmental language policy forum report: The common European framework of reference for languages (CEFR) and the development of language policies: Challenges and responsibilities* (pp19-29). Strasbourg: Council of Europe Language Policy Division.
- North, B., Ortega, A., & Sheehan, S. (2010). *A core inventory for general English*. British Council & EAQUALS. Available from <https://englishagenda.britishcouncil.org/sites/default/files/attachments/books-british-council-eaquals-core-inventory.pdf>
- Tono, Y. (Ed.) (2012). *Research into development and validation of English language proficiency guidelines for Japanese learners of English at primary, secondary, and tertiary education*. Grant-in-Aid for Scientific Research (A): No. 20242011. An unpublished final report and collected papers. Tokyo University of Foreign Studies.
- Tono, Y. (Ed.) (2013). *The CEFR-J handbook: A resource book for using CAN-DO descriptors for English language teaching*. Tokyo: Taishukan.
- Tono, Y. (2013). Criterial feature extraction using parallel learner corpora and machine learning. In A. Diaz-Negrillo, N. Ballier, & P. Thompson (Eds.). *Automatic treatment and analysis of learner corpus data* (pp. 169-203). Amsterdam: John Benjamins.
- Tono, Y. (2014). *TALC in action: ten years on*. A plenary speech given at Teaching and

Language Corpora 2014. Lancaster, U.K.

Tono, Y. (2015). *Linguistic feature extraction and evaluation using machine learning to identify “criterial” grammar constructions for the CEFR levels*. A paper presented at Corpus Linguistics 2015. Lancaster, U.K.

Tono, Y. (2016). *The CEFR-J RLD project: Developing grammar, text and error profiles using textbook & learner corpora. Grant-in-Aid for Scientific Research (A): No. 24242017*. An unpublished final report and collected papers. Tokyo University of Foreign Studies.

Van Ek, J.A. & Trim, J.L.M. (1991). *Threshold level 1990*. Strasbourg: Council of Europe.

# **Selected Papers**

Research Articles





## **Linguistic Features of Discussion Sections of English Medical Research Paper Genres and Their Pedagogical Implications**

Motoko Asano

Osaka University

Judy Noguchi

Kobe Gakuin University

### **Abstract**

This study investigated the linguistic features of the discussion sections of medical research paper genres from the viewpoint of English for specific purposes (ESP) and English as an international language (EIL) to determine whether or not the use of language in original articles from the Japanese cardiovascular journal *Circulation Journal* (*CircJ*) could be a model for Japanese scientists involved in discipline-specific research article (RA) writing. Used as control corpora were RAs from the US cardiovascular journal *Circulation* (*Circ*), the *New England Journal of Medicine* (*NEJM*) and case reports (CRs) from the *NEJM*. Analyses of the moves and steps in the discussion section revealed common features; however, the *CircJ* texts appeared to be more similar to the *NEJM* RA texts than to the *Circ* texts and tended to use modal auxiliary verbs in a manner different from those in the control articles. The findings suggested that the *CircJ* articles in our corpora differed significantly enough, at least with regard to the use of words, to propose the building of mini-corpora consisting of authentic material of the target genre in a pedagogical setting as a way to become aware of rhetorical patterns and expressions frequently used by more experienced writers.

**Keywords:** English for specific purposes, English as an international language, corpus linguistics, medical research paper genres, move analysis

## **Introduction**

Writing up research in English for presentation to the global professional community is an essential skill for those conducting research in order to share their findings for scrutiny and consideration toward the construction of field knowledge (Robinson, Stoller, Costanza-Robinson, & Jones, 2008; Noguchi, 2016). At the top of the information hierarchy among genre texts in science is the journal research paper. Other important genres in science include research proposals and scientific posters mainly for members of the community in each specialized field, textbooks and laboratory manuals for students, and popular science articles for the general audience (Robinson et al., 2008).

Another important consideration with respect to the publication of research is the use of English as an international language (EIL). An international language is defined by Smith (1976, p. 38) to be ‘one which is used by people of different nations to communicate with one another.’ In such an environment, varieties of English are noted. Quirk and Stein (1990, p. 49) maintain that ‘any use-related variety must be expressed in terms of a particular user-related variety.’ One clear example of user-related variety is Singaporean English (Smith, 1976, p. 38). A use-related variety is reportedly employed to achieve a certain specific communicative goal, such as that observed with air traffic controllers in the aviation community (Quirk & Stein, 1990). Based on Quirk and Stein (1990), the research paper genre can also be thought of as an example of a use-related, discipline-specific variety of English, with the users sharing a scientific culture (Swales, 1990).

The theoretical framework proposed by Swales helped develop the idea of ESP. In ESP, genre is considered to ‘comprise a class of communicative events, the members of which share some set of communicative purposes’ (Swales, 1990, p. 58). Swales (1990) states that:

‘As a teacher of English, my own experience leads me to suppose that students can get a better handle on communicative affairs by concentrating, at least initially, on the sui generis feature of particular genre texts.’ (p. 18)

Working on the premise that genre texts for a specific field would display similar features, we compiled computer-based specialized corpora, which are ‘now beginning to provide some satisfactory solution to the problems’ related to identifying professional English usage (Noguchi, 2010).

Genre-based analyses of specialized corpora have been conducted extensively in the area of ESP. As a seminal work, the move analysis of 48 journal research papers of various fields including molecular physics, chemical pathology, and educational psychology led to the ‘Create a Research Space (CARS) model’ (Swales, 1990, p. 140) for the introduction sections, characterized by:

Move 1: Establishing a territory

Move 2: Establishing a niche

Move 3: Occupying the niche.

Each move has subcategories called ‘steps’ (Swales, 1981, 1990).

Nwogu (1997) applied the CARS model to the entire text of medical RAs investigating 15 UK and US medical research articles (RAs) of high-stakes journals such as *Lancet*, the *British Medical Journal (BMJ)*, the *New England Journal of Medicine (NEJM)*, the *Journal of Clinical Investigation (JcL Inv)* and the *Journal of the American Medical Association (JAMA)*. He determined the moves and steps in each of the introduction, methods, results, and discussion (IMRD) sections (Swales, 1990, p. 134) of the articles.

Based on the moves proposed by Nwogu (1997), El Malik and Nesi (2008) studied 10 medical RAs each written by UK and Sudanese researchers and found that the Sudanese researchers tended to write shorter sentences than the UK researchers whereas the UK researchers tended to use a greater number of the nominal triplets *the (noun) of* and *a (noun) of* than the Sudanese researchers.

In 2010, Maci studied the moves and steps of the discussion sections in 50 US and Italian medical RAs in cardiology. Maci’s examination into the discussion sections of articles

from the *Italian Heart Journal* or the *Journal of Cardiovascular Medicine* and the US journal *Circulation* (*Circ*) revealed that Italian authors tended to use hedges less frequently in the presence of supporting evidence from previous studies (Maci, 2010). In a study of engineering articles by Maswana, Kanamaru, and Tajino (2015), analyses of the entire text of technical research papers were performed by six engineering researchers based on the Swales analytical framework (1990), and the moves and steps of the abstract and IMRD (A-IMRD) texts were identified.

The medical research paper genre is regarded as having several subgenres including medical RAs and case reports (CRs). Salager-Meyer (1989, p. 23) defines subgenres, maintaining that research articles are characterized by ‘an enquiry style’ (Salager-Meyer, 1989, p. 27) as they include the process of answering the proposed research question or hypotheses based mainly on randomized clinical trials while CRs generally involve a narrative description and explanation of one or multiple patients with an illness. Medical CRs have been considered less important than medical RAs in terms of a genre but have certain important features (Helan, 2012, p. 58). Today, CRs play a substantial role in changing specific medical conventions in the health care community, including urging necessary steps to be taken by the regulatory authorities, such as the Food and Drug Administration of the United States (Taylor, 2005, p. 144). Medical CRs are also important from a pedagogical viewpoint as they are one of the first article types that medical students or clinicians write and could be ‘a pathway to an early publication on their curriculum vitae’ (Taylor, 2005, p. 143).

Quantitative analysis is useful for studying specialized corpora. Cluster analysis was used by Umesaki (2000) to examine linguistic similarities and differences of oral presentation transcriptions recorded at international conferences and the written versions published in the proceedings of the conferences in a natural science field. Cluster analysis was also used in a study by Kobayashi (2011) to illustrate different uses of metadiscourse markers in textbooks of English for Japanese secondary school students and in the International Corpus of Learner

English written by Japanese undergraduates (ICLE-JP). Another method of interest is the machine-learning technique called Random Forests, which has been used in a study ‘to detect rhetorical features that characterize Dickens’s text’ (Tabata, 2016, p. 22) and could classify the Dickens’s texts and major 18th- and 19th-century British fiction texts with good accuracy (Tabata, 2016). The Random Forests technique has also been used to classify RAs written by native and non-native speakers with metadiscourse markers as variables (Kobayashi, Tanaka, Tomiura, 2011).

RAs from the US and Japanese cardiovascular journals *Circ* and *Circulation Journal* (*CircJ*), respectively, and the *NEJM*, have been examined in a previous study (Asano, 2016a) by means of Random Forests, with the articles being classified using words as variables. The RAs from *Circ* and *CircJ* formed two distinct clusters; one with almost all of the RAs from *Circ* and the other from all of the RAs from *CircJ* when analyzed using cluster analysis with the 200, 500, and 1000 most frequent words as variables in another previous study (Asano, 2016b). The results of these two studies indicated that RAs of the journals studied may have distinct linguistic features that are related to the target audience.

In the present study, we compared the features of language used in the texts of the RAs as a whole as well as that in the discussion section from *CircJ* with those from *Circ* and from the *NEJM* and case reports from the *NEJM* for the purpose of investigating whether or not the RAs from *CircJ* could be a model for Japanese researchers. Our goal is to provide suggestions for the teaching of medical research paper genres and also offer implications for building an EIL model of these genres.

The discussion section was chosen for close analysis of moves and steps (Swales, 1981, 1990) because it is regarded as being the most argumentative among all of the sections in RAs (Horton, 1995). This section is where research paper writers must use their linguistic abilities to express their viewpoint by employing rhetorical devices such as modal auxiliary verbs *may* and *might* for stating various degrees of probability (Huckin & Olsen, 1983; Maci, 2010).

This section is especially important for persuading the audience to accept one's research findings and their implications (Swales, 1990; Swales & Feak, 2012). Our findings illuminate the process of communication via a journal published by a Japanese professional society and reveal an EIL use-related type of communication.

## Materials and Methods

### Corpora

Corpora were compiled with ten randomly-chosen articles each from the 50 RAs randomly downloaded from the *CircJ* and *Circ websites* as well as from *NEJM* called the *NEJM Original Articles (NEJM-OA)* websites, and ten case reports (CRs) from the *NEJM* case reports called the *NEJM Brief Reports (NEJM-BR)*, prepared in a similar manner (Table 1).<sup>1</sup>

Table 1

### Corpora

Journal Name	Country	Subgenre	Presumed main audience
<i>Circulation Journal (CircJ)</i>	Japan <sup>2</sup>	RA	Cardiovascular experts
<i>Circulation (Circ)</i>	US	RA	Cardiovascular experts
<i>NEJM Original Articles (NEJM-OA)</i>	US	RA	Physicians
<i>NEJM Brief Reports (NEJM-BR)</i>	US	CR	Physicians

Note. RA = Research Articles; CR = Case Reports.

*CircJ* is the official journal of the Japanese Circulation Society (*CircJ*, 2017), and *Circ* is affiliated with the American Heart Association (*Circ*, 2017); the two journals mainly target experts in cardiovascular medicine. *NEJM* belongs to a division of the Massachusetts Medical Society (*NEJM*, 2017: para 3) and is reportedly 'the most widely read, cited, and influential general medical periodical in the world' for physicians.

<sup>1</sup> The bibliographic information of the RAs and CRs in the corpora is available upon request.

<sup>2</sup> All of the first authors had Japanese names and were at Japanese institutions (Wood, 2001).

### **Quantitative Analyses of the Entire Text**

The corpora for this study were analyzed quantitatively and qualitatively according to the ESP approach. The entire text was observed macroscopically to examine whether or not the text had an abstract and the IMRD structure.

The word frequency of each text was determined to obtain the number of word types and word tokens, and the type/token ratio (TTR; Biber, 1988) was calculated by dividing the number of word types by the number of word tokens. The number of types and tokens was further analyzed to obtain the Guiraud's Index (GI), which is calculated by the number of word types divided by the square root of the number of word tokens (Hulstijn, 2010). GI is considered to provide a more accurate representation of the type versus token characteristics in longer texts in which several of the same words, such as articles and prepositions, are used repeatedly (Daller, Van Hout, & Treffers-Daller, 2003). Types and tokens were determined by CasualConc (Imao, 2015), and TTR and GI were calculated using Microsoft Excel 2013.

For the entire text, a hierarchical cluster analysis was performed using the 1000 most frequent words as variables. Hierarchical cluster analysis is an unsupervised method that groups data into a hierarchy according to similarities between the members (McEnery & Hardie, 2012, p. 53). The hierarchical cluster analysis undertaken in this study used scaled Euclidean distance matrix with Ward's method (Ward, 1963; Kobayashi, 2011) with the hierarchical cluster analysis package (hclust) of R (for Mac OS X Cocoa GUI, Version 3.2.3).

### **Qualitative and Quantitative Analyses of the Discussion Section**

The discussion section of the articles was examined both qualitatively and quantitatively. TTR and GI in the discussion section were calculated in the same manner as those calculated for the entire text. Moves and steps of the discussion section were coded by three coders, i.e., one of the authors and two experts in the medical field under the supervision of the other author. All coders as well as the supervisor shared one spreadsheet, using Microsoft Excel 2013, in which each sentence in the discussion section was placed in a cell

for coding. When the coding was completed, the coders met to discuss handling of the codes for those sentences for which different codes were given by the individual coders and came to a final agreement regarding the coding for each sentence.

The texts in the discussion section were then further analyzed into moves and steps, and the 200 most frequent words were used to perform a hierarchical cluster analysis in the same manner as that for the entire text.

## Results

### Quantitative Analyses of the Entire Text

Texts from all four corpora had identifiable structures which differed slightly among the corpora. *CircJ*, *Circ*, and *NEJM-OA* had the IMRD structure. The two cardiovascular journals, *CircJ* and *Circ*, had separate discussion and conclusion sections whereas the two were combined in the discussion section in the original articles of *NEJM-OA*. *NEJM-BR* had several other structural patterns.<sup>3</sup>

The types, tokens, TTR, and GI are shown in Table 2. TTR and GI were higher in the *NEJM-BR* than in the other three corpora. *CircJ*, *Circ*, and *NEJM-OA* had similar GI values although the maximum number of words allowed for each journal differs.

Table 2

#### *Types and Tokens of the Entire Texts of Each Corpus*

Journal Name (Country)	Number of Articles	Maximum Words <sup>a</sup>	Type	Token	TTR (%)	GI
<i>CircJ</i> (Japan)	10	6,000	3,179	32,141	9.9	17.7
<i>Circ</i> (US)	10	7,000	3,829	42,287	9.1	18.6
<i>NEJM-OA</i> (US)	10	2,700	3,497	35,660	9.8	18.5
<i>NEJM-BR</i> (US)	10	2,000	3,946	25,364	15.6	24.8

Note. TTR = type/token ratio; GI = Guiraud's Index.

<sup>a</sup>Maximum number of words allowed for each text.

<sup>3</sup>The structural variations included methods, case report, results, and discussion (*NEJM-BR2*) and case report(s), methods, results, and discussion (*NEJM-BR6*, *NEJM-BR7*, *NEJM-BR8*, and *NEJM-BR9*).



A cluster analysis based on the 1000 most frequent words showed a group of *NEJM-BR* (Rectangle (a) in Figure 1) and a group of *CircJ*, *Circ*, and *NEJM-OA* as two large distinct clusters, indicating that the words used in the CRs for *NEJM-BR* are different from those used in RAs for the three journals.<sup>4</sup>

One RA from *NEJM-OA* (*NEJM-OA17*) in a small rectangle (b) in the cluster of *NEJM-BR* described research of a drug for pregnant women, with descriptions of the course and complications in the women and their children; the style of this RA seemed to be similar to that of CR articles.

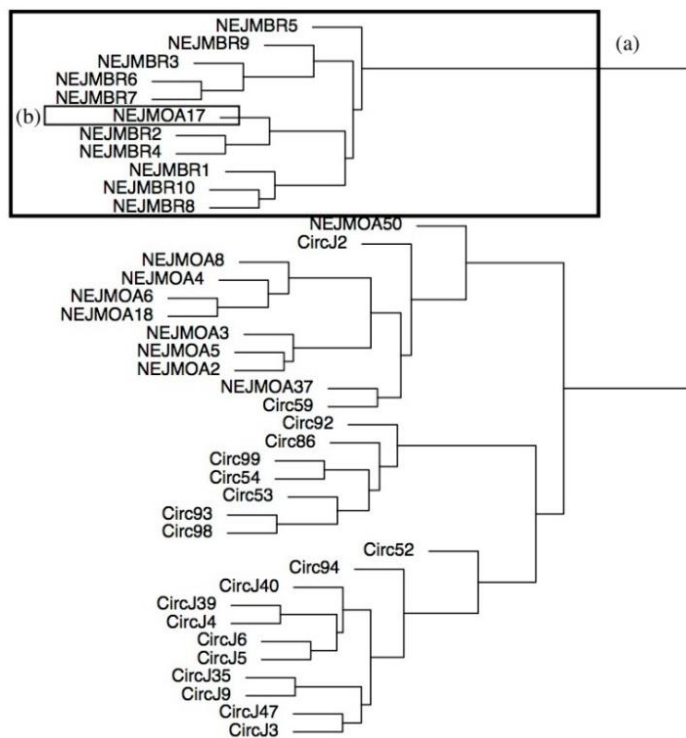


Figure 1. Cluster analysis of all texts from the four corpora.

## Qualitative and Quantitative Analyses of the Discussion Section

**Lexical items.** The types and tokens in the discussion section were determined, and the TTR and GI were calculated (Table 3).

<sup>4</sup> The 1000 most frequent words used for cluster analysis is available upon request.

Table 3

*Types and Tokens of the Discussion Section*

Journal	Entire Text		Discussion					
	Type	Token	% Type in Discussion	% Token in Discussion	Type	Token	TTR (%)	GI
<i>CircJ</i>	3,179	32,141	58.1%	34.8%	1,846	11,169	16.5	17.5
<i>Circ</i>	3,829	42,287	56.9%	30.5%	2,178	12,883	16.9	19.2
<i>NEJM-OA</i>	3,497	35,660	47.2%	21.7%	1,650	7,747	21.3	18.7
<i>NEJM-BR</i>	3,946	25,364	45.6%	25.4%	1,800	6,446	27.9	22.4

The discussion sections of the four corpora accounted for 21.7% to 34.8% of the entire text, with the discussion section of *CircJ* being the longest among the four. The reason for this lengthiness was explained by one specialist informant as possibly being due to the fact that articles in *CircJ* were likely to have been initially submitted to an overseas journal such as *Circ*, but having been rejected, were rewritten by adding more information to the discussion section and then submitted to *CircJ*. Similar to the entire text, TTR and GI were higher for *NEJM-BR* than for the original article genre. GI values were similar among *CircJ*, *Circ*, and *NEJM-OA*.

**Moves and steps.** The moves and steps of the discussion section of technical or medical RAs have been identified by several researchers (Nwogu, 1997; Maci, 2010; Maswana et al., 2015). In the present study, the discussion sections of the journal articles that we analyzed were found to have the following moves and steps (Table 4).

Table 4

*Proposed Moves and Steps for the Discussion Section*

Move / Step
<u>Move JD-A: Highlighting overall research purpose and results as well as their significance</u>
Move JD-A10: Highlighting overall research purpose
Step JD-A11: Presenting background information (established knowledge)
Step JD-A12: Reference to previous literature (and eventually to a gap within previous literature)
Step JD-A13: Reference to main research purpose (including hypothesis)
Move JD-A20: Highlighting overall results and their significance
<u>Move JD-B: Explaining specific research methods and outcomes</u>
Move JD-B10: Explaining specific research outcomes
Step JD-B11: Stating a specific outcome
Step JD-B12: Interpreting the outcome
Step JD-B13: Indicating significance of the outcome
Step JD-B14: Contrasting present and previous outcomes
Step JD-B15: Indicating limitations of outcomes
Move JD-B20: Explaining specific methods
Step JD-B21: Stating specific methods
Step JD-B22: Interpreting the methods
Step JD-B23: Indicating significance of the methods
Step JD-B24: Contrasting present and previous methods
Step JD-B25: Indicating limitations of methods

(table continues)

Table 4 (continued)

Move / Step
<u>Move JD-C: Stating research conclusions</u>
Move JD-C10: Stating research conclusions
Step JD-C11: Indicating research implications
Step JD-C12: Promoting further research
<u>Move JD-Z: Introducing new information</u>
Move JD-Z10: Introducing new information

Moves JD-A, B, and C, as well as the steps in each move, in Table 4 correspond to Moves 9, 10, and 11, respectively, that were proposed by Nwogu (1997, p. 135). J represents *journals* and D is for *discussion*.

***Move JD-A: Highlighting overall research purpose and results as well as their significance.*** Move JD-A10 is similar to Maci's Move 1 (Maci, 2010, p. 349) and reports overall research purposes frequently by referring to previous literature. In Step JD-A11, the authors present background information:

(1) Genetic subtypes of *T. cruzi* *vary according to* the geographic location. [emphasis added] (*NEJM-OA4*)

In the subsequent step (Step JD-A12), the authors refer to previous literature:

(2) Variable responses to benznidazole *have been reported previously* and may have contributed to our neutral findings. [emphasis added] (*NEJM-OA4*)

In Step JD-A13, the authors refer to the main research purpose. In this step, the authors

sometimes refer to the research hypothesis:

- (3) *Our goal was to identify* promising signals of activity in individual tumor types that could be pursued in subsequent studies with statistically robust efficacy end points or through protocol amendment and expanded enrollment in the current study.  
[emphasis added] (*NEJM-OA3*)

Move JD-A20 is similar to Move 9 (Nwogu, 1997, p. 125; Maswana et. al., 2015, p. 7) and states the overall results and their significance. RAs of both *CircJ* and *Circ* have similarities in form, showing all major findings in an itemized manner in one sentence:

- (4) *The major findings of the study were that* (1) QOL improved after LVAD implantation, and those patients' QOL was better than that of extracorporeal LVAD patients, and (2) lower postoperative albumin levels and RVF are independent factors related to lower physical QOL at 3 months after LVAD implantation. [emphasis added] (*CircJ5*)

**Move JD-B: Explaining specific research methods and outcomes.** Step JD-B11 for stating the results objectively was one of the steps that appeared most frequently in the discussion section. This move is slightly different from those of Maci (2010) or Maswana et al. (2015) in that Step JD-B11 includes the statement of specific research methods as well as that of specific outcomes; only the statement of specific research outcomes is given by Maci (2010) or Maswana et al. (2015). We found that the inclusion of a step for statement of research methods would facilitate understanding of the discussion section and could be used to explain the advantages or disadvantages of the methods used for the study in the discussion section. In this step, reporting verbs such as *found* and *observed* were often used.

- (5) *We found* that more than one fourth of ischemic stroke patients did not receive statins at hospital discharge. [emphasis added] (*Circ* 98)

The phrase *we found* appeared seven times in the *Circ* articles but only twice and once, respectively, in *NEJM-OA* and *CircJ*. The phrase *we observed* was also found four times and once in *Circ* and *NEJM-OA*, respectively, but not in *CircJ*. The phrase *be observed* appeared four times in *CircJ* but only once each in *Circ* and *NEJM-OA*. These observations point to more frequent use of the active voice in the American journals when describing findings (Ichihara, 1982; Lock, 1977).

In Step JD-B12, the authors provided explanations for specific results and used modal auxiliary verbs as boosters or softeners. In this step, we found the modal auxiliary verb *should* four times in *Circ* and once each in *NEJM-OA* and *NEJM-BR* but not in *CircJ* (Standardized frequency is shown in Table 6). In most cases, *should* was used to express obligation (Huckin & Olsen, 1983, p. 542); however, in one sentence in *NEJM-BR*, *should* was used to express probability (Sentence (6)).

- (6) As a marker of gonadal differentiation, the transcription factor GATA4 *should be absent in* adrenal cells, so its elevated levels point to dedifferentiation toward their common adrenal-gonadal precursor cell type. [emphasis added] (*NEJM-BR*8)

Steps JD-B13 and JD-B14 were not seen very frequently, but both steps appeared to help authors argue for their results. In Step JD-B13, the authors expressed the strength of the results obtained in their study (Sentence (7)), and in Step JD-B14, the authors compared their study results with those in other studies (Sentence (8)). One of the hint expressions, or phrases most frequently used to guide readers identify moves (Tojo, Hayashi, & Noguchi, 2014), in

Step JD-B14 was *consistent with ...*. This phrase was used to support the study results by referring to similar results obtained in previous studies and appeared in move JD-B14 three times in *CircJ*, 5 times in *Circ*, and 4 times in *NEJM-OA*, but only once in *NEJM-BR*.

- (7) Therefore, *we are confident that* our results are robust *and that we would have detected clinically significant effects of* hypoglycemia on neurocognitive processing. [emphasis added] (*NEJM-OA8*)
- (8) The findings in our trial *are consistent with* those from nonrandomized comparisons of these strategies. [emphasis added] (*NEJM-OA8*)

The statement of the limitations of the study being reported appears in Step JD-B15. As explained by Maci (2010, p. 358), ‘the author is merely a narrator, voicing possible limitations.’ Modal auxiliary verbs are used to point out presumed limitations or biases (Sentence (9)). However, the authors of the *CircJ* articles used *could* as the past tense of *can* and did not use it to make hypothetical statements (Huckin & Olsen, 1983, p. 547).

- (9) Although our study was a large trial involving patients with established Chagas’ cardiomyopathy, *we could have missed* small differences in risk (e.g., a relative risk reduction of 10%). [emphasis added] (*NEJM-OA4*)

**Move JD-C: Stating research conclusions.** Move JD-C is similar to the move labeled as ‘Stating research conclusions’ that includes ‘c1: indicating research implications’ and ‘c2: promoting further research’ (Maswana et al., 2015, p. 10).

Step JD-C11 represents a step for indicating research implications. Most of the *NEJM-OA* articles appeared to have a sentence that starts with *In conclusion* to introduce the research

implications. The phrase *In conclusion* is frequently followed by a sentence that includes a reporting verb such as *found* and *suggest*. As *Circ* and *CircJ* articles had a separate conclusion section, such a phrase was not needed.

- (10) *In conclusion, we found that the BRAF V600 mutation is a targetable oncogene in some, but not all, cancer types. [emphasis added] (NEJM-OA3)*

Step JD-C12, which states author suggestions for promoting further research, was present only in seven of the ten *CircJ* articles and six of the ten articles each for *Circ*, *NEJM-OA*, and *NEJM-BR*. This finding was similar to that reported by El Malik and Nesi (2008), who reported that this step occurred only in seven of the ten RAs written by Sudanese authors. El Malik and Nesi (2008, p. 93) concluded that ‘Sudanese writers may be unwilling to promote further research ... for fear of encouraging rival research groups in an environment where there is intense competition for funding.’ The findings in the present study are consistent with those of El Malik and Nesi (2008).

- (11) Their effect on clinical outcome *remains to be determined* in phase 3 studies.  
[emphasis added] (*NEJM-OA18*)

**Move JD-Z: Introducing new information.** Although the total word count was small for Move JD-Z (Figure 2), this move is considered to be very important from the pedagogical perspective because the sentence in Move JD-Z is sometimes given in the present tense, a structure that is rather unfamiliar to Japanese writers in this context:

- (12) *Our study has some potential limitations. [emphasis added] (NEJM-OA6)*



**Word count by move and step.** The total number of words, or tokens, were determined by move and by step (Figure 2). Sentences coded for more than two steps were excluded.

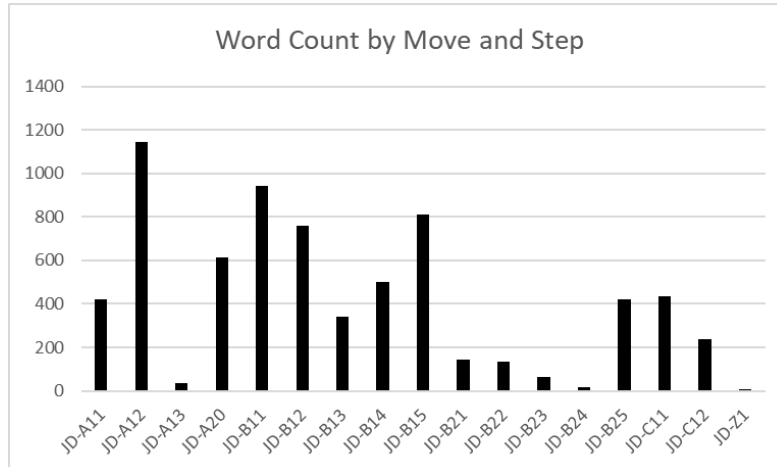


Figure 2. Bar plot of word count by move and step.

Table 5

*Major Moves and Steps in the Discussion Section*

Move/Step
Move JD-A20: Highlighting overall research outcome
Step JD-B11: Stating a specific outcome
Step JD-B12: Interpreting the outcome
Step JD-B15: Indicating limitation of outcome
Step JD-C11: Indicating research implications
Step JD-C12: Promoting further research

Table 5 shows the proposed major moves and steps in the discussion section based on discussions with the two specialist informants according to the following criteria:

- Having high word counts within moves JD-A, B, and C;
- Reporting or discussing the methods or outcomes of the study.

These moves and steps more or less occur in the order shown in Table 4 although step JD-C12, for stating promotion of further research, does not always appear at the end of the discussion section.

**Cluster analysis by move and step.** The discussion section of *CircJ*, *Circ*, *NEJM-OA*, and *NEJM-BR* was divided into small groups of texts by move and step, using the text editor *TextWrangler*, and word frequencies of each text were calculated with *CasualConc* (Imao, 2015). A cluster analysis was performed using the 1000 most frequent words as variables (Figure 3). The 200 most frequent words and the word frequency profiles of the moves and steps are shown in Appendices A and B, respectively. As shown in Appendix B, the modal auxiliary verbs *may*, *might*, *can*, *would*, and *should* appear in the 200 most frequent words found in the discussion section.

The cluster analysis showed that most of the *Circ* texts were in one cluster, as shown in the rectangle (a) in Figure 3, with a group of the texts from *CircJ*, *NEJM-OA*, and *NEJM-BR* and the remaining texts from *Circ* were in another large cluster. The results indicated the similarity of the texts from *CircJ*, *NEJM-OA*, and *NEJM-BR* when the 1000 most frequent words were used as variables.

The results also showed similarity of the words used to introduce new information (JD-Z10) in the RAs of the two cardiovascular journals and the *NEJM* (Rectangle (b) in Figure 3). Note the cluster of steps for indicating limitation of outcomes and methods (JD-B15 and JD-25) for *CircJ* and JD-25 for *NEJM-OA*, as shown in rectangle (c) in Figure 3, suggesting that the words used to describe the limitation of the study in the *CircJ* RAs seem to be similar to those in the *NEJM-OA* RAs.

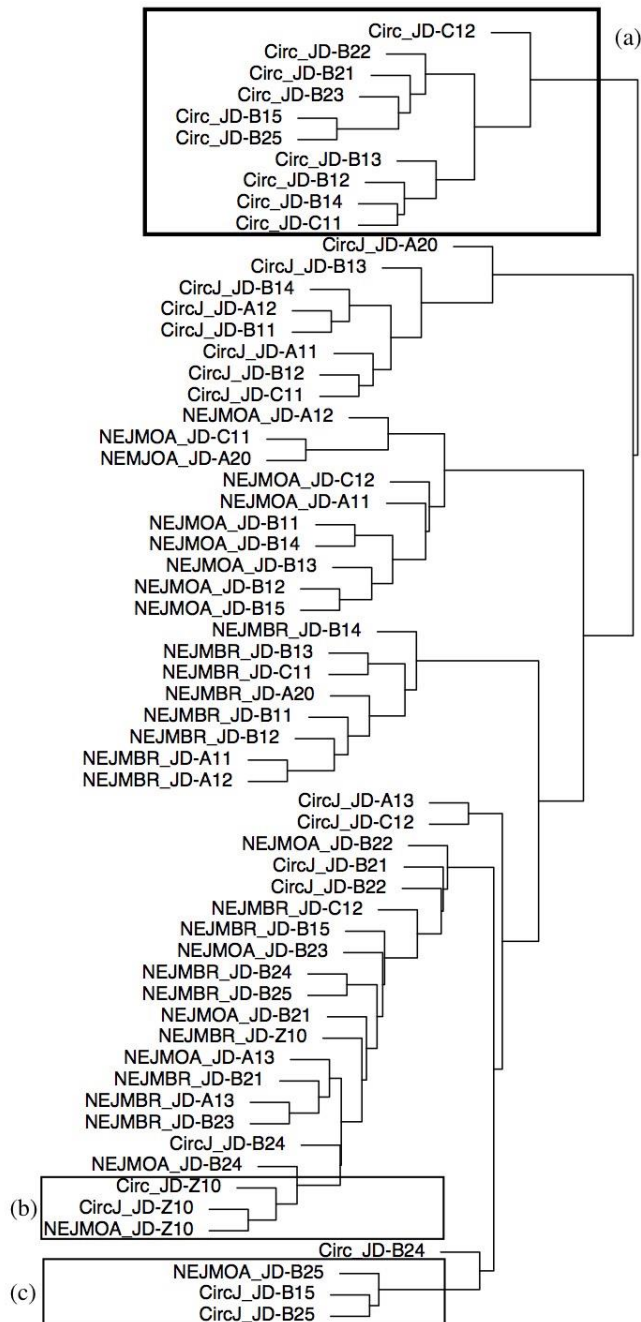


Figure 3. Cluster analysis using the most frequent 1000 words in the moves and steps of the discussion section.

**Word frequency of modal auxiliary verbs.** Table 6 shows the word frequency of the modal auxiliary verbs in the major moves and steps. The frequency was substantially low in Move JD-A20 (Highlighting overall results and their significance) and Step JD-B11 (Stating a specific outcome). In both Move JD-A20 and Step JD-B11, the authors tended to describe the study results objectively as in sentences (4) and (5), respectively.

Table 6

*Word Frequency of Modal Auxiliary Verbs in the Discussion Section Standardized per 1000 words*

Move/ Journal	Modal Auxiliary Verbs								
	may	might	would	could	can	should	cannot	must	will
JD-A20: Highlighting overall results and their significance									
<i>CircJ</i>	0.00	0.00	0.00	3.82	0.00	0.00	0.00	0.00	0.00
<i>Circ</i>	1.53	0.00	0.00	0.00	1.53	0.00	0.00	1.53	0.00
<i>NEJM-OA</i>	1.74	0.00	0.00	0.00	1.74	0.00	0.00	0.00	0.00
<i>NEJM-BR</i>	0.00	0.00	0.00	0.00	2.70	0.00	0.00	0.00	0.00
JD-B11: Stating a specific outcome									
<i>CircJ</i>	0.00	0.00	0.00	0.00	0.69	0.00	0.00	0.00	0.00
<i>Circ</i>	0.77	0.00	1.53	0.00	0.00	0.00	0.00	0.00	0.00
<i>NEJM-OA</i>	1.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>NEJM-BR</i>	2.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
JD-B12: Interpreting the outcome									
<i>CircJ</i>	11.07	10.28	2.37	1.58	4.74	0.00	0.00	0.00	0.00
<i>Circ</i>	8.62	7.29	5.97	3.32	1.33	2.65	2.65	0.00	0.00
<i>NEJM-OA</i>	16.11	2.68	1.34	2.68	2.68	4.03	2.68	1.34	0.00
<i>NEJM-BR</i>	8.67	1.73	3.47	3.47	0.00	1.73	1.73	0.00	0.00
JD-B15: Indicating limitation of outcome									
<i>CircJ</i>	5.68	3.79	0.00	3.79	1.89	3.79	1.89	0.00	1.89
<i>Circ</i>	4.53	2.27	3.40	1.13	1.13	1.13	6.80	0.00	0.00
<i>NEJM-OA</i>	10.22	1.28	1.28	5.11	1.28	1.28	0.00	1.28	0.00
<i>NEJM-BR</i>	0.00	0.00	11.11	11.11	0.00	0.00	11.11	0.00	0.00
JD-C11: Indicating research implications									
<i>CircJ</i>	4.72	3.14	0.00	0.00	4.72	6.29	0.00	1.57	0.00
<i>Circ</i>	1.22	0.00	1.22	0.00	1.22	1.22	1.22	1.22	0.00
<i>NEJM-OA</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>NEJM-BR</i>	9.01	3.00	0.00	0.00	0.00	3.00	0.00	0.00	3.00
JD-C12: Promoting further research									
<i>CircJ</i>	6.94	0.00	0.00	3.47	3.47	20.83	0.00	0.00	0.00
<i>Circ</i>	0.00	4.85	4.85	4.85	4.85	0.00	0.00	0.00	0.00
<i>NEJM-OA</i>	12.99	4.33	0.00	0.00	8.66	4.33	0.00	0.00	4.33
<i>NEJM-BR</i>	4.52	9.05	4.52	4.52	4.52	0.00	0.00	0.00	9.05

In step JD-B12 for interpreting the outcome (JD-B12), the frequency of *might* is the highest in *CircJ* and is as high as 10.28 per 1000 words (Table 6, Sentence (13)). Neither *should* nor *cannot* were found in *CircJ*. The use of *might* in *CircJ* is discussed below after sentence (17).

(13) The preoperative physiological and psychological conditions of those patients *might* be different from patients who received an implantable LVAD directly, which *might* influence QOL after LVAD implantation. [emphasis added] (*CircJ5*)

(14) An important implication is that conventional tumor nosology based on organ site (with molecular subtypes) *cannot be* entirely replaced by molecular nosology (e.g., BRAF-mutated cancers). [emphasis added] (*NEJM3*)

The authors of *CircJ* articles did not use *would* in the step for indicating limitation of outcomes (Step JD-B15), but did use *will* for this.

(15) Women were asked in middle age to recall their age at menarche, but the likely effect of reporting errors *would* be to attenuate risk estimates, ... [emphasis added] (*Circ86*)

(16) It is unclear from the present study whether these patients *will* further progress to HFrEF over a longer period of time. [emphasis added] (*CircJ6*)

In Step JD-C11, in which authors summarize their study, *might* was used only in *CircJ* among the RAs. The modal auxiliary verb *might* was used in *NEJM-BR*.

(17) Given our findings, circulating levels of the hormone that appear to be normal in relation to body-mass index and fat mass do not rule out disease-causing mutations in the gene encoding leptin and *might* obscure the correct diagnosis. [emphasis added] (*NEJM-BR10*)

The articles published in *NEJM-BR*, according to the specialist informants, often include novel findings of patients with rare cases of diseases that require, for example, DNA analyses for determination or confirmation of diagnosis. Taking this into consideration, it would be reasonable to find the use of *might* in the summary. The use of *might* in *CircJ* may represent the reluctance of the local culture to use definite expressions or may be a result of regional pedagogy.

- (19) Additional administration of IV-mexiletine *might be* a therapeutic option for the treatment of IV-amiodarone-refractory VT storms. [emphasis added] (*CircJ*47)

For promoting further research (Step JD-C12), all authors except those of the *CircJ* articles used *might* (Sentence (20)) and *will* (Sentence (21)). Also, the authors of the *CircJ* articles were found to have substantially overused *should* (Sentence (22)).

- (20) A genomewide association study *might be* able to identify genetic factors associated with this observation. [emphasis added] (*NEJM-OA*5)

- (21) Our next step *will be* to examine these proposed pathways between early-life psychosocial factors and cardiac health outcomes later in life. [emphasis added] (*CircJ*99)

- (22) In the future, intervention studies *should be* performed to investigate whether the correction of GV effectively prevents RP in patients with ACS. [emphasis added] (*CircJ*9)

## Discussion

This study investigated the features of English in medical research paper genres from the viewpoint of ESP and EIL and examined whether or not the use of language in the RAs from the Japanese cardiovascular journal *Circulation Journal* (*CircJ*) could be used as a model for Japanese scientists involved in discipline-specific RA writing. A cluster analysis of the entire texts of the RAs with the 1000 most frequent words as variables revealed that the RAs from the three journals *CircJ*, *Circ*, and *NEJM-OA* and the CRs from *NEJM-BR* form two separate categories. The presence of RAs from the Japanese cardiovascular journal *CircJ* in the corpora produced a finding consistent with the studies defining RAs and CRs as two different subgenres (Helan, 2012; Salager-Meyer, 1989).

Close examination of the discussion section showed that the moves and steps have common linguistic features such as the use of the present tense in Move JD-Z and the hint expression *consistently with* in Step JD-B14. However, cluster analysis using the 1000 most frequent words in each of the moves or steps as variables revealed that most of the *Circ* texts formed a distinct cluster, with the *CircJ*, *NEJM-OA*, and *NEJM-BR* texts and the remaining texts from *Circ* appeared in another large cluster. The cluster analysis also suggested the similarities of the words used in the steps for indicating limitation of the study in the *CircJ* and *NEJM-OA* texts. The results suggested that the words and some rhetorical features of the *CircJ* discussion texts may have greater similarities to *NEJM-OA* than to *Circ*. These findings are understandable as *NEJM-OA* is widely referred to in textbooks for medical RA writing and in teaching materials for undergraduate and graduate medical students.

The use of modal auxiliary verbs in some of the steps in the *CircJ* articles appeared to be different from that in other articles:

- The *CircJ* authors tended to overuse *should* to discuss further research. This finding was considered to be due to direct translation of Japanese phrase representing obligation *subekidearu*. The *Circ* and *NEJM* authors tend to use phrases such as

*A...study might be able to* and *Our next step will be* (Sentences (20) and (21));

- The *CircJ* authors did not use *would* or *could* to make hypothetical statements (Huckin & Olsen, 1983, p. 547);
- Among the RA authors, only the authors of *CircJ* used *might* to state the most important summary point of the study.

Such usage of modal auxiliary verbs should be of value to local instructors as examples of rhetorical devices chosen by the discourse community members of the target research paper genres.

One of the limitations of this study is the use of words as variables to quantitatively study the RAs and CRs. Also, we studied moves and steps only in the discussion section of ten articles each from *CircJ*, *Circ*, *NEJM-OA*, and *NEJM-BR*. Further studies with larger corpora examining moves and steps of other sections may reveal linguistic features that could contribute to developing a model for Japanese researchers.

The findings of this study suggested that the Japanese *Circulation Journal* articles in our corpora differed significantly enough, at least with regard to the use of words, from the articles of the widely-read US counterpart *Circulation* to recommend building mini-corpora consisting of authentic materials of a target genre as a way to become aware of rhetorical patterns and expressions frequently used by more experienced writers (Anthony, 2015; Casanave, 2003; Lee & Swales, 2006).

EIL authors, who struggle with the burden of English as a second or foreign language, produce at least 50% of the publications in most of the peer-reviewed journals of the highest quality according to Benfield and Feak (2006, p. 1728). *CircJ* is one of the best peer-reviewed English medical journals in Japan with an impact factor of 4.124 in 2015 (*Circulation Journal*, 2017). Our present study has shown how research findings are communicated in a discipline-specific journal published by a Japanese expert society in the field of medicine and revealed an example of use-related EIL communication with RAs as the medium.



## Acknowledgments

We are grateful to the anonymous *JACET Selected Papers* reviewers for their insightful comments on earlier drafts of this article. We also thank Nobuyuki Hino, Tomoji Tabata, Bor Hodošček, Hisashi Iwane, Yasuhiro Imao, Maki Miyake, Hironori Ueda, and Tomoko Wakasa for their invaluable discussions and help throughout this project.

## References

- Anthony, L. (2015). The changing role and importance of ESP in Asia. *English as a Global Language Education (EaGLE) Journal*, 1(1), 01–21.
- Asano, M. (2016a). Igaku ronbun ni okeru senmon eigo bunseki no tameno yobiteki kento (Preliminary study for analyzing linguistic features of English in medical research articles). In T. Tabata (Ed.), *The Institute of Statistical Mathematics Cooperative Research Report 364: Text Mining and Text Analysis*. (pp. 47–68). Tokyo: The Institute of Statistical Mathematics.
- Asano, M. (2016b). Nichibei no igaku ronbun ni okeru gengoteki tokucho no yobi bunseki (A preliminary analysis of linguistic features in Japanese and American medical research articles). In T. Tabata (Ed.), *Language and Culture Cooperative Research Project 2015: Osaka Studies in Corpus Linguistics 2015–2016*. (pp. 81–118). Osaka, Japan: Graduate School of Language and Culture, Osaka University.
- Benfield, H. R., & Feak, C. B. (2006). How authors can cope with the burden of English as an international language. *Chest*, 129, 1728–1730.
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Casanave, C. P. (2003). Multiple uses of applied linguistics in a multi-disciplinary graduate EAP class. *ELT Journal*, 57(1), 43–50.
- Circulation. (2017). General Preparation Instructions. Retrieved from

- <http://www.wkadcenter.com/mediakit/circulation>
- Circulation Journal. (2017). Retrieved from <http://www.j-circ.or.jp/english/cj/>
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222. doi: 10.1093/applin/24.2.197
- El Malik, T., & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *English for Academic Purposes*, 7, 87–96. doi: 10.1016/j.jeap.2008.02.007
- Helan, R. (2012). Analysis of Published Medical Case Reports: Genre-based Study. PhD Dissertation. Department of English and American Studies, Faculty of Arts, Masaryk University. (unpublished). Retrived from [https://is.muni.cz/th/18899/ff\\_d/](https://is.muni.cz/th/18899/ff_d/DISSERTATION_-_ROBERT_HELAN.pdf) DISSERTATION\_-\_ROBERT\_HELAN.pdf
- Horton, R. (1995). The rhetoric of research. *British Medical Journal*, 310, 985–987. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2549363/pdf/bmj00588-0039.pdf>.
- Huckin, T., & Olsen, L. (1983). *Technical Writing and Professional Communication for Nonnative Speakers of English*. New York: McGraw-Hill College.
- Hulstijn, J. H. (2010). Measuring second language proficiency. In E. Blom & S. Unsworth (Eds.), *Experimental Methods in Language Acquisition Research*. (pp. 185–200). Amsterdam: John Benjamins.
- Ichihara, E. (1982). *Life Science ni okeru Eigo Ronbun no Kakikata*. (How to Write Life Science Papers in English). Tokyo: Kyoritsu Shuppan.
- Imao, Y. (2015). CasualConc (Version 2.0.6) [Computer Software]. Osaka, Japan: Osaka University. Retrieved August 9, 2016 from <https://sites.google.com/site/casualconc/>
- Kobayashi, Y. (2011). *Toukeiteki Text Mining ni yoru Gakushusha Sakubun ni okeru Meta Danwa Hyoshiki no Kenkyu*. (An Investigation of Metadiscourse Markers in L2 Composition Using a Statistical Text Mining Approach). Unpublished doctoral

- dissertation. Department of Language and Culture, Osaka University.
- Kobayashi, Y., Tanaka, S., & Tomiura, Y. (2011). Random Forest wo mochiita eigo kagaku ronbun no bunrui to hyoka. (Classification and assessment of English scientific papers using Random Forests). *Joho Shori Gakkai Kenkyu Houkoku (Information Processing Society of Japan, IPSJ) SIG Technical Report*, 6, 1–8.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25, 56–75. doi: 10.1016/j.esp.2005.02.010
- Lock, S. (1977). *Thorne's Better Medical Writing*. (2nd Ed.). London: Pitman Press.
- Maci, S. M. (2010). Argumentative structure in American and Italian medical research articles. In G. Garzone & J. Archibald (Eds.): *Discourse, Identities and Roles in Specialized Communication*. (pp. 343–363). Bern, Switzerland: Peter Lang.
- Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, 2, 1–11. doi: 10.1016/j.amper.2014.12.002
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Methods, Theory and Practice*. Cambridge, England: Cambridge University Press.
- Noguchi, J. (2010). Exploiting ESP frontiers: Systemic literacy, life-long learning, ESP bilingualism. *Annual Report of JACET-SIG on ESP*, 12, 3-13.
- Noguchi, J. (2016). Revisiting the construction of knowledge in science. *Journal of Global Communication Studies*, 1, 3–17.
- Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16(2), 119–138. doi: 10.1016/S0889-4906(97)85388-4
- Quirk, R., & Stein, G. (1990). *English in Use*. London: Longman.
- Robinson, M. S., Stoller, F. L., Costanza-Robinson, M. S., & Jones, J. K. (2008). *Write Like a Chemist*. Oxford, UK: Oxford University Press.

- Salager-Meyer, F. (1989). Principal component analysis and medical English discourse: An investigation into genre analysis. *System*, 17(1), 1989.
- Smith, L. E. (1976). English as an international auxiliary language. *RELC Journal*, 7(2), 38-53.
- Swales, J. (1981). *Aspects of Article Introductions*. Birmingham, UK: The Language Studies Unit, Aston University.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, England: Cambridge University Press.
- Swales, J. M., & Feak, C. B. (2012). *Academic Writing for Graduate Students, Essential Tasks and Skills*. 3rd ed. MI, USA: University of Michigan Press.
- Tabata, T. (2016). Computing rhetorical features in texts: Dickens, 18th- and 19th-century British fiction. In T. Tabata (Ed.), *Language and Culture Cooperative Research Project 2015: Osaka Studies in Corpus Linguistics 2015–2016*. (pp. 21–53). Osaka, Japan: Graduate School of Language and Culture, Osaka University.
- Taylor, R. B. (2005). *The Clinician's Guide to Medical Writing*. New York: Springer.
- The *New England Journal of Medicine*. (2017). Retrieved from <http://www.nejm.org/page/about-nejm/history-and-mission>
- Tojo, K., Hayashi, H., & Noguchi J. (2014). Linguistic dimensions of hint expressions in science and engineering research presentations. *JACET International Convention Selected Papers, I*, 131-163.
- Umesaki, A. F. (2000). Syntactic differences in the discourse of oral and written papers. *English Corpus Studies*, 7, 39-59.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 58, 236–244.
- Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes*. (pp. 71-83). Cambridge, England: Cambridge University Press.

## Appendix A

### The Most Frequent 200 Words in the Discussion Section

1-20		21-40		41-60		61-80		81-100		101-120		121-140		141-160		161-180		181-200	
the	55.58	our	5.18	had	2.35	did	1.79	those	1.53	health	1.21	artery	1.09	significantly	0.97	even	0.82	recent	0.76
of	45.90	we	5.06	been	2.32	high	1.79	age	1.50	both	1.18	mellitus	1.09	some	0.97	factor	0.82	related	0.76
in	34.19	risk	4.53	reported	2.32	found	1.77	can	1.47	cancer	1.18	their	1.09	glucose	0.94	most	0.82	showed	0.76
and	23.80	are	4.47	coronary	2.29	other	1.77	chd	1.44	care	1.18	ischemic	1.06	lvef	0.94	rates	0.82	small	0.76
to	19.21	from	4.41	data	2.27	after	1.74	outcomes	1.41	events	1.18	statin	1.06	pause	0.94	virus	0.82	acute	0.74
with	18.77	on	4.33	more	2.27	present	1.74	analysis	1.38	first	1.18	therefore	1.06	any	0.91	bias	0.79	aortic	0.74
a	17.15	by	4.27	who	2.21	group	1.71	only	1.38	mg	1.18	addition	1.03	one	0.91	bridging	0.79	et	0.74
that	12.42	at	4.06	also	2.15	heart	1.71	patient	1.38	population	1.18	among	1.03	opcab	0.91	cabg	0.79	infection	0.74
for	11.42	an	4.00	during	2.15	significant	1.71	there	1.38	renal	1.18	cells	1.03	warfarin	0.91	cohort	0.79	pregnancy	0.74
patients	10.97	may	3.74	stroke	2.09	therapy	1.71	trial	1.38	should	1.18	could	1.03	dose	0.88	complications	0.79	receiving	0.74
was	9.77	which	3.41	although	2.06	effect	1.68	bleeding	1.32	compared	1.15	early	1.03	number	0.88	follow-up	0.79	reduced	0.74
is	9.47	associated	2.97	clinical	2.03	increased	1.65	higher	1.32	failure	1.15	important	1.03	qol	0.88	incidence	0.79	blood	0.71
not	7.56	these	2.88	findings	2.03	observed	1.65	mortality	1.32	including	1.15	increase	1.03	sle	0.88	pauses	0.79	children	0.71
study	7.15	than	2.77	no	2.03	such	1.65	survival	1.32	potential	1.15	levels	1.03	cases	0.85	psychosocial	0.79	given	0.71
be	6.62	however	2.74	but	2.00	all	1.62	women	1.29	previous	1.15	multiple	1.03	effects	0.85	randomized	0.79	sample	0.71
this	6.33	disease	2.62	use	1.97	has	1.62	cardiovascular	1.27	duration	1.12	previously	1.03	likely	0.85	suggest	0.79	analyses	0.68
were	5.74	it	2.53	treatment	1.94	association	1.59	used	1.27	lower	1.12	overall	1.00	menarche	0.85	cin	0.76	based	0.68
or	5.44	studies	2.53	because	1.91	cardiac	1.56	would	1.27	lvad	1.12	case	0.97	outcome	0.85	groups	0.76	identify	0.68
as	5.30	between	2.44	factors	1.85	might	1.56	time	1.24	major	1.12	death	0.97	they	0.85	normal	0.76	ii	0.68
have	5.30	results	2.38	rate	1.85	similar	1.53	diabetes	1.21	years	1.12	function	0.97	vt	0.85	primary	0.76	included	0.68

*Note:* The word frequencies are standardized per 1000 words.

## Appendix B

### Word Frequency Profiles of the Moves and Steps in the Discussion Section:

#### Standardized per 1000 Words

Group	the	of	in	and	to	with	a	that	for	patients	was	is	not	study	be	this	were	or	as	have
Circ_JD-A20	38.34	55.21	35.28	26.07	4.60	35.28	18.40	10.74	13.80	9.20	7.67	3.07	4.60	3.07	4.60	6.13	4.60	4.60	3.07	6.13
Circ_JD-B11	59.72	41.35	31.39	32.92	13.78	22.97	13.78	9.19	22.21	4.59	13.78	1.53	9.19	3.06	0.77	3.06	10.72	3.06	5.36	0.77
Circ_JD-B12	49.07	43.77	29.18	22.55	15.92	15.92	9.28	21.22	12.60	10.61	3.32	7.29	5.31	3.98	12.60	11.27	2.65	2.65	4.64	8.62
Circ_JD-B15	37.37	32.84	31.71	27.18	16.99	12.46	13.59	12.46	13.59	5.66	6.80	7.93	23.78	6.80	7.93	11.33	9.06	6.80	2.27	7.93
Circ_JD-C11	41.51	45.18	40.29	26.86	19.54	37.85	20.76	7.33	8.55	4.88	12.21	8.55	6.11	2.44	6.11	8.55	2.44	4.88	4.88	3.66
Circ_JD-C12	33.98	43.69	38.83	33.98	19.42	29.13	9.71	4.85	0.00	19.42	0.00	0.00	0.00	4.85	4.85	0.00	0.00	0.00	9.71	4.85
Circ_JD-A20	53.44	61.07	26.72	19.08	15.27	22.90	11.45	19.08	11.45	22.90	15.27	15.27	3.82	22.90	3.82	7.63	7.63	11.45	7.63	0.00
Circ_JD-B11	68.78	50.21	55.02	20.63	13.07	21.32	14.44	5.50	7.57	21.32	24.07	1.38	9.63	13.76	1.38	3.44	11.69	8.25	6.88	1.38
Circ_JD-B12	61.66	40.32	26.88	17.39	22.13	17.39	15.02	15.02	11.86	20.55	10.28	8.70	9.49	7.11	15.02	5.53	2.37	7.11	4.74	3.95
Circ_JD-B15	66.29	39.77	28.41	20.83	20.83	5.68	22.73	13.26	13.26	26.52	13.26	9.47	13.26	17.05	11.36	5.68	11.36	5.68	5.68	5.68
Circ_JD-C11	56.60	39.31	31.45	28.30	17.30	17.30	20.44	9.43	18.87	26.73	3.14	12.58	4.72	3.14	18.87	6.29	7.86	7.86	11.01	6.29
Circ_JD-C12	59.03	55.56	24.31	20.83	45.14	24.31	10.42	3.47	20.83	24.31	3.47	3.47	0.00	10.42	27.78	6.94	3.47	3.47	0.00	3.47
NEJMBR_JD-A20	56.60	45.82	35.04	18.87	5.39	21.56	37.74	21.56	5.39	5.39	13.48	13.48	0.00	0.00	0.00	2.70	0.00	2.70	0.00	2.70
NEJMBR_JD-B11	62.91	48.57	35.32	30.91	9.93	9.93	7.73	4.42	5.52	5.52	11.04	3.31	4.42	1.10	1.10	3.31	9.93	8.83	2.21	2.21
NEJMBR_JD-B12	65.86	38.13	29.46	15.60	32.93	8.67	13.86	15.60	12.13	3.47	5.20	17.33	8.67	1.73	19.06	10.40	5.20	1.73	6.93	5.20
NEJMBR_JD-B15	133.33	66.67	22.22	0.00	33.33	0.00	11.11	11.11	11.11	11.11	0.00	11.11	11.11	0.00	11.11	33.33	11.11	0.00	0.00	0.00
NEJMBR_JD-C11	57.06	57.06	39.04	24.02	21.02	9.01	27.03	24.02	6.01	3.00	9.01	0.00	3.00	3.00	6.01	15.02	0.00	3.00	3.00	6.01
NEJMBR_JD-C12	36.20	36.20	22.62	22.62	27.15	4.52	18.10	9.05	22.62	4.52	0.00	18.10	0.00	0.00	0.00	9.05	13.57	0.00	9.05	4.52
NEMJOA_JD-A20	55.56	60.76	48.61	19.10	19.10	24.31	27.78	17.36	10.42	8.68	13.89	1.74	8.68	10.42	3.47	10.42	1.74	5.21	1.74	0.00
NEJMOA_JD-B11	78.48	67.26	37.00	15.70	24.66	19.06	16.82	12.33	8.97	12.33	21.30	2.24	7.85	3.36	1.12	2.24	11.21	6.73	1.12	1.12
NEJMOA_JD-B12	71.14	61.74	37.58	21.48	22.82	14.77	13.42	10.74	6.71	10.74	4.03	12.08	6.71	6.71	17.45	8.05	4.03	5.37	8.05	12.08
NEJMOA_JD-B15	68.97	54.92	26.82	12.77	24.27	12.77	14.05	14.05	7.66	14.05	12.77	10.22	17.88	5.11	12.77	10.22	11.49	7.66	3.83	7.66
NEJMOA_JD-C11	43.06	47.85	43.06	14.35	16.75	35.89	14.35	11.96	14.35	14.35	9.57	2.39	11.96	2.39	0.00	0.00	7.18	14.35	0.00	0.00
NEJMOA_JD-C12	38.96	43.29	17.32	30.30	34.63	8.66	21.65	12.99	4.33	0.00	4.33	0.00	0.00	4.33	30.30	8.66	0.00	4.33	4.33	0.00



**Assessing the Effectiveness of the COLT Scheme as a Reflection Tool  
for High School Teachers of English**

Aiko Sano

Hokkaido Bunkyo University

Noriaki Katagiri

Hokkaido University of Education

Yuko Sakai

Hokkaido Sapporo Intercultural and Technological High School<sup>i</sup>

Akinobu Shimura

Hokkaido University of Education

**Abstract**

One of the major challenges English teaching in Japan faces is how to make language classrooms more communicative. Teacher development plays an important role in improving English education, and helping teachers reflect on their practices is one of the most powerful ways through which to achieve this. This study reports on how making use of Communicative Orientation of Language Teaching observation scheme (COLT; Spada & Frölich, 1995) enables three English teachers at Japanese high schools to reflect on their teaching practices through an objective lens, alongside a qualitative analysis of their teacher beliefs. Their teacher beliefs and self-reflections were examined through semi-structured interviews, and analysed through a coding procedure inspired by Grounded Theory Approach (Saiki- Craighill, 2008). Despite the limitations of small sample size, this study sheds light on how analytical tools such as COLT for teachers' self-reflections can benefit teacher development.

**Key words:** COLT, teacher development, language teaching, self-reflection

<sup>i</sup>Tokai University

## **Introduction**

English proficiency constitutes one of the core skills needed by students in order to flourish in a rapidly globalising era. At the same time, it seems widely accepted that English education in Japan has seen a series of failures, especially in terms of speaking and writing. While being aware of the danger of commodification of English skills and its teaching, teachers of English are constantly searching for better teaching methods.

In Japan, in order to become a teacher, one needs to attend a university that offers a teaching certificate degree. Teaching certificates are granted upon completion of all the required teaching courses, as well as completing a teaching practicum, and taking care of the elderly. The minimal requirements for this course includes twenty credits in subjects related to English, such as linguistics, literature and applied linguistics. Thus, it is possible for a student with only four credits in English teaching to become a qualified English teacher. To make the situation even worse, the high school practicum required by the Ministry of Education lasts only for two weeks. This insufficient training has been identified as one of the causes of unpreparedness of new teachers (Ito, 2016).

It then becomes crucial to provide in-service teachers with opportunities for professional development. In addition to such training courses offered by local boards of education, it has now become mandatory in Japan for teachers to take courses every 10 years in order to keep their certificate valid. Teachers thus, at least in theory, have opportunities to update their knowledge and skills in order to adapt to the rapidly changing needs of society. However, teacher development is not an easy task. Guskey (2002), reviewing studies on professional development, summarises: “Despite the general acceptance of professional development as essential to improvement in education, reviews of professional development research consistently point out the ineffectiveness of most programs (pp.381-382)”.

Wallace (1991), in consonance with Guskey, emphasises the importance of the role played by teacher education, and identifies three major models of it: the craft model, the



applied science model, and the reflective model. The latter, he argues, is best suited to language teacher training. Larrivee (2000), synthesising previous studies, likewise illustrates the importance of “critical analysis and reformulation” of personal experiences in reflective practices (p.296). In the same vein, Crandall (2000), while acknowledging the importance of all three types of teacher development identified by Wallace, points out that the lack of due attention to the role reflection model plays, and states: “Long ignored, teacher inquiry and reflection are now viewed as important to the development of language teaching theory and appropriate language teacher education (p.40).”

As there appears to be a general consensus that teacher reflection plays an important role in teacher development, the question of how such reflective practices should be conducted arises. One approach is through action research where in-service-teachers become self-aware of challenges they face in their classes and try to resolve the issue by themselves. While such an approach has seen much success, it requires use of excessive time and effort which most teachers simply cannot spare. Thus, there is a need for a framework through which in-service-teachers can reflect on their teaching practices in an objective manner. This framework needs to link their practice to theory, as well as to deepen their understandings of their own teacher beliefs. This study attempts to make use of a theoretically driven tool in order to analyse teaching practices so that the teachers themselves can better reflect upon their teaching. For that purpose, we limited our focus on how communicative the classes are, which reflects the main challenge that English teachers in Japan face, as a way to gauge improvements in English teaching.

### **Literature Review**

As summarised in the introduction, there is a call for more objective ways to help teachers reflect upon their teaching practices. In the present study, our focus is on helping teachers of English make their classes more communicative, and therefore, we will begin with

reviewing measures of communicativeness of language classes.

### **Measures of Communicativeness of Language Classes**

In order for researchers to investigate the degree of communicative orientations of language classes, several indices have been devised; Flanders' Interaction Analysis Categories (FIAC) proposed by Flanders (1970), Interactional Analysis in the Foreign Language Classroom by Wragg (1970), and Foreign language interaction analysis (Flint) by Markowitz (1971) to name a few. All of these measures attempt to ensure that observation of language classes would not be based on any subjective and impressionistic views held by researchers but would be objective and valid. Among these tools, the most widely applied to research is the Communicative Orientation of Language Teaching observation scheme (COLT), first proposed by Frölich, Spada, & Allen (1985), and revised in Spada and Frölich (1995). This is a tool comprised of two parts, Part A and B. In Part A, observers name each activity taking place in a class, and record the length of time devoted to that activity. The observers then describe what kind of activity it is against the checklist of criteria that are theoretically driven to reflect communicative orientations of language teaching. In Part B, researchers record and transcribe students-teacher interactions in class and classify them into categories.

### **Adaptability of COLT as an Analytical Tool in a Japanese Context**

COLT has been applied widely to classroom observation and analysis of the English teaching in Japan. Ishizuka et al. (2005) was one of the first to apply COLT scheme to the Japanese context, using Part A to distinguish four English-native instructors' teaching styles. They found that it was indeed effective in describing the differentiated lesson objectives of the four classes designated as "Oral English," "Reading," and "The Art of Writing" and distinguishing their in-class activities. Following Ishizuka et al. (2005), Kawai, Sakai, Yokoyama, Ishizuka, Aoki (2007) and Aoki, Ishizuka, Yokoyama, Sakai, Kawai (2008) further investigated the effectiveness of the COLT scheme to analyse English programmes at a Japanese university using Part B. Kawai et al. (2007) concluded that Part B provides more

precise information regarding communicative orientation of the class than Part A. Combined with sequential analysis, the scheme provided a very thorough picture of what was happening in the classes, although Part B analysis is admittedly a labour intensive one. Aoki et al. (2008) revealed that Part B was sensitive enough to detect the teacher variations across instructors teaching the same English language programs. Together, these studies illustrate that COLT can indeed be an effective tool in analysing the degree of communicative orientation of English teaching in a Japanese context.

### **Learners' and Teachers' Perceptions and the Communicative Orientations of Classes**

Classroom observation using COLT has also been tested to see if it reflects students' perceptions of the classes. Yokoyama et al. (2012) reported that student not only preferred classes with more group work but also perceived these classes to be more effective. In addition, although students showed general preference for having some content control, having too much of it was not necessarily perceived to be the ideal.

In order to further investigate the link between students' motivation for language learning and the degree of communicative orientation of English classes, Shimura, Sano, Sakai, Yokoyama and Kawai (2015) combined Motivation Orientation of Language Teaching (MOLT) scheme (Guilioteaux & Dörnyei, 2008) perspectives to the COLT scheme to assess English classes at a Japanese university. This study illustrated how classes given low scores in COLT were also given low scores in MOLT. They also reported some items scoring high in COLT that were rated low in MOLT, indicating that COLT and MOLT are capturing different aspects of language teaching. In investigating the teacher belief about communicative teaching and their actual teaching practices, Shimura (2010) linked the teacher belief portrayed through interviews with classroom observation utilising both parts of COLT.

The studies reviewed have all made use of COLT as an analytical tool of classroom observation, and suggest a link between the communicative orientations of the classes as depicted in COLT and the preferences, perspectives, or the beliefs of learners and teachers.

## **Technological Adaptations of COLT Scheme and Its Applications for Research**

Ishizuka and Yorozyu (2014) created an automatic classroom analysis tabulating system called VOD COLT, which uses COLT categories. Users upload classroom video clips and analyse them using COLT Part A through a web-based interface. This system enables the classroom researchers to both observe the video and tabulate their analysis without changing tabs on the computer. Katagiri and Kawai (2015) adapted the COLT Part B scheme for computational tabulation, and demonstrated the effectiveness of quantifying the conventional literal coding on the spread sheets for advanced computational analyses.

Ishizuka and Yorozyu (2016) further developed their VOD COLT into a multi-modal on-line system for classroom analysis, and named it CollaVOD. This system enables on-line classroom video analysis and is equipped with four features; 1) tele-conferencing, 2) providing video lectures to enable distant learning, 3) task-based learning with uploaded materials on the platform, and 4) tabulation tables for COLT Part A analysis. The final feature enables multiple users, i.e., researchers and teacher-trainers/trainees as well as the teachers who share videos of their classes, to share the same on-line video material and contribute to the class analysis on COLT Part A using the same internet platform. Ishizuka and Yorozyu (2016) demonstrated the usefulness of using the CollaVOD for classroom observation.

This utilisation of technology was a major driving force for applying the COLT scheme to various contexts. It can be applied not only as an analytical tool to investigate the communicative orientations of classes for research purposes, but can now be further applied as a teaching tool in applied linguistic classes at universities and graduate schools, and furthermore as a tool for teacher development.

As presented in the introduction, there is a call for reflective practices in teacher development. Wallace (1991) emphasises the importance of structured framework for reflection in order for such reflective practices to be effective. The COLT scheme (Spada &

Frölich, 1995) appears to be the best framework for the purpose of this study as it has been widely implemented in classroom observations in English teaching in Japanese contexts. It can also be easily applied to research thanks to the technological advancement made with CollaVOD.

### Research Questions

The demonstrated effectiveness of the COLT scheme as an analytical tool for English classes, together with various technological advancements, has opened up the way for its use as a reflection tool in teacher training. The research questions that guided the present study are:

1. How effective is the COLT scheme as a reflection tool for high school teachers of English?
2. How do teachers vary in the ways they reflect on their teaching using the COLT scheme?

### Method

#### Participants

Three in-service non-native English teachers at public high schools in Japan were recruited for the study. In order to have a balanced view of the diverse aspects of teaching in the area, this included two experienced teachers, and one novice teacher from three different schools in one city of Hokkaido. Table 1 shows the profiles of the participants. Teacher 1 was an experienced female teacher teaching at an *advanced* school, where the students are university entrance-examination oriented and highly-motivated to study English. Teacher 2 was an experienced male teacher teaching an *international course*, in which students are motivated to learn English language as a means of communication. Teacher 3 was a novice

male teacher in his third year of teaching at a *regular course*, where students range from those highly motivated to unmotivated. We observed two classes per teacher with the exception of Teacher 2, who invited us to a third class after observing the second class. We decided to include the third class observation into our data as well, since this study does not involve any average figures that could be affected by differing numbers of observed classes. In our research design, which will be described in more detail in the following sections, adding one more observation simply means more materials for the teacher to reflect upon, and does not distort the results.

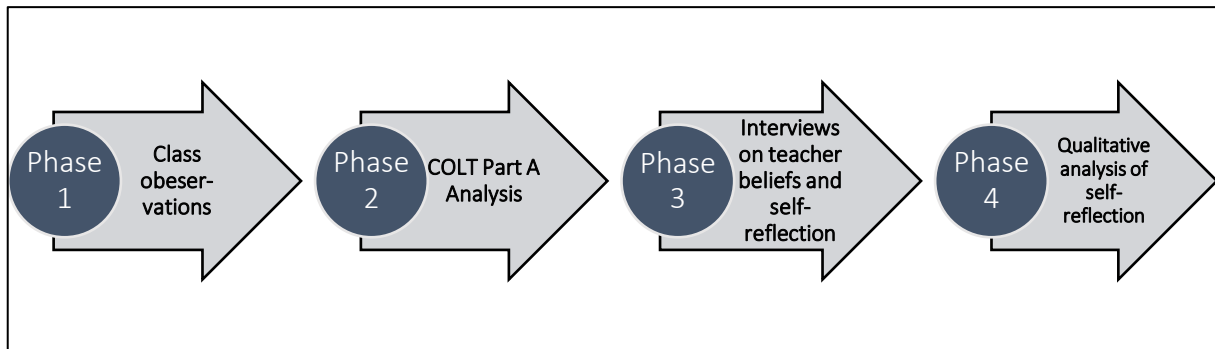
Table 1  
*Participants' Profiles*

ID	Sex	Years of teaching experience	Type of high school for the class observation	Number of classes observed
1	Female	15	Advanced	2
2	Male	18	International Course	3
3	Male	3	Regular Course	2

### **The Overall Construct of the Present Study**

The present study had four phases as depicted in Figure 1. Firstly, two to three classes of each participant were observed and recorded on video. The recorded classes were then analysed in the second phase through COLT Part A to investigate how communicatively oriented they were, using the CollaVOD system. In the third phase, the participants were interviewed, firstly about their teacher beliefs of language teaching. They were then asked to reflect on their performance while viewing the results of COLT analysis of their teaching. The recorded and transcribed interviews were analysed in the last phase, and then qualitatively analysed following the procedures in Grounded Theory Approach (Saiki-Craighill, 2008). The Grounded Theory Approach (GTA) is a data-driven approach to discover theory emerging from data, and hence repeated data collection and data analysis is required until theoretical saturation is achieved. In this study however, GTA was employed purely as an analytic

method used to analyse the interview data to avoid subjective interpretation of the data by researchers. Each of these phases will be described in detail in the following sections.



*Figure 1.* The overall construct of the study.

**Phase 1: Observing and video-recording the classes.** The three participants were specifically recruited to represent both the diversity in education within the same city of Hokkaido, as well as diversity of experience. We visited the participants' schools in the first semester of 2015 and the classes were recorded with the consent of the teachers and students. Two video cameras were set up at the back corners of the classroom.

**Phase 2: Analysing the recorded classes through COLT Part A.** COLT Part A, as proposed by Spada and Frölich (1995), was designed to observe the classes under the following five categories: Participant Organisation, Content, Content Control, Student Modality, and Materials.

CollaVOD is a multi-modal on-line platform developed by Ishizuka and Yorozyu (2016) on which its users can store lesson videos with audio, and analyse the class using COLT Part A. Since it has a time slider that is connected to the video, it automatically counts the time allocated for each activity and calculates automatically to devise pie charts instantly.

**Phase 3: Semi-structured interview on teacher beliefs and self-reflections.** In the third phase of the study, semi-structured interviews were conducted. The first half of the interview centred on the teacher beliefs of the three participating teachers. The questions were

asked in relation to the categories of COLT part A. In the second half of each interview, the teachers were presented with and asked to reflect on the results of COLT part A analysis of their classes.

The interview sessions were conducted in the participants' first language (Japanese) in the hope that this L1 use may help them reflect on their lessons with more honesty. In the sections that follow, all the interview data has been translated into English. The interviews lasted from around one hour to ninety minutes for each teacher. All the interviews were video recorded and transcribed with the participants' consent.

#### **Phase 4: Analysis of self-reflection through Grounded Theory-inspired coding.**

The transcribed interviews were then analysed through coding schemes inspired by the procedures of Grounded Theory Approach (Saiki- Craighill, 2008). In this approach, all the transcribed data is first segmented according to theme. Each segment is then qualitatively analysed to extract the *properties* and *dimensions*. *Properties* and *dimensions* constitute the founding stones of the theory derived through this approach, with *property* indicating the viewpoint of the researcher and *dimension* the interpretation of the data through the perspective described in the properties. *Properties* and *dimensions* are then condensed into more abstract *labels*, and finally combined together to form *categories* which are used for building a theory. We applied this approach in order to systematically extract concepts from the vast amount of data accumulated in the interviews. A chart was then created in order to make comparisons between the three teachers easier. In order to ensure the reliability of the present study, it was made sure that three of the four researchers jointly analysed all the segments together, referring frequently back to the actual video-recordings whenever clarification was required.

All of the three participating teachers read and signed consent forms to allow video recording of their classes and interviews. All the students in class were also informed of the study purposes and gave their consent to have their classes video recorded.



## Results

### The Individual Differences in the Teaching Practices across the Participants

**Participant organisation.** Cross-tabulation by Fisher's exact test of the number of minutes spent for each categories of Participant Organisation in COLT A detected differences at a statistically significant level ( $p = .00$ ). Table 2 presents the number of seconds each teacher spent doing the whole class activities, group activities, or individual work. The results of the post-hoc residual analysis are shown where ▲ indicates the results of the residual analyses being larger than the rest at a statistically significant level, and ▽ indicating it to be smaller.

Table 2  
*Variations in Participant Organisation Among the Three Teachers*

Teacher	Class	Class	Group	Individual
T1	A	1451	510 **▽	983 **▲
	B	1287 **▽	1017 **▽	927 **▲
T2	C	1283 **▽	1166 **▲	569 **▲
	D	936 **▽	2132 **▲	0 **▽
	E	1727 * ▲	1561 **▲	0 **▽
T3	F	1995 **▲	564 **▽	442
	G	2284 **▲	394 **▽	316 **▽

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

The figures in Table 2 indicate that Teacher 1 spent less time on group work and more time on individual work compared to the other two teachers. Teacher 2 spent much less time teaching the entire class, and much more time was allowed for group work. In the classes conducted by Teacher 2, there was very little time allotted for individual work. Teacher 3, the novice teacher, spent much more time teaching to the entire class, and much less time for activities in the form of group work.

**Content.** Cross-tabulation by Fisher's exact test of the number of minutes spent for each categories of Content in COLT A detected differences at a statistically significant level ( $p$

= .00). Table 3 presents the number of seconds spent by each teacher on activities related to class management, teaching of language, and the meaning. It also shows the results of the post-hoc residual analysis. COLT part A was also effective in depicting the differences amongst the teachers in terms of content of the classes. As can be seen in Table 3, Teacher 1 used most of the class explaining about language and not so much on the meaning-based contents. A similar pattern where more emphasis is placed on language teaching over meaning-based contents can be seen in Teacher 3, while Teacher 2 shows the opposite trend of teaching very little about language.

Table 3  
*Variations in Content Among the Three Teachers*

Teacher	Class	Management		Language		Meaning	
T1	A	81	**▽	1920	**▲	943	**▽
	B	397	**▲	1668	**▲	1168	**▽
T2	C	271	**▽	101	**▽	2656	**▲
	D	434	**▲	0	**▽	2634	**▲
	E	393	**▲	0	**▽	2895	**▲
T3	F	410	**▲	1324	**▲	1267	**▽
	G	245	**▽	1364	**▲	1385	**▽

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

**Content control.** Cross-tabulation by Fisher's exact test of the number of minutes spent for each categories of Content Control in COLT A detected differences at a statistically significant level ( $p = .00$ ). Table 4 presents the number of seconds each teacher spent on teacher and text controlled activities, as well as the number of seconds spent on activities controlled by students, teacher and the text. The results of the post-hoc residual analysis are also shown. As can be seen in Table 4, Teacher 1 and Teacher 3 were found to present a similar pattern in terms of content control as well. Both teachers spent more time being controlled by teacher and the text while Teacher 2 spent more time controlled by the students.

Table 4

*Variations in Content Control Among the Three Teachers*

Teacher	Class	Teacher / Text		Teacher / Text / Student	
T1	A	2629	**▲	315	**▽
	B	3231	**▲	0	**▽
T2	C	1342	**▽	1686	**▲
	D	862	**▽	2206	**▲
	E	393	**▽	2895	**▲
T3	F	1829	**▲	1172	**▽
	G	1871	**▲	1123	**▽

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

**Student modality.** Cross-tabulation by Fisher's exact test of the number of minutes spent on each category of Student Modality in COLT A detected differences at a statistically significant level ( $p = .00$ ). Table 5 presents the number of seconds spent on activities the students were engaged in, sorted according to the main types of the modality used. The results of the post-hoc residual analysis are also shown. In terms of student modalities, students in Teacher 1's classes spent more time reading and writing compared to students in the classes of the other two teachers, as can be seen in Table 5. Students in Teacher 3's classes were not given any opportunities to read in those particular classes, and spent far more time listening to

Table 5

*Variations in Student Modality Among the Three Teachers*

Teacher	Class	Listening		Speaking		Reading		Writing	
T1	A	1438	**▽	380	**▽	455	**▲	674	**▲
	B	1339	**▽	355	**▽	738	**▲	803	**▲
T2	C	1533	**▽	584	**▲	335	**▽	539	**▲
	D	1692	**▽	756	**▲	623	**▲	0	**▽
	E	2262	**▲	535		653	**▲	0	**▽
T3	F	2218	**▲	342	**▽	0	**▽	442	**▲
	G	1871	**▲	463	*▲	0	**▽	316	

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

the teacher. Also evident is that there are some variations shown by each individual teacher depending on the class taught.

**Types of materials.** Cross-tabulation by Fisher's exact test of the number of minutes spent for each category of Materials in COLT A detected differences at a statistically significant level ( $p = .00$ ). Table 6 presents the number of seconds each teacher spent on activities using various types of materials. Materials used included minimal texts, extended texts, audio materials, visual materials or none. The results of the post-hoc residual analysis are also shown. The teachers present different preferences for the types of materials they utilise in class. As can be seen in Table 6, only Teacher 1 made use of audio materials. Both Teacher 1 and 3 use more minimal text than extended text, while Teacher 2 was the opposite and made use of extended texts. Furthermore, in two of Teacher 2's classes, the time students spent without utilising materials is longer when compared to other classes, with the exception of Class G taught by Teacher 3.

Table 6  
*Variations in the Types of Materials Among the Three Teachers*

Teacher	Class	Minimal Text		Extended Text		Audio		Visual		None	
T1	A	1689	**▲	272	**▽	292	**▲	614	**▲	77	**▽
	B	1109	*▲	1248	**▲	222	**▲	402	**▽	282	**▽
T2	C	0	**▽	1047	**▲	0	**▽	0	**▽	1981	**▲
	D	203	**▽	1868	**▲	0	**▽	0	**▽	997	**▲
	E	0	**▽	1477	**▲	0	**▽	1450	**▲	361	**▽
T3	F	2087	**▲	0	**▽	0	**▽	87	**▽	827	**▲
	G	1926	**▲	0	**▽	0	**▽	656	**▲	412	**▽

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

**Sources of materials.** Cross-tabulation by Fisher's exact test of the number of minutes spent for each category of Sources of Materials in COLT A detected differences at a statistically significant level ( $p = .00$ ). Table 7 presents the number of seconds each teacher

spent on activities using materials of various sources. These sources included teaching materials devised for non-native speakers of English, authentic materials for native speakers of English, materials devised by the students themselves, or none of the above. The results of the post-hoc residual analysis are also shown. The three teachers in the study also showed different preferences in terms of the sources of materials used in classes as indicated in Table 7. Only Teacher 1 made use of authentic materials, and even then it was only for one of her classes (Class A). In Class B, she made use of student-produced materials, which was also a significant trend with regards to Teacher 3.

Table 7  
*Variations in the Sources of Materials Among the Three Teachers*

Teacher	Class	L2-NNS		L2-NS		Student-made		None	
T1	A	2309	**▲	241	**▲	146	**▽	248	**▽
	B	2296	**▲	0	**▽	564	**▲	371	**▽
T2	C	830	**▽	0	**▽	0	**▽	2198	**▲
	D	2071	**▲	0	**▽	0	**▽	997	
	E	1593	**▽	0	**▽	0	**▽	1695	**▲
T3	F	1828	*▲	0	**▽	345	**▲	827	**▽
	G	1816		0	**▽	701	**▲	477	**▽

*Note.* The figures in the table indicate the number of seconds.

▲: the figure is larger than the rest at a statistically significant level.

▽: the figure is smaller than the rest at a statistically significant level.

\*  $p < .05$ , \*\*  $p < .01$ .

In this section, we have presented how COLT analysis is effective in illustrating the differences between the teaching practices of the three teachers, and sometimes the differences in teaching practices displayed by the same teacher depending on the class taught. This finding is in line with previous studies, which have utilised the COLT scheme as an analytical tool for various research purposes. In this study however, data obtained from the COLT analysis was used in order to allow teachers to self-reflect on their teaching practices. The next section will compare the teacher beliefs held by each participant prior to the study, to their reactions to the objectively analysed data obtained through COLT.

## Teacher Beliefs of the Three Teachers

Table 8 summarises the properties and dimensions of the analysis through Grounded Theory Approach on the interviews related to the teacher beliefs regarding teaching English in relation to COLT categories. This analysis yielded 7 labels and 16 properties altogether, with 23 dimensions for teacher 1, 22 dimensions for Teacher 2, and 17 dimensions for Teacher 3.

Table 8  
*Properties and Dimensions Extracted From the Interviews About Teacher Beliefs of the Three Teachers*

Categories	Properties	Dimensions		
		Teacher 1	Teacher2	Teacher 3
The most important in teaching English	The main actors of the class	<ul style="list-style-type: none"> <li>• should be students</li> <li>• teacher-centred classes are boring</li> <li>• students-centred classes are effective</li> </ul>	<ul style="list-style-type: none"> <li>• should be students</li> <li>• interactions between the teacher and the students are important too</li> </ul>	<ul style="list-style-type: none"> <li>• the teacher</li> </ul>
	The ideals	<ul style="list-style-type: none"> <li>• actively vocal students</li> <li>• highly motivated students</li> </ul>	<ul style="list-style-type: none"> <li>• classes that flows in an unexpected way</li> </ul>	<ul style="list-style-type: none"> <li>• classes with a flow and a good tempo</li> </ul>
	The important features	<ul style="list-style-type: none"> <li>• various activities</li> <li>• active communication with the students</li> <li>• motivational strategies to increase students' willingness to communicate</li> </ul>	<ul style="list-style-type: none"> <li>• no extended silence</li> <li>• opportunities for students to practise English</li> <li>• active interactions between the teacher and the students</li> </ul>	<ul style="list-style-type: none"> <li>• activities aimed at learning grammar implicitly</li> </ul>
Participant Organisation	Overall views on Participant Organisation	<ul style="list-style-type: none"> <li>• pair and group activities are important</li> <li>• little individual work preferred</li> </ul>	<ul style="list-style-type: none"> <li>• the balance between individual work and the class work is important</li> </ul>	<ul style="list-style-type: none"> <li>• at least one group work a class is necessary</li> <li>• tasks for meaning negotiations</li> </ul>
	Views about teacher talks	<ul style="list-style-type: none"> <li>• effective when students are unmotivated</li> </ul>	<ul style="list-style-type: none"> <li>• should be kept little</li> <li>• sometimes necessary</li> </ul>	<ul style="list-style-type: none"> <li>• should be kept short</li> <li>• ineffective</li> </ul>
Participant Organisation	The purposes of pair works	—	<ul style="list-style-type: none"> <li>• less mental pressures to the students than with bigger groups</li> </ul>	—
	The purposes of group works	—	<ul style="list-style-type: none"> <li>• less mental pressures to the students than making class presentations</li> </ul>	<ul style="list-style-type: none"> <li>• increases students' motivations</li> <li>• enhances collaborative learnings</li> </ul>

**Table 8**  
*Continued*

Categories	Properties	Dimensions		
		Teacher 1	Teacher2	Teacher 3
Content	Views of meaning related teaching	• more important than language teaching	—	—
	Views of language related teaching	• teaching grammar out of context are meaningless • focused instruction on forms at the end of each reading materials	• accuracy in students' speeches little prioritised • grammar teaching less prioritised	• should be intentionally blended into content comprehension related tasks
Content Control (CC)	Views on students' CC	• difficult • only possible with students of high proficiencies	• students' ideas and thoughts are important • problem-solving type tasks are important	• important • enhances students' motivations
	How much CC given to students	—	• want to increase	• not as much as it should be
Student Modality	Overall views on Student Modality	• activities related to output are important • four skills need to be well balanced in classes	• speaking is the most important	• not pay much attention to the balance of the four skills
	Views on writing	• effective as comprehension check	• not so important	• not much writing components in class
Materials	Views on the use of materials other than the textbook	• use of various materials are important	• use of supplementary materials are effective to help students understand the text • internet articles related to text is effective	• infrequent use of supplementary materials other than the textbook
	Uses on the use of visual materials	• video clips of You Tubes and related work sheets • visual materials are effective to help students understand the content of texts	• hard to find appropriate materials	• use pictures and video clips related to the class topic
Sources of Materials	Preferences in the sources of materials	• preference in the use of authentic materials	• preference in the use of authentic materials	—

### Teacher Reflections on the COLT A analysis

In this section, we will closely investigate the teachers' self-reflection upon seeing the results of COLT analysis in relation to their teacher beliefs presented in Table 8.

**Participant organisation.** Teacher 1, as presented in Table 8, stated a strong preference for making use of group work in her classes, based on her belief in student-centred

classes. However, when she was presented with the result of the COLT A analysis of her class which indicated that in the case of Class A, only 17% of her class was devoted for group work activities as shown in Table 2, it made her realise that the reality did not reflect her ideals. She said, “It might be better to have more group work in my class. I thought I had more time for it.”

Teacher 2 came to a similar unexpected realisation of his own teaching style. In the interview he expressed his belief in student-centred classes. He also stated that having a variety of participant organisational styles from pair work to group work, and to whole class was important when planning a class. When he saw the result of COLT analysis of his classes, which revealed that he was teaching to the whole class for 31% to 53% of the time as can be seen in Table 2, he came to a realisation that he was talking far more than he had intended. He said,

Unexpectedly, I was spending much time for this T/SC. To come to think about it, yes, I was speaking a lot, but it did not occur to me that way before I was presented with these results. It is really surprising... My first reaction to this is like, ‘Oops, I talked too much’. It really makes me think that way... Whenever my class isn’t going well, it is when I am doing all the talking. It has been pointed out that I try too much to make my students understand what I am talking about; I try to repeat it in many different ways and end up speaking too much, and looking at this result, it really tells that I am speaking far more than I should be. (Teacher 2)

The novice teacher on the other hand, somewhat defended the small proportion of group work in his class. As shown in Table 8, he stated that he believed group work motivated students, and that time spent with the teacher lecturing to the whole class should be kept to a minimum because it was not effective in language teaching. In reality, as shown in Table 2,



66 % of class F and 76% of Class G was spent by him talking to the class.

However, he did not find that to be too problematic. He stated that he aimed to increase the amount of group work in his classes as his students gained more proficiency in English in the higher grades. For the time being he was satisfied with what he was doing in his classes, while admitting that his classes were probably too teacher-centred.

**Content.** In terms of content, Teacher 1 stated that the COLT analysis did not come as a surprise, although it seems to somewhat contradict her beliefs on form teaching being meaningless, as presented in Table 8. She spent 65% of Class A on Language as indicated in Table 3, but defended herself by stating that the focus of that particular lesson happened to be on vocabulary.

Teacher 2 likewise found few problems with his classes in terms of Content, and for Teacher 3, the COLT result actually gave him a sense of self-assurance. He stated:

It is a bit of a surprise that I spent this much of my teaching on Meaning. I guess I can claim my classes were meaning-focused to some extent since I spent half my classes on Meaning and half on Form. I was half convinced that I must have spent more time on Form since the focus of both classes was to introduce new grammatical features. I guess it worked that I tried to keep my grammar explanations as little as possible. (Teacher 3)

**Content Control.** Teacher 1 stated that the level of content control she could give to her students depended on how good they were. She seemed reluctant to give more control of the content of classes to her students. However, the COLT result in terms of content control as indicated in Table 4, revealed that she had let her students control the content of the class to a much lesser degree compared to the other two teachers, and in the case of Class B, 100 % of the content was controlled by the text and herself. Seeing this striking result seemed to force

her into thinking about giving more control over class content to her students. She displayed a sense of confusion in how to go about doing this:

I thought I was trying to get my students to have control, but seeing these results, I am not actually doing so. I guess there are times when teachers need to be in control, but when we want to give more control to the students, I am not quite sure how that should be done... (Teacher 1)

Teacher 2 came to alter his belief about content control as well, but interestingly in the opposite direction. Originally, he believed strongly in giving content control to students, as shown in Table 8, and stated that content control by the teacher is not ideal. However, seeing that nearly 90% of his class content in the case of class E was being controlled by students, as indicated in Table 4, he commented “Maybe I have given too much control of the content to my students”.

For Teacher 3, the result of COLT analysis on Content Control was somewhat satisfying. In the interview he demonstrated the view that giving control to the students was important to activate classes. In reality, however, as was shown in Table 4, his students did not have content control over 60% of his class time. He was not shocked to see the results however, because being a novice teacher, he had been aware that giving more content control to his students was a challenge he had to overcome, and thus when he found that around 40% of his class was controlled by his students he was more relieved than shocked.

**Student Modality.** In terms of Student Modality, both Teacher 1 and Teacher 3 came to realise that, contrary to their intentions, their classes spent much less time with students engaged in speaking. As indicated in Table 5, students in both of Teacher 3’s classes were engaged in listening activities for over 70 % of the time. On seeing the result, he commented that such heavy concentration of listening had to be changed. He also mentioned that he had

thought he allocated more time in both classes for students' speaking, but in fact it was less than one fifth. He found it to be problematically short. He commented:

I thought my classes were very communicative and that both my students and myself made use of English, but seeing this result made me realise that my students were not speaking much. This will be something I need to work on next. (Teacher 3)

Teacher 2 reflected on his teaching practices in terms of Student Modality in a unique way. He realised that out of the three classes he had conducted, Class D was closest to his ideal when comparing the balance of the four skills. The remark was striking because he regarded Class D and not Class C as his ideal teaching practice, despite Class C presenting more balanced distribution of the four skills as indicated in Table 5. He preferred Class D over Class C because his emphasis was on listening and speaking without scripts in his classes, and he was not, at that stage, focused on writing.

## **Discussion**

An obvious advantage of having one's own teaching practices analysed through the COLT scheme is that the teacher can self-reflect on his or her teaching through an objective lens. Through the use of COLT analysis as a reflection tool, these three teachers came to realise that some aspects of their teaching practices contradicted their beliefs. The novice teacher realised that contrary to his belief of balanced student modality in his classes, his students were spending far more time listening, something which he felt the need to change. This was not limited to the novice teacher since both of the experienced teachers discovered they were talking more than they had expected which needed to change in order to fulfil their endeavour to make their classes more student-centred.

Using COLT analysis as a reflection tool not only helped the teachers realise that their teaching practices do not reflect their beliefs, but also alter parts of their teacher beliefs. Previous studies on teacher beliefs pointed out their persistent natures, (Bolster, 1983; Shimura, 2010), and therefore this finding can be said to highlight the particular efficacy of using COLT analysis as a self-reflection tool. Teacher 1, who believed that giving content control to her students was not easy and did not think of its significance in her teaching, altered that belief after seeing the COLT analysis. The novice teacher for whom student-centred classes were the ideal and therefore sought to give as much content control to his students as possible, came to think it might be better to have more balance in content control. In his case, proportion of content control may not have been on his list of the issues to be addressed before he saw the COLT analysis, but this interview possibly encouraged him to start seeking the ideal balance.

Using COLT analysis for teachers to self-reflect on their teaching practices was observed to be effective because it was a bottom-up realisation that teachers came to themselves. The realisations were not forced on them from a third person's perspective. In the present study, there were instances where the teacher's actual teaching practices contradicted the teacher beliefs of what an ideal language class should be like. However, not all of the seemingly problematic aspects were considered to be challenges by the teachers themselves. One such example was seen with Teacher 1, who spent nearly two thirds of her class teaching Form, which appeared to contradict her belief that form teaching is pointless. However, this finding did not appear to be a problem to her since in that particular class the focus was on vocabulary. A similar case was found with Teacher 3, a strong proponent of group work. Although the COLT analysis revealed that around 70% of his class was spent in a teacher-centred manner, he acknowledged little dissatisfaction with stating that he was planning to increase the amount of group work as his students become more proficient. It is easy to criticise the practices of teachers, but if the teachers themselves do not agree with the

critique, they are of no use. Using COLT as a self-reflection tool enables teachers to pinpoint their improvements to the places in their teaching practise they find fault with and leave the areas they do not deem to be problematic. Furthermore, COLT is also effective in giving a sense of self-assurance. As was the case with Teacher 3, who found that he managed to give content control to his students for over 40 % of his class. He was satisfied with these results for the time being and aimed to improve on this at a later stage, and so teachers are able to evaluate their teaching practices based on their present teaching levels. It can, as was the case with Teacher 2, be used as a way to compare a teacher's own teaching practices across three different classes in order to find out which class most closely represents the ideal. This can be used as a reference point for future classes. Self-reflection does not always require teachers to change their teaching practices, but can work to support and encourage their practices, or to shed light on their current teaching pathway. Such encouragement and guidance is the strongest when it comes from the teacher themselves, and as Guskey (2002) pointed out, the sense of affirmation or the evidence of improvement is a prerequisite for teachers when they make changes in their teaching practices.

### **Conclusion and Pedagogical Implications**

In response to the first research question, the COLT scheme on CollaVOD was indeed observed to be an effective reflection tool for high school teachers of English in Japan because of its objectivity and its visual representation of the results. Because it was used as self-reflection tool and not as evidence for critique from authoritative figures, it was easily accepted by the teachers and was easily applied when evaluating or changing their teacher beliefs.

In Japan, junior high school and high school teachers are offered many opportunities to observe and critique classes given by their teaching peers. While we acknowledge that these teacher training methods can successfully contribute to the improvement of teaching

practices in Japan, we propose that using the COLT analysis as a self-reflection tool can create another approach to teacher-training, at least in making the language teaching of both experienced and novice teachers more communicative.

Such was the view shared by the teachers who participated in the present study. Teacher 1 commented that she appreciated the objectification of the communicative orientation of her teaching practise which she found to be very convincing. Teacher 2 likewise stated that the objective nature of COLT analysis as well as the visualisation of his teaching practices through CollaVOD was effective in helping him self-reflect on his classes. He also pointed out that because class videos can be easily uploaded onto the CollaVOD platform, teachers who rarely have enough time for self-reflection every day, can easily reflect on their classes over a longer period of time. For the novice teacher it was certainly helpful, and he commented that it was better than the system of peer class-observation and discussion, because his intentions and individual teaching styles can be taken into account. This expresses the effectiveness of self-reflective activities very succinctly.

In response to the second research question that asked if there were teacher variations in the ways they reflect on their teaching practices using the COLT scheme, the answer remains tentative due to the limited number of participants in this study, all of whom displayed different reactions to each other.

During the study, all three participants became aware of elements in their teaching that they had previously been ignorant of, as well as instances where their teaching practices contradicted their teacher beliefs in ways they did not find problematic. The latter point is particularly important as it comes from the teachers themselves and therefore it allows the teachers to focus on the issues they find to be important. Another important finding was that using the COLT scheme as a reflective tool can empower the teachers, assuring them that some of their teacher beliefs are actually being implemented in their practices.

Also, in this study it was documented that even for experienced teachers with more

solid teacher beliefs, making use of COLT succeeded in altering some of their teacher beliefs, indicating its robustness as a teacher training method. Using COLT as self-reflection tool is a powerful approach to teacher training because it gives more control of personal development to the teacher themselves. Larrivee (2000), arguing the importance of having a reflective component in teacher development programmes states:

Becoming an effective teacher involves considerably more than accumulating skills and strategies. Without tying teaching and management decisions to personal beliefs about teaching, learning, and development, a teacher will have only the bricks. The real “stuff” of teaching is the mortar-what holds the bricks in place and provides a foundation. Being successful in today’s classroom environment goes beyond taking on fragmented techniques for managing instruction, keeping students on-task, and handling student behavior. It requires that the teacher remain fluid and able to move in many directions, rather than stuck only being able to move in one direction as situations occur. Effective teaching is much more than a compilation of skills and strategies. It is a deliberate philosophical and ethical code of conduct (pp.293-294).

Using COLT as a reflection tool thus helps English teachers to link communicative teaching theory to their practices, and gives teachers opportunities to reflect on their teaching practices, while simultaneously assuring and altering their teacher beliefs.

There are many limitations to the present study, the small sample size being the most prominent issue, in relation to the second research question in particular. In the future study, this methodology should be replicated with more teachers in various teaching settings.

Despite its limitations, the present study shed light on the implementation of the COLT scheme in teacher training. In the future studies, it will be very indicative to conduct the study in a longitudinal manner with repeated sessions of self-reflection using the COLT

scheme to capture actual teacher development supported by such practices. Another aspect of the study we would like to explore is to conduct group discussion for the teacher training purposes using COLT analysis as stimuli for the discussion. Vescio, Ross, & Adams (2008), synthesising ten articles studying the impacts of professional learning communities on teaching practices and students learning, illustrated the power of teachers learning together to improve their teaching, and how it positively affects the students' achievements as well. Crandall (2000) likewise points to the importance of learning from others, not only novice teachers learning from the experienced but also for the experienced teachers to be able to learn from their new counterparts. Such group discussion based on teachers' self and peer reflections using objective tools such as COLT will benefit group learning in teacher development.

Lastly, objective tools for reflection in teacher development should not be limited to COLT by any means. When we asked the participants of our study about the direction in which they would like to see the research go, two teachers answered that they were interested in how to increase motivation in their students when it comes to language learning. This has inspired us to replicate this study using the MOLT scheme as a self-reflection tool, which may yield findings in another dimension in the improvement of language teaching.

### **Acknowledgement**

The authors would like to express their gratitude to the three teachers who participated in the study. We are also immensely grateful to the anonymous reviewers whose invaluable comments helped us improved the earlier version of the manuscript.

### **References**

Aoki, C., Ishizuka, H., Yokoyama, Y., Sakai, Y., & Kawai, Y.(2008).Using COLT part B classroom observation scheme to analyze communication-oriented English language



- programs[ COLT Part B *ni yoru komyunikeishon wo shikou shita eigo puroguramu no jyugyou bunseki*]. *Research bulletin of English teaching*, 5, 1-25.
- Bolster, A. S. (1983). Toward a more effective model of research on teaching. *Harvard Educational Review*, 53, 294-308. doi: 10.17763/haer.53.3.0105420v41776340
- Crandall, J. (2000). Language teacher education. *Annual Review of Applied Linguistics* 20, 34–55. doi: 10.1017/S026719050020003
- Flanders, N. A. (1970). *Analyzing teaching behavior*. New York: Addison-Wesley.
- Frölich, M., Spada, N. & Allen, P. (1985). Differences in the communicative orientation of L2 classroom. *TESOL Quarterly*, 19(1), 27-57. doi: 10.1002/j.1545-7249.2008.tb00207.x
- Guilioteaux, M. J. & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, 42(1), 55-77. doi: 10.1002/j.1545-7249.2008.tb00207.x
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381-391. doi: 10.1080/135406002100000512
- Ishizuka, H., Yokoyama, Y., Hirata, Y., Aoki, C., Ito, Y., Kawai, Y., Takai, O., & Arai, Y. (2005). Using COLT part A classroom observation scheme to analyze communication-oriented English language programs [COLT Part A *ni yoru komyunikeishon wo shikou shita eigo puroguramu no jyugyou bunseki*]. *Research bulletin of English teaching*, 2, 41-63.
- Ishizuka, H. & Yorozyua, R. (2014). Collaborative VOD platform for classroom observation. In J. Viteli & M. Leikomaa (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology 2014*, 2427-2432.
- Ishizuka, H. & Yorozyua, R. (2016). Jyugyou Kenkyuu wo Kaeru CollaVOD – VOD Onrain Kyoudou Gakushuu Puratto Foomu Kaihatsu. [CollaVOD to change classroom

observation research- developing online platforms for collaborative learning].

*Proceedings of JASELE 42nd National Convention*, 160-161.

Ito, Y. (2016). Nihon no Eigoka Kyouin Yousei no Genjou to Kadai-Senmonsei Kijun • Kijun to Kyouin Yousei Sutanda-do no Shiten kara. [The present state and problems of the curriculum for English teacher education: From a perspective on professional criterion, standard, and teacher education standard.]. *The Bulletin of the Faculty of Expression, Wako University*, 16. 11-20. Retrieved from <http://id.nii.ac.jp/1073/00004046/>

Katagiri, N. & Kawai, G. (2015). Tabulating transcripts and coding on COLT Part B scheme to quantify classroom interaction analysis categories. *HELES Journal* 14, 23-41.

Kawai, Y., Sakai, Y., Yokoyama, Y., Ishizuka, H. & Aoki, C. (2007). Observation scheme of COLT Part B and its challenges. [COLT Part B ni yoru Kansatsu Houhou to sono Mondaiten]. *Media and Communication Studies*, 53,99-113. Retrieved from: <http://hdl.handle.net/2115/34562>

Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice*, 1(3), 293-307. doi: 10.1080/713693162

Moskowitz, G. (1971). Interaction analysis: A new modern language for supervisors. *Foreign Language Annals* 5(2), 211-221. doi: 10.1111/j.1944-9720.1971.tb00682.x

Saiki-Craighill, S. (2008). *Shitsuteki Kenkyu Hou Zemina-ru: Grounded theory approach wo Manabu*. [Study meeting for qualitative research methods: Studying Grounded Theory Approach.] Tokyo: Igaku Shoin.

Shimura, A. (2010). Relationship between English teacher beliefs and their communicatoion-orientedness [Eigo Jyugyou ni okeru Kyoushi no Kangaekata to Komyunikeishon Shikou tono Kanren]. *STEP BULLETIN*, 22, 216-231.

Shimura, A., Sano, A., Sakai, Y., Yokoyama, Y., & Kwaia, Y. (2015). Combining MOLT

perspectives to COLT schemes in assessing instructional events. *Research Bulletin of English Teaching*, 12, 1-25.

Spada, N. & Frölich, M. (1995). *COLT Communicative orientation of language teaching observation scheme, coding conventions and applications*. Sydney: NCELTR Publications.

Vescio, V., Ross, D. & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and students learning. *Teaching and Teacher Education*, 24, 80–91. doi:10.1016/j.tate.2007.01.004

Wallace, M. J. (1991). *Training foreign language teachers: A Reflective Approach*. Cambridge: CUP.

Wragg, E. C. (1970). Interaction analysis in the foreign language classroom. *Modern Language Journal*, 54(2), 116-120. doi: 10.1111/j.1540-4781.1970.tb02244.x

Yokoyama, Y., Shimura, A., La Fay, M., Ishizuka, H., Kawai, Y., & Aoki, C. (2012). Assessing student evaluation of English activities using the COLT observation scheme. *Research Bulletin of English Teaching*, 9, 23 – 44.



## **Exploring Japanese College Student Perceptions of Native and Nonnative English-Speaking Teachers: The Case of Repeaters**

Takaharu Saito

Ryutsu Keizai University

### **Abstract**

The present study examined college repeater attitudes toward native English-speaking teachers (NESTs) and nonnative English speaking teachers (NNESTs) in an EFL context in Japan. The study employed the discourse analytic technique, asking student participants to provide written responses about their opinions on NESTs and NNESTs. The results suggest that college repeaters recognize both positive and negative aspects of both teacher types, illustrating 1) their conception of NESTs as ideal linguistic models and 2) the important role of NNESTs' shared linguistic background to facilitate students' understanding.

**Keywords:** native English-speaking teachers, nonnative English-speaking teachers

### **Introduction**

Native English-speaking teachers (NESTs), known for their native English proficiency, currently play a major role in English language teaching in Japan, given the prevailing assumption that English should be taught and learned monolingually, without use of the learners' local language (Hall & Cook, 2013). NESTs hold a dominant position in English language teaching worldwide (Canagarajah, 1999), with their knowledge and proficiency typically regarded as a point of reference (Stern, 1983).

In contrast, Rajagopalan (2005) claims that nonnative English speaking teachers (NNESTs) are treated as second-class citizens in the TESOL field, because English is "the hottest selling commodity on the foreign language teaching market" (p. 283), privileging NESTs as reliable models for learners. In addition, the perceived credibility issues of

NNESTs, though challenged by both experts and learners, have been widely discussed by NNEST professionals (e.g., Braine, 2010; Kubota, 2002).

The research on student perceptions of NESTs and NNESTs in English language teaching, however, has shown that students in fact appreciate the advantages of both types of teachers, in terms of specific classroom tasks in EFL and ESL settings. Ma's (2012) study in Hong Kong (an EFL setting), for example, notes that the advantages of NNESTs, as perceived by students, include their proficiency in the students' local language, their knowledge of and sensitivity to students' difficulties in learning, and the relative ease felt by students in communicating with them and understanding their teaching. NESTs, meanwhile, are perceived by students to have good English proficiency and the ability to facilitate student learning. The disadvantages of one category of teachers seem to be the reverse of the advantages of the other.

Mahboob (2004) reveals that ESL students at a U.S. university (ESL setting) perceived NESTs as having good oral skills and NNESTs as being good at teaching grammar, and appreciated the NNESTs' experience as foreign language learners. He concludes that ESL students do not have a clear preference for either NESTs or NNESTs, accepting the strengths and unique attributes of both types of teachers.

Medgyes (1994) acknowledges that NNESTs tend to be regarded as less proficient users of English than NESTs, and unable to acquire native linguistic competence, but highly values the following NNEST pedagogical strengths: providing a good learner model, teaching language learning strategies effectively, providing pertinent information about the English language, anticipating and preventing difficulties in learning, showing empathy, and exploiting the shared local language.

Students in EFL contexts, in particular, recognize that some NNESTs have high English proficiency levels, and their shared linguistic background plays an important role in conveying meaning efficiently, maintaining class discipline, reducing student anxiety, and enhancing teacher-learner rapport in the classroom (e.g., Hall and Cook, 2013; Ma, 2012; Saito, 2014). Thus, the theoretical assumption that learning is likely to be more efficient if

teachers draw students' attention to the similarities and differences between the two languages should be given due consideration (Cummins, 2007).

Examining student viewpoints of NESTs and NNESTs is indispensable because, as consumers, their views of both types of teacher are important for improving the latter's collaborative effectiveness in teaching English and understanding their linguistic and pedagogical differences. In light of this, the present study investigated Japanese college repeater perceptions of NESTs and NNESTs, seeking the key factors that influenced such perceptions.

### **The study**

This study examined college repeater attitudes toward NESTs and NNESTs, in an EFL context in Japan. The study employed Mahboobs' (2004) discourse analytic technique, asking student participants to provide written responses to cues seeking their opinions on NESTs and NNESTs, expecting that the students' deeper and more nuanced perceptions would best observed in the dynamics of their free written expression.

The study focused on college repeaters: students who reregistered for classes because they had failed in the previous year for reasons such as obtaining bad grades on mid-term or final exams, long absence from school, and/or failure to establish a good relationship with their English teachers. In the author's teaching experience, such students tend to have low English proficiency, with low motivation in general, and suffer from a sense of inferiority in learning English. Due to the decreasing number of school-age children in Japan, most Japanese colleges, except for highly competitive academic institutions, have recently faced serious enrollment problems, resulting in the acceptance of applicants with lower motivation and/or academic proficiency. Especially in light of this, it is highly important to investigate such students' perception of NESTs and NNESTs, in the hope of critically assisting in the design of optimal English education for such students.

### **Participants**

The author, who taught three repeater English classes at a middle-scale private college in Japan in the fall of 2014, asked his students to complete a relevant questionnaire, and also

obtained permission from three other instructors, teaching the same type of classes, to ask their students to complete the same questionnaire. In total, all 80 students enrolled in repeater English classes were invited to complete a questionnaire on the given topic. Of the six classes, two were courses in basic English conversation, called English D, and the other four were English E which prepared students for the TOEIC test. The repeaters were third- or fourth-year students. The 80 students were asked to write their responses to an open question in their first language, Japanese, as this enabled them to express their thoughts more freely, smoothly, and adroitly than if forced to use their limited English. The administration of the questionnaire was performed in class, at the end of the fall semester, in November and December 2014.

### **Procedure**

Research participants were given the following topic, a slight modification of the stimulus topic in Mahboob's (2004) study<sup>1</sup> and asked to write their responses:

Some students think that only native speakers can be good English teachers. Other students think that Japanese English teachers can also be good English teachers. What is your opinion about this issue? Please feel free to provide details, including your own experiences and examples.

The above topic was presented in Japanese<sup>2</sup>, and participants were asked to answer, in their first language, within approximately 15 minutes. Of the 80 student comments collected, 20 were discarded, as some students only expressed their preference for NESTs or NNESTs without detailed exploration of the issue, while others had no experience of taking lessons from NESTs. Therefore, a total of 60 comments from college repeaters were analyzed in this study.

Replicating Mahboob (2004), discourse analysis was employed in analyzing the students' written comments. The researcher first carefully read the student comments several times, and then sorted them into four types: (1) positive or (2) negative comments regarding NESTs, and (3) positive or (4) negative comments regarding NNESTs. Next, the researcher coded the comments using different highlighters, and generated a relevant list of categories,



which were then labeled and sorted into major groups, with the total number of comments in each category being counted. Therefore, instead of using predetermined categories in the analysis, the categories emerged through the process of data analysis itself, on the assumption that the students' distinctive and animated perceptions would be best revealed in the uncategorized dynamics of the data itself.

### Findings and discussion

Three broad category groups, including eight individual categories, emerged from the data analysis. The first group, "linguistic factors," included "oral skills," "literacy skills," "grammar," and "vocabulary." The second group, "teaching styles," included "ability to use language intelligible to students" and "teaching methods." The third group, "personal factors," included "experience as an FL (foreign language) learner" and "providing emotional support." The results of the analysis follow. As the student responses were written in Japanese, the quoted passages have been translated into English by the author.

#### Linguistic Factors

As noted above, four linguistic factors emerged from student responses about NESTs and NNESTs. The distribution of student comments across these factors is shown in Table 1.

In total, 52 comments were distributed over the linguistic factors: 33 positive (32 for oral skills, 1 for vocabulary) comments about NESTs; and 12 positive (5 for oral skills, 1 for literacy skills, 4 for grammar, and 2 for vocabulary) and 7 negative (oral skills) comments about NNESTs. No negative comments were observed about NESTs.

Table1

*Distribution of Linguistic Factor Comments*

Linguistic factors	NESTs		NNESTs	
	Positive comments (N)	Negative comments (N)	Positive comments (N)	Negative comments (N)
Oral skills	32	0	5	7
Literacy skills	0	0	1	0
Grammar	0	0	4	0
Vocabulary	1	0	2	0
Total	33	0	12	7

**Oral skills.** The oral skills category included comments focusing on the teaching of listening, speaking, and pronunciation. The teaching of oral skills was undoubtedly considered the greatest strength of NESTs, with 32 positive comments noting NESTs' oral skills, the highest number of all the individual categories. The following three examples typify student perceptions about NEST oral-skills teaching:

- I can learn real and proper pronunciation from NESTs. (Student #13)
- I can listen to real English as it is spoken by native speakers. (Student #19)
- The fact that NESTs are native speakers of English is very important. If I were a parent, I would really want my kid to listen to NESTs' real English. (Student #39)

These comments show that NESTs were preferred for teaching oral skills because students felt they could learn accurate and natural pronunciation from them. The frequent use of words like "accurate," "beautiful," "practical," "proper," and "real," in describing NESTs' English, appears to illustrate the students' strong belief that native speakers can provide an ideal linguistic model, especially for English pronunciation.

As compared to NESTs, NNESTs received mixed responses from students in terms of oral-skills teaching. Out of 12 comments about NNEST oral-skills teaching, 5 were positive and 7 negative. The following illustrate positive perceptions of NNEST oral-skills teaching:

- It is easier for me to understand NNESTs' pronunciation. (Student #20)
- NNESTs can give me an intelligible sense of how to pronounce English words. (Student #35)

Some students may have had difficulty in keeping up with NESTs' natural speed of English pronunciation, and might feel more ease in understanding NNESTs' pronunciation with a Japanese accent. Others may have valued NNESTs' ability to identify problems with students' pronunciation, and to explain how to pronounce English words, using the local language, Japanese.

On the other hand, the negative comments about NNEST oral-skills teaching focused primarily on their perceived nonstandard English pronunciation; for example:

- NESTs' English pronunciation is authentic, as it is spoken by native speakers.

NESTs can teach English intonation and stress, while NNESTs cannot. NNESTs' pronunciation is not as good as NESTs'. (Student #19)

The negative responses regarding NNESTs suggest that NNESTs were perceived to be inferior to NESTs in oral-skills teaching simply because NNESTs were themselves nonnatives, without the students expressing unequivocal disapproval of their nonstandard English pronunciation.

**Literacy skills.** Literacy skills here included reading and writing, and received the fewest comments among all the categories. The only comment praised NNESTs' capability for writing instruction:

- NNESTs are good at teaching English writing. It is easy to understand their lessons. (Student #32)

Based on the author's own experience of teaching college repeaters, one reason for this single positive evaluation of NNESTs, and the absence of any positive comments about NESTs, may be that repeaters often do not reach a high level of English literacy, and are thus unable to describe their literacy achievements, or difficulties, in detail. The complexity of literacy acquisition may, in the end, simply exceed the scope of such repeaters' understanding of NEST/ NNEST issues.

**Grammar.** In contrast to the NESTs' forte as oral-skills teachers, grammar was regarded as a NNEST strength, with no comments about NESTs, and four positive comments about NNESTs, in this category. The following were typical of the positive comments:

- It is easy to understand NNESTs' explanation of English grammar. NNESTs should teach grammar-oriented lessons. (Student #2)
- NNESTs can explain English grammar in detail. I feel easy about asking them questions. (Student #16)

These students' use of the word "explain" focuses on NNESTs' ability to explain the rules of English grammar in detail, using the local language, Japanese. In EFL contexts, as in Japan and Hong Kong, NNESTs and their students share the same local language, which enables learners to better understand difficult concepts and important information in lessons. In

contrast, the absence of any positive comments on NESTs, in this respect, suggests that, in the students' perception at least, NESTs may not have explicit awareness of English grammar rules, due to their lack of experience in explicitly learning these rules, though they have natural facility in the language.

**Vocabulary.** A total of three comments were recorded in this category, with one positive comment about NESTs' vocabulary teaching, and two positive comments about NNESTs' vocabulary teaching. No negative comments were observed in the category. The following is the only comment about NESTs as good vocabulary teachers:

- NESTs can kindly teach English vocabulary that I have difficulty understanding, through their own stories. (Student #18)

This student valued the fact that NESTs' vocabulary instruction exploited their own creative stories, unfettered by textbook restrictions. The following is an example of a comment praising NNESTs' vocabulary teaching:

- Since NNESTs are familiar with English vocabulary and its meaning in a Japanese context, it is easy for weak learners to understand their lessons. (Student # 28)

This suggests that NNESTs' knowledge of the foreign language within a Japanese context enhances their ability to teach the vocabulary and explain its meaning to weak learners. Given NESTs' non-textbook-bound vocabulary teaching, and NNESTs' familiarity with English vocabulary in a Japanese context, both teacher types appeared to have their own distinct advantages and were considered by the students to be effective vocabulary teachers.

### **Teaching Styles**

The second broad category, "teaching styles," included two individual categories of comments: ability to use language intelligible to students, and teaching methods. Table 2 shows the distribution of student comments, with a total of 61 comments distributed over the two categories: 6 positive and 21 negative comments about NESTs, and 32 positive and 2 negative comments about NNESTs. The analysis results are discussed below.

Table 2

*Distribution of Teaching Styles Comments*

Teaching Styles	NESTs		NNESTs	
	Positive comments (N)	Negative comments (N)	Positive comments (N)	Negative comments (N)
Ability to use language intelligible to students	0	20	24	1
Teaching methods	6	1	8	1
Total	6	21	32	2

**Ability to use language intelligible to students.** This category refers to teachers' ability to use English and/or the local language (Japanese) appropriately, in order to meet the needs of students in English lessons. A total of 45 comments were recorded in this category, with striking differences in the student perceptions of NESTs and NNESTs: 20 negative comments regarding NESTs, and 24 positive comments (with 1 negative) regarding NNESTs. The following extracts typify the negative student responses regarding NESTs:

- I have difficulty in understanding NESTs' explanations with little use of Japanese. (Student #2)
- I had difficulty in understanding lessons of a NEST from England, because he focused too much on the Queen's English. (Student #3)
- I have no idea what to do in NESTs' English-only lessons. In addition, I have difficulty reading their writing on the blackboard. (Student #9)

These students, as repeaters with low English proficiency in general, had difficulty in understanding NESTs lessons, mainly due to NESTs' English-only rapid speech, and wished to learn the difficult concepts of the English language in the local language, Japanese. In addition, some students expressed difficulty in communicating with NESTs:

- I had difficulty asking NESTs questions. It is difficult to communicate with NESTs. (Student #5)

This comment shows that communication between the NESTs and their students failed because the students were not able to code-switch to Japanese effectively. Thus, it would appear that the NESTs' relative inability to use their students' local language causes some pedagogical difficulty, in both understanding and communicating, for repeaters with limited English proficiency.

In contrast to the overwhelming negativity of student perceptions to NEST teaching in this category, 24 comments regarding NNESTs were positive, with 1 negative. The following comments typify the positive attitudes to NNESTs:

- I can ask NNESTs questions in Japanese. It is helpful to communicate in Japanese. (Student #4)
- Some NNESTs can give me detailed English lessons in Japanese while sometimes communicating in English. (Student #8)
- I think NNESTs, with the shared mother tongue, give a clear explanations to those students who want to learn English. (Student #12)
- It is easier to communicate with NNESTs than NESTs. (Student #40)

Such responses clearly suggest that the students found it easier to understand and communicate with NNESTs than NESTs. In the students' mind, these NNEST strengths were closely associated with, and even dependent on, their appropriate use of the local language, as this enhanced their ability to ensure student understanding in lessons, when explaining difficult concepts for example, and facilitated student-teacher communication. Being able to code-switch to Japanese is a great advantage, for both NNESTs and their students, if communication in English fails. The only negative comment about NNESTs focused on their overuse of written English, without sufficient oral communication in English.

Thus, notably, in this category, NESTs' weaknesses were NNESTs' strengths: students' difficulty in understanding NESTs' instruction was regarded as a disadvantage of their teaching style, while students' ease in understanding NNESTs was considered an advantage of the latter's approach; likewise, students' difficulty in communicating with NESTs was regarded

as a weakness of their style, while students' ease in communicating with NNESTs was perceived as a strength.

**Teaching methods.** In this category, NESTs received 6 positive and 1 negative comment, while NNESTs received 8 positive and 1 negative comment. Typical positive comments regarding NESTs included:

- Taking lessons from NESTs is a good opportunity to practice English since I do not have much opportunity to communicate in English with foreigners in daily life.  
(Student #9)
- NESTs can teach English through their own stories. (Student #18)
- NESTs' English lessons are enjoyable, and increase my eagerness to learn English. English-only lessons give me an opportunity to frequently think in English.  
(Student #41)

Some students appreciated the English-only environment created by NESTs, where students were forced to practice English, and enjoyed NESTs' creative, less textbook-bound lessons. The following was the only negative comment about NESTs:

- It seems that NESTs simply keep speaking English, sometimes speaking broken Japanese. I feel sleepy, have difficulty understanding, and cannot enjoy their lessons. (Student #6)

This student was not engaged by the NESTs' monotonic English lessons with occasional broken Japanese. In contrast, the following comments focused on positive aspects of NNESTs' teaching methods:

- NNESTs can give detailed explanations of English to students, and better understand students' questions. NNESTs are good at teaching the basics of English. (Student #1)
- NNESTs have better teaching skills to engage learners. (Student #6)
- NNESTs can teach English in an effective way, so I can understand well.  
(Student #23)
- NNESTs can give us detailed explanations of English from the shared Japanese

point of view. (Student #45)

This NNEST strength, as perceived by the students, would appear to derive from the linguistic, cultural, and educational experiences they share with their students, which result in greater effectiveness in meeting their students' English learning needs. In particular, sharing the same linguistic background enabled learners to receive detailed explanations for learning English.

The only negative comment about NNESTs was:

- NNESTs can only teach English expressions in the textbook. (Student #18)

This student notes a commonly perceived NNEST weakness: their traditional and textbook-bound teaching approach, which tends to be old-fashioned and inflexible. Since NNESTs themselves learned English as a foreign language, their tendency to rely on textbooks would appear to be unavoidable.

### Personal Factors

Two categories of student comments, experience as an FL (foreign language) learner, and providing emotional support, were grouped together as personal factors. A total of 15 comments were recorded in this group. The comment distribution is shown in Table 3.

Table 3

*Distribution of Personal Factor Comments*

Personal Factors	NESTs		NNESTs	
	Positive comments (N)	Negative comments (N)	Positive comments (N)	Negative comments (N)
Experience as an FL learner	0	0	5	0
Providing emotional support	0	5	5	0
Total	0	5	10	0

**Experience as an FL learner.** This category received a total of 5 comments, all of them positive comments regarding NNESTs, typically characterizing the latter as more effective English teachers because they themselves had had the experience of learning English as a foreign language; for example:



- NNESTs can understand us and teach in an effective way regarding difficulties in learning, because they also had the experience of difficulties in learning.

(Student #36)

Some students in this study felt that NNESTs' experience of learning English as a foreign language made them aware of the problems these students may face, and thus that NNESTs were better equipped to teach them by exploiting this experience.

**Providing emotional support.** This category received 5 negative comments regarding NESTs, and 5 positive regarding NNESTs. The following is illustrative of the students' negativity to NESTs in this respect:

- I have no idea what to do in NESTs' English-only lessons. I tend to have a little fear of NESTs. (Student #9)

- I sometimes shrink in front of NESTs. I feel at ease in asking NNESTs questions. (Student #25)

These students reported that they experienced anxiety when encountering NESTs; or, put another way, NESTs were perceived as weak at creating an emotionally supportive classroom atmosphere which can reduce learners' anxiety level. The students' positive perception of NNESTs in this regard is illustrated by the following comments:

- I always face difficulty in learning English since I am not good at it. NNESTs can respond with care to my problems in learning. (Student #5)
- NNESTs have made a great effort to learn English, and can sympathize with weak learners. (Student #10)

These students' emotional satisfaction with NNESTs is closely related with the shared language learning experience. NNESTs can provide emotional support to their students, because they have gone through the learning process themselves and have a sensitive understanding of the experience.

### **Summary of Findings and Discussion**

The findings of the present study are collated in Table 4.

Table 4

*Distribution of Positive and Negative Comments about NESTs and NNESTs*

Categories	NESTs		NNESTs	
	Positive comments (N)	Negative comments (N)	Positive comments (N)	Negative comments (N)
<b>1. Linguistic factors</b>				
• Oral skills	32	0	5	7
• Literacy skills	0	0	1	0
• Grammar	0	0	4	0
• Vocabulary	1	0	2	0
<b>2. Teaching styles</b>				
• Ability to use language intelligible to students	0	20	24	1
• Teaching methods	6	1	8	1
<b>3. Personal factors</b>				
• Experience as an FL learner	0	0	5	0
• Providing emotional support	0	5	5	0
<b>Total</b>	39	26	54	9

Considering the respective perceived strengths of NESTs and NNESTs, such as NESTs' forte in teaching oral skills (in the linguistic factors), and NNESTs' advantages in teaching styles and personal factors, the two types of teachers seem clearly to complement one another, a result generally corroborating those of Mahboob (2004) and Saito (2014).

The students' overwhelming admiration for NESTs' oral-skills teaching, based on the latter's perceived English proficiency, and in particular their accurate pronunciation, is echoed in a number of other related studies (Lasagabaster & Sierra, 2005; Ma, 2012; Medgyes, 1994; Saito, 2014). In this regard, the conception of NESTs as an ideal linguistic model is highly influential. However, this conception should not be uncritically endorsed, given that the goal of learning English is not merely to conform to the native speaker model, but to use the language internationally as a tool of global communication for mutual intelligibility.

NNESTs' perceived positive traits in teaching grammar, and in the two categories of teaching styles and personal factors, would appear to be closely related to the linguistic, cultural, educational, and language-learning background they share with their students. In

particular, NNESTs' use of the local language, Japanese, plays a crucial role in the effectiveness of their English instruction. These results corroborate the work of Ma (2012) and Medgyes (1994) in EFL contexts, where NNESTs and their learners typically share the same linguistic background.

A number of studies have identified key pedagogical functions of the use of learners' local language, including the efficient conveyance of meaning, maintenance of classroom discipline, teacher-learner rapport, and expressing personal concern and sympathy (e.g., Cook, 2001; Kim & Elder, 2008). These pedagogical functions are also endorsed by the present study, given students' positive responses regarding NNESTs in the categories of teaching styles and personal factors.

At the same time, given students' difficulty in understanding NESTs, and their anxiety in communicating with them, as noted in the comments on teaching styles and personal factors, it is recommended that NESTs at least make adjustments in their speech rate, choice of vocabulary, and writing on the blackboard, in order to embrace these weaker students, who typically suffer from a sense of inferiority.

### **Conclusion**

The results of the present study suggest that college repeaters recognize both positive and negative aspects of both NESTs and NNESTs, in their own learning experience, without expressing a clear preference for either teacher type. The students' admiration for NESTs' pronunciation illustrates their conception of NESTs as ideal linguistic models; yet, at the same time, student comments suggest that NESTs should consider making some pertinent adjustments in their teaching style. On the other hand, NNESTs' shared linguistic background plays an important role in meeting the needs of their students; however, they should consider providing their students with more opportunities to practice English, resisting the tendency to rely overmuch on the shared local language.

Ideally, a collaborative educational system should be established, both to assist NNEST trainees in becoming highly proficient users of English, and to assist NESTs in achieving a greater and more nuanced understanding of their learners' problems. Finally, both theoretical

and pedagogical studies of team teaching at the Japanese college level are encouraged, in order to maximize the strengths, and minimize the weaknesses, of both teacher types.

### Notes

1 The following is the stimulus topic used in Mahboob's study (2004):

“Some students think that only native speakers can be good language teachers. Others think that nonnatives can also be efficient teachers. What is your opinion about this issue? Please feel free to provide details and examples.”

2 The following is a Japanese translation of the stimulus topic, used in the present study.

「ある学生たちはネイティブ・スピーカーの英語の先生だけが、よい先生だと思っています。また他の学生たちは、日本人の英語の先生もよい先生だと思っています。この問題について、あなたはどのように思いますか。あなた自身の経験や事例も含めて、自由にくわしく述べてください。」

### References

- Braine, G. (2010). *Nonnative speaker English teachers: Research, pedagogy, and professional growth*. New York: Routledge.
- Canagarajah, A. S. (1999). Interrogating the “native speaker fallacy”: Non-linguistic roots, non-pedagogical results. In G. Braine (Ed.), *Non-native educators in English language teaching* (pp. 77-92). Mahwah, NJ: Lawrence Erlbaum.
- Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review*, 57(3), 402-423.
- Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics*, 10(2), 221-240.
- Hall, G., & Cook, G. (2013). *Own-language use in ELT: Exploring global practices and Attitudes* (British Council ELT Research Papers 13-01). London: British Council.
- Kim, S. H., & Elder, C. (2008). Target language use in foreign language classrooms:

- Practices and perceptions of two native speaker teachers in New Zealand. *Language, Culture, and Communication*, 21(2), 167-185.
- Kubota, R. (2002). Marginality as an asset: Toward a counter-hegemonic pedagogy for diversity. In L. Vargas (Ed.), *Women faculty of color in the white classroom* (pp. 293-307). New York: Peter Lang.
- Lasagabaster, D., & Sierra, J. M. (2005). What do students think about the pros and cons of having a native speaker teacher? In E. Llurda (Ed.), *Nonnative language teachers: Perceptions, challenges and contributions to the profession* (pp. 217-241). New York: Springer.
- Ma, L. P. N. (2012). Advantages and disadvantages of native-and nonnative-English-speaking teachers: Student perceptions in Hong Kong. *TESOL Quarterly*, 46(2), 280-305.
- Mahboob, A. (2004). Native or nonnative: What do students enrolled in an intensive English program think? In L.D. Kamhi-Stein (Ed.), *Learning and teaching from experience* (pp. 121-147). Ann Arbor, MI: University of Michigan Press.
- Medgyes, P. (1994). *The non-native teacher*. London: Macmillan Publishers.
- Rajagopalan, K. (2005). Non-native speaker teachers of English and their anxieties: Ingredients for an experiment in action research. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 283-303). New York: Springer.
- Saito, T. (2014). Exploring Japanese college students' perceptions of native and nonnative speaker English teachers. *The Journal of Ryutsu Keizai University*, 48(3), 43-54.
- Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford University Press.



## **Students' Uptake and Task Activities as Measures of Effective Learning**

**Yukiko Ohashi**

**Yamazaki Gakuen University**

### **Abstract**

This study explores whether students' uptake can lead to learning and the relationship between students' uptake and the languages and associated activities used in the classroom. The effectiveness of the L1 and the L2 as well as the activities observed in classes were compared. The examination is based on pre-test, post-test, delayed test, and the results of uptake questionnaires administered after classes, which focused on task, mechanical drill, and grammar translation. There was a strong positive correlation between the number of items written in the uptake questionnaire and those correctly answered in the post-test, which implied that students' uptake can lead to their learning and an uptake questionnaire can be used as a measure to evaluate students' learning. Furthermore, the effects of having students work on task activities using the target language instead of grammar translation or mechanical drills were evaluated by administering repeated-measured ANOVAs. The findings show that students' written uptake leads to their learning and task activities facilitated students' uptake more than mechanical drill and grammar translation.

**Keywords:** uptake, task, drill, translation

### **Introduction**

Language teachers have seen a number of different approaches to designing a lesson. As a communicative approach is becoming more prevalent, Task-Based Language Teaching (TBLT) has also been supported. The effectiveness of language education varies according to

the activities the teachers choose, however, accurate judgement of which activity should be chosen is difficult. While some approaches used in language educational settings in Japan require students to translate literary texts after learning rules of grammar, others focuses on having students imitate the correct form and practice mechanically. Although these activities are chosen by the teachers according to the purpose of their classes, teachers are not sure which activity can better lead to students' understanding. In a study of the use of Japanese (L1) and English (L2) in language teaching classrooms, Ohashi (2013) recorded three language classes. It was found that for each class, the languages used by the teachers and students as well as activities, were different, although the purpose of language learning was almost the same. Following this finding, Ohashi (2015) continued the study using a corpus approach which showed that: (a) the activities conducted in class can be categorized into three types: drill, task, and translation; and (b) the differences in language teaching classroom contexts influence the quality and quantity of students' uptake. The findings of these studies confirm that classroom context influences students' uptake, however, whether the students' uptake can lead to their learning or not has not been identified. Also, it implied that the teacher's choice of language and activity influenced the amount of uptake by students, however, a quantitative analysis that can show the effectiveness of the L1 and the L2 as well as the activities in class was not undertaken. McNamara (2008) and Larsen-Freeman and Anderson (2011) argue that a study of activities as well as teaching methods is a way of professional thinking that enables the teacher to know that the classroom activities chosen by them from the alternatives available will most effectively meet the learners' needs according to the aim of the lesson. Thus, by investigating in this current study the relationship between students' uptake and learning, the author will examine whether students' uptake leads to their learning and which type of language and activity best facilitates students' learning.



## **Literature Review**

For the analysis of the study, the author will compare the effectiveness of the three activities: mechanical drill, task, and grammar translation which were also observed in Ohashi (2015) as the activities chosen by the teachers. Definitions are reviewed below.

### **Drill**

Paulston (1970) and Paulston and Bruder (1976) distinguished among three different types of drills: (a) drill, (b) meaningful drills, and (c) communicative drills. Drill is defined as drills where there is only one correct way of responding with complete control of the response. This type of drill is the same as what Hok (1962) called ‘oral drills.’ Hok (1962) defined an oral drill as “the pattern that the students are to imitate either by simply mimicking or by more complicated procedures of combining something new with something already learned” (p.47). The second type, ‘meaningful drills’ is defined by Paulston and Bruder (1976) as the drills that the student cannot complete without fully understanding structurally and semantically what is said. Unlike these drills, communicative drills involve communicative process in which students make sense of expression according to the grammar. The drills that can be observed in EFL (English as a Foreign Language) classes in Japan are either mechanical drills or meaningful drills. Lightbown (2000) suggests that if learners’ practice provides learners with opportunities for meaningful language use, the role of practice is beneficial and even essential.

### **Task**

Ur’s (2012) definition of ‘task’ also includes both focus on using the forms of the language and focus on meaning in completing the task. Ur (2012) states that a ‘task’ has two objectives: (a) learning of some aspect of the language, and (b) an outcome to be evaluated, calling those tasks as ‘language-learning task.’ She states “A good task produces good learning” (p.43) and points out the importance of validity, quality, and success-orientation as underlying practical principles required for tasks. As for validity, tasks should activate

students primarily to use the language items or skills taught in class, which Ellis (2003) also pointed out. The tasks given in Japanese educational settings are within the definitions introduced by Ur (2012), which focus on the connection between the language form and sentence meaning.

### **Grammar Translation**

Larsen-Freeman and Anderson (2011) state that Grammar-Translation Method has been used by language teachers for many years. Japanese students learning English in Japan also experience this method. Teachers teach grammar points through translation. Characteristics of grammar translation introduced by Larsen-Freeman and Anderson (2011). They are: (a) teachers have students translate each sentence and add explanation after students' translation and (b) teachers explicitly teach grammar rules with less attention to speaking and listening. The translation activities observed in Japanese EFL classes can be termed 'Grammar translation' defined above, in which teachers explain grammatical points through the work of translation.

### **Languages Used in Classrooms**

The language used in classrooms changes depending on historical background, the teaching methodology and purpose. Some research showed significant amounts of L1 use in classes (Kaneko, 1991; Polio & Duff, 1994), while there are also studies that show the results of relatively low frequencies of L1 use (Macaro, 2001). Polio and Duff (1994) researched six university EFL classes and revealed that most teachers used the L1 rather than the L2. They described problems that teachers encounter in their use of the L1 and the L2 in the language classroom, such as: (a) Teachers have little idea about when, how, and how often the L1 should be used, (b) Using the L2 requires more time in 'negotiation of meaning,' so teachers tend to resort to the L1, which deprives students of the opportunities to negotiate in the L2 that can develop their strategies.

## **Uptake**

According to Van Lier (1988), language learning occurs through the process of exposure, input, and intake. Exposure includes learners' attention, participation, and interaction. Uptake can be considered to happen in this process. Slimani (1987) and Kaneko (1991) tried to examine how interaction in the classroom affects language acquisition. The aim of these studies was to find out what learners have managed to learn in the midst of the lesson. They used an uptake chart to measure students' reported uptake. In this study as well, the term 'uptake' is synonymous with 'what students claim to have learned in class.' Mackey (2012) points out that by using an uptake chart, it is possible to elicit information regarding learners' perceptions and what they notice in class. The studies using uptake questionnaire such as Slimani (1987) and Kaneko (1991) provided the findings on the relationship between the interaction and uptake. However, whether students' uptake can lead to learning was not shown by these results.

## **Research Questions**

In the above studies using the uptake questionnaires in Slimani (1987) and Kaneko (1991), whether the students' uptake can be regarded as their understanding was not examined in these studies because the required testing of the subject students could not be conducted. As well, in the study of Ohashi (2015) a qualitative analysis to examine whether the items written by the students' as 'uptake' can be identified as their learning was not realized due to the constraints of time and class management. This was pointed out at the time as one of the limitations of the study. As shown in some case studies of Ohashi (2015), it is difficult to conduct the same check test for the subject students when the classes used for the research are managed by different teachers with different opinions. For this reason, uptake questionnaires have not yet shown that students' uptake is synonymous with their learning, which is an issue for further study. If "uptake" should occur in the course of understanding, it is important to know whether the uptake actually leads to learning. Moreover, if students' uptake can be

considered synonymous with their learning, the uptake questionnaire could be used as a tool to measure how well the students understand. This study therefore seeks to investigate the following research questions.

1. Will learners' uptake lead to their learning?
2. Which language of instruction (L1 or L2) is more effective to facilitate learners' uptake and learning?
3. Which activities is the most effective among drill, task, and grammar translation to facilitate learners' uptake and learning?

### **Method**

This study adopted two instruments; (a) three written tests comprised of pre-test, post-test, and delayed post-test, and (b) an uptake questionnaire administered to the participating students. The participants are 40 university students. A written pre-test was administered before each class started and the participants were required to finish it in ten minutes. After a series of treatments was over, the post-test was administered, in which the students were also required to finish in ten minutes. After one week, the delayed post-test was given to the participants. With regard to the vocabulary questions, 18 questions in total were given; nine questions from the part taught in the L1, and nine questions from the part taught in the L2. All vocabulary questions were translating from English words to Japanese words or vice versa. For sentence questions to check sentence uptake, four questions were given; two questions from each part taught in the L1 or L2. The questions to check sentence uptake consisted of filling in the blanks and completing sentences. For grammar questions, ten questions were given; five questions from each part taught in the L1 or L2. Eight of ten grammar questions consisted of filling in the blanks and two questions asked the participants to explain points of grammar from a given sentence. Thus, the participants were supposed to answer 32 questions in total. A perfect score was 32 points. The same questions were used for the pre-test, post-

test, and delayed post-test. The purpose of these tests was to examine whether the students really understood what they wrote in the uptake questionnaire, so all the questions were intentionally made simple and explicit. All the lexical items, sentences, and grammatical points focused on in classes as well as in the pre-test, post-test, and delayed post-test were selected in accordance with those observed in the textbook used in classes. Ten questions are shown in Appendix A as a sample of these tests.

For the uptake questionnaire, the participants were asked to write what they learned in the midst of class without looking at any textbook or materials used during the class. What they wrote in the uptake questionnaire served as data for this study to examine whether the written uptake items were truly learned. In the questionnaire, participants were asked three questions: what new points have come up in today's lesson in terms of (a) vocabulary, (b) sentences, and (c) grammar. The participants were supposed to write English vocabulary, sentences, and grammatical points that they thought they learned or remembered. A picture of a students' answer as written in the uptake questionnaire is shown in Appendix B. The frequencies of items written in each part of the questionnaire were counted. The instructional treatments were *task*, *grammar translation*, and *drill*. Tables 1 and 2 show the descriptions of activities conducted in the study. The procedure of each 90 minutes' class is shown in Table 3. There were six classes in total: two classes focusing on *drill* (first week), two classes of *task* (second week), and two classes of *translation* (third week). The effects of languages and activities used in each class were compared. In this study, the author uses meaningful drills in which mechanical drills were mixed. The tasks used for this study are called 'Language-learning tasks.' The aim of the language-learning tasks used in this study is to have participants use the form of language that the participants learned through the task. Translation activities used for this study include grammar explanation in addition to having participants translate English sentences into Japanese, which is commonly done in Japanese language classes. To measure the participants' improvement in scores, one-way repeated-

measures ANOVA was conducted with the type of activities, namely, *drill*, *language-learning tasks*, and *translation*, as the within-subjects factor with the three scores of the pre-test, post-test, and delayed post-test, which were set as dependent variables. The results include the descriptive statistics for the ANOVA and *t* tests, and pairwise comparison tests, to compare the differences among groups. Next, to compare the effect of languages (L1 or L2) used in class and the three activities: *drill*, *language-learning tasks*, and *translation*, two-way repeated-measures ANOVA was conducted. To examine whether the uptake written in the uptake chart by the participants leads to their learning, the relationship among the three variables below was investigated. They are: (a) the frequency of items written in the uptake chart and observed on the test, (b) the frequency of items written in the uptake chart and correctly answered items on the post-test, and (c) the frequency of items written in the uptake chart and correctly answered items in the delayed post-test. Pearson product-moment correlation coefficients were used for the analysis.

Table 1

*Activity Definitions*

Activities	Definition
Drill	Mechanical and meaningful activities in which students acquire the forms taught by the teachers through examples or explanation.
Language-learning tasks	Non-mechanical activities in which learners use the learned forms of the target language with emphasis on meaning.
Translation	Activities in which teachers explain grammatical points through the work of translation.

Table 2

*Activity Details in Classes*

Activities	Activities done in class	Main language	Grammar points
Drill	1. Students fill in the blanks following the grammatical rules of infinitives with the combination of meaningful drill.	L1	Infinitives
	2. Students fill in the blanks following the grammatical rules of gerunds with the combination of meaningful drill.	L2	Gerunds
Language-learning task	1. Students listen to a story from the teacher and draw a picture of the story, then, exchange information in a group and complete a perfect picture together.	L1	Prepositions
	2. Students read a letter and discuss what kind of advice students can give to the writer in a group, and finally, write a letter back to the writer.	L2	Making a suggestion
Translation	1. Students translate sentences with the grammar points of the comparative or superlative.	L1	Comparative and superlative
	2. Students translate sentences with the grammar points of the relative pronoun.	L2	Relative pronoun

Table 3

*Class Procedure*

Group 1 (20 students)	Group 2 (20 students)
Informed consent & pretest (10 minutes)	Informed consent & pretest (10 minutes)
↓	↓
Lesson focusing on drill, task, or grammar translation using L1 (25 minutes)	Lesson focusing on drill, task, or grammar translation using L2 (25 minutes)
↓	↓
Lesson focusing on drill, task, or grammar translation using L2 (25 minutes)	Lesson focusing on drill, task, or grammar translation using L1 (25 minutes)
↓	↓
Filling in uptake chart (10 minutes)	Filling in uptake chart (10 minutes)
↓	↓
Posttest (10 minutes)	Posttest (10 minutes)
↓	↓
(one week later)	(one week later)
Delayed posttest	Delayed posttest

**Results****One-Way Repeated Measures ANOVA Results**

To measure test-reliability, a split-half coefficient expressed as a Spearman-Brown

corrected correlation and Cronbach alpha coefficient were computed. The 32 post-test questions items for each activity (*drill*, *task*, and *translation*) were split into the odd and even numbers and a correlation was calculated for the two sets of scores. There was a strong positive correlation between the two variables in the *drill* test ( $r = .68$ ,  $\alpha = .81$ ), in the *task* test ( $r = .67$ ,  $\alpha = .80$ ), and in the *translation* test ( $r = .68$ ,  $\alpha = .81$ ). The Cronbach alpha coefficient was .81 for *drill*, and .80 for *task*, and .81 for *translation*.

Table 4 shows the descriptive statistics for the participants' scores on the pre-test, the post-test, and the delayed post-test. A one-way repeated measures ANOVA was conducted to evaluate the effect of activities on the pre-test, post-test, and delayed post-test scores. The independent variables were the instructional treatment: *drill*, *task*, and *translation*. The dependent variables were the participants' scores on the post-test, and delayed post-test.

Regarding the *drill*, the *test* main effect was significant, Wilks's  $\Lambda = .027$ ,  $F(2, 38) = 682.30$ ,  $p < .001$ ,  $\eta^2 = .97$ . The *test* main effect was significant for *task* as well, Wilks's  $\Lambda = .022$ ,  $F(2, 38) = 844.03$ ,  $p < .001$ ,  $\eta^2 = .98$ . The *test* main effect was also significant for *translation*. Wilks's  $\Lambda = .043$ ,  $F(2, 38) = 421.94$ ,  $p < .001$ ,  $\eta^2 = .96$ . The Multivariate Test Results and the univariate test results for the differences between the participants' scores on the pre-test, post-test, and delayed post-test, are shown in Tables 5 and 6, were in accord with the multivariate test results. In the *drill*, the *test* main effect was significant,  $F(2, 78) = 1069.2$ ,  $p < .001$ ,  $\eta^2 = .86$ . The *test* main effect in *task* was significant,  $F(2, 78) = 667.15$ ,  $p < .001$ ,  $\eta^2 = .81$ . Also, the *test* main effect in *translation* was significant,  $F(2, 78) = 619.05$ ,  $p < .001$ ,  $\eta^2 = .81$ . Follow-up paired-samples *t* tests were conducted in order to determine which means differed from each other. Table 7 displays the results. For *drill*, the mean of the immediate post-test, 21.43 ( $SD = 3.98$ ), was significantly higher than the mean of the pre-test, 2.15 ( $SD = 1.96$ ),  $t(39) = 37.417$ ,  $p < .001$ ,  $r = .99$ . The mean of the delayed post-test, 14.68 ( $SD = 3.43$ ), was higher than the pre-test mean,  $t(39) = 12.53$ ,  $p < .001$ ,  $r = .98$ . Regarding *task*, the post-test mean of 24.05 ( $SD = 5.18$ ) was significantly higher than the pre-test mean



of 2.03 ( $SD = 1.42$ ),  $t(39) = 31.66$ ,  $p < .001$ ,  $r = .98$ . The delayed post-test mean of 17.28 ( $SD = 5.57$ ) was higher than the pre-test mean,  $t(39) = 19.93$ ,  $p < .001$ ,  $r = .96$ . For *translation*, the post-test mean of 16.03 ( $SD = 3.61$ ) was significantly higher than the pre-test mean of 2.05 ( $SD = 1.75$ ),  $t(39) = 28.90$ ,  $p < .001$ ,  $r = .98$ . The delayed post-test mean of 10.48 ( $SD = 2.89$ ) was higher than the pre-test mean,  $t(39) = 20.49$ ,  $p < .001$ ,  $r = .96$ . All three activities improved the students' scores, however, the post-test mean was also significantly higher than the delayed post-test mean, (*drill*:  $t(39) = 21.52$ ,  $p < .001$ ,  $r = .96$ ; *task*:  $t(39) = 24.74$ ,  $p < .001$ ,  $r = .97$ ; *translation*:  $t(39) = 20.02$ ,  $p < .001$ ,  $r = .96$ ), implying that the effect of those activities were not sustained for some students. Comparing the descriptive statistics of three activities, *task* had a stronger positive influence on the participants' longest-term memory more than the other activities.

Table 4

*Descriptive Statistics of One-way Repeated Measures ANOVA*

		Activity 1 Drill	Activity 2 Task	Activity 3 Translation
Pre-test	<i>M</i>	2.15	2.03	2.05
	<i>SD</i>	1.96	1.42	1.75
Post-test	<i>M</i>	21.43	24.05	16.03
	<i>SD</i>	3.98	5.18	3.61
Delayed test	<i>M</i>	14.68	17.28	10.48
	<i>SD</i>	3.43	5.57	2.89

Note.  $N = 40$ .

Table 5

*Multivariate Test Results of the One-Way Repeated Measures ANOVA*

Effect			Value	<i>F</i>	<i>p</i>	$\eta^2$
Drill	Test	Pillai's trace	.97	682.30	.00	.97
		Wilks' lambda	.03	682.30	.00	.97
		Hotelling's trace	35.91	682.30	.00	.97
		Roy's largest root	35.91	682.30	.00	.97
Task	Test	Pillai's trace	.98	844.03	.00	.98
		Wilks' lambda	.02	844.03	.00	.98
		Hotelling's trace	44.42	844.03	.00	.98
		Roy's largest root	44.42	844.03	.00	.98
Translation	Test	Pillai's trace	.957	421.94	.00	.96
		Wilks' lambda	.043	421.94	.00	.96
		Hotelling's trace	22.21	421.94	.00	.96
		Roy's largest root	22.21	421.94	.00	.96

Note.  $\alpha = .05$ .

Table 6

*Univariate Test Results of the One-Way Repeated Measures ANOVA*

		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta^2$
Drill	Test	7652.85	2	3826.43	1069.178	.000	0.86
	Residual	946.50	39	24.27			
	Residual	279.15	78	3.58			
Task	Test	10180.85	2	5090.43	667.148	.000	0.81
	Residual	1739.70	39	44.61			
	Residual	595.15	78	7.63			
Translation	Test	3961.12	2	1980.56	619.048	.000	0.81
	Residual	705.30	39	18.08			
	Residual	249.55	78	3.20			

*Note.*  $\alpha = .05$ .

Table 7

*Pair-wise Comparisons Results*

		<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Drill	Pre-test X Post-test	-19.28	3.26	-37.42	.000
	Pre-test X Delayed test	-12.53	2.63	-30.11	.000
	Post-test X Delayed test	6.75	1.98	21.52	.000
Task	Pre-test X Post-test	-22.03	4.40	-31.66	.000
	Pre-test X Delayed test	-15.25	4.84	-19.93	.000
	Post-test X Delayed test	6.78	1.73	24.74	.000
Translation	Pre-test X Post-test	-13.98	3.06	-28.89	.000
	Pre-test X Delayed test	-8.43	2.60	-20.49	.000
	Post-test X Delayed test	5.55	1.75	20.02	.000

*Note.*  $\alpha = .05$ .

### **‘Uptake’ Indicated on the Uptake Chart and the Items Correctly Answered in the Tests**

To examine the correlation between ‘uptake’ written in the uptake chart by the participants and their actual uptake observed in class, Pearson product-moment correlation coefficient tests were conducted. The relationships among the three variables shown below were investigated:

1. The frequency of items written in the uptake chart and also seen on the post-test.
2. The frequency of items written in the uptake chart and also correctly answered on the post-test.
3. The frequency of items written in the uptake chart and also correctly answered on the delayed post-test.

First, the reliability of the results shown in the uptake chart counted by two raters was evaluated. The result of the kappa coefficient between two raters was  $k = .824$ , which means the results counted by two raters showed a strong correlation. Next, the results of the correlational analyses are shown in Table 8. There was a strong positive correlation between all pairs of variables (1, 2, and 3 shown above); in vocabulary uptake, between 1 and 2,  $r = .90$ ,  $n = 40$ ,  $p < .01$ , between 1 and 3,  $r = .75$ ,  $n = 40$ ,  $p < .01$ , and between 2 and 3,  $r = .93$ ,  $n = 40$ ,  $p < .01$ ; in sentence uptake, between 1 and 2,  $r = .91$ ,  $n = 40$ ,  $p < .01$ , between 1 and 3,  $r = .69$ ,  $n = 40$ ,  $p < .01$ , and between 2 and 3,  $r = .75$ ,  $n = 40$ ,  $p < .01$ ; and in grammar uptake, between 1 and 2,  $r = .83$ ,  $n = 40$ ,  $p < .01$ , between 1 and 3,  $r = .71$ ,  $n = 40$ ,  $p < .01$ , and between 2 and 3,  $r = .93$ ,  $n = 40$ ,  $p < .01$ . The results of a one-way repeated measures ANOVA indicated that there was a significant main test effect for tests, and the results of correlation analysis showed there was a strong positive relationship between ‘uptake’ written by the participants in the uptake charts and actual uptake.

Table 8

*Correlations Among Three Items*

Scale	Vocabulary uptake			Sentence uptake			Grammar uptake		
	1	2	3	1	2	3	1	2	3
1. Frequency of items written in the uptake chart and seen in the test	-			-			-		
2. Frequency of items written in the uptake chart and correctly answered in the posttest	.899**	-		.911**	-		.832**	-	
3. Frequency of items written in the uptake chart and correctly answered in the delayed test	.747**	.925**	-	.694**	.746**	-	.706**	.933**	-

Note. \*\*  $p < .01$  (2-tailed).

**Two-Way Repeated Measures ANOVA on Activity and Test Effect**

A two-way within subjects repeated measures ANOVA was conducted to evaluate the effects of instructional treatments: *drill*, *task*, and *translation*, and the language types used in

class. Independent variables were the instrumental treatment with three levels (*drill*, *task*, and *translation*) that the participants received in the classroom and the languages with two levels (*L1* and *L2*) used in class. The dependent variables were the participants' gain scores (pre-test scores subtracted from post-test scores) in the areas vocabulary, sentence, grammar, and the total scores. Table 9 displays the descriptive statistics for the total gain scores for L1 and L2. The mean of *task* was higher than the other two activities in the L2 score. Tables 10 and 11 show the results of the multivariate and univariate tests respectively.

Table 9

*Descriptive Statistics for Total Gain Scores*

Activity			Activity		
L1 Drill	<i>M</i>	6.98	L2 Drill	<i>M</i>	11.08
	<i>SD</i>	2.25		<i>SD</i>	1.94
L1 Task	<i>M</i>	9.58	L2 Task	<i>M</i>	12.53
	<i>SD</i>	2.73		<i>SD</i>	2.01
L1 Translation	<i>M</i>	5.65	L2 Translation	<i>M</i>	8.28
	<i>SD</i>	1.97		<i>SD</i>	1.89

Note. *N* = 40.

Table 10

*Multivariate Test Results on the Total Gain Scores*

Effect		Value	<i>F</i>	<i>p</i>	$\eta^2$
Language	Pillai's Trace	.85	226.47	.00	.85
	Wilks' Lambda	.15	226.47	.00	.85
	Hotelling's Trace	5.81	226.47	.00	.85
	Roy's Largest Root	5.81	226.47	.00	.85
Activity	Pillai's Trace	.75	58.35	.00	.75
	Wilks' Lambda	.25	58.35	.00	.75
	Hotelling's Trace	3.07	58.35	.00	.75
	Roy's Largest Root	3.07	58.35	.00	.75
Language * Activity	Pillai's Trace	.21	4.90	.01	.21
	Wilks' Lambda	.79	4.90	.01	.21
	Hotelling's Trace	.26	4.90	.01	.21
	Roy's Largest Root	.26	4.90	.01	.21

Note. *df* = 1,  $\alpha$  = .05.

Table 11

*Univariate Test Results on the Total Gain Scores*

Effect	SS	df	MS	F	p	$\eta^2$
Language	626.04	1	626.04	226.47	.00	.85
Error (language)	107.46	39	2.76			
Activity	668.33	2	334.16	64.73	.00	.62
Error (activity)	402.68	78	5.16			
Language * Activity	24.02	2	12.01	4.62	.01	.11
Error (Language * Activity)	202.98	78	2.60			

Regarding the multivariate test, the  $F$ -values,  $p$ -values, and partial eta squared values were identical for all criteria. The *activity* main effect was significant, Wilks's  $\Lambda = .25$ ,  $F(2, 38) = 58.35$ ,  $p < .001$ ,  $\eta^2 = .75$ . The *language and activity* interaction was also significant, Wilks's  $\Lambda = 0.79$ ,  $F(1, 38) = 4.9$ ,  $p < .05$ ,  $\eta^2 = .21$ . The univariate test associated with the *language* main effect was significant,  $\Lambda = .147$ ,  $F(1, 39) = 226.47$ ,  $p < .001$ ,  $\eta^2 = 0.85$ .

In order to follow up the significant main and interaction effects, the means of the languages and three activities were computed and pairwise comparisons were conducted. Holm's sequential Bonferroni adjustment was used to control for Type One errors.

Table 12 shows the results of pair-wise comparisons in each test. The mean for *task* ( $M = 11.1$ ,  $SD = 2.17$ ) was significantly higher than the mean for *drill* ( $M = 9.03$ ,  $SD = 2.17$ ),  $t(39) = 5.33$ ,  $p = .000$  ( $< .017$ ),  $r = .65$ . The mean for *drill* was significantly higher than the mean for *translation*,  $t(39) = 6.65$ ,  $p = .000$  ( $< .025$ ),  $r = .73$ , and the mean for *task* was significantly higher than the mean for *translation* ( $M = 6.96$ ,  $SD = 1.50$ ),  $t(39) = 10.67$ ,  $p = .000$  ( $< .05$ ),  $r = .86$ . Considering the results including the descriptive statistics, it was *task* that was the most effective among the three activities, and *drill* follows next. To follow up the significant language main effect, the means of the L1 and L2 scores were computed, and a

paired-samples *t* test was conducted. The mean of the L2 scores on the three tests ( $M = 10.63$ ,  $SD = 1.19$ ) was significantly higher than the mean of the L1 scores on the three tests ( $M = 7.40$ ,  $SD = 1.63$ ),  $t(39) = 15.05$ ,  $p = .000$  ( $< .05$ ),  $r = .92$ . These results provided the evidence that using the L2 is more effective than using the L1.

Table 12

*The Results of Activity Pair-wise Comparisons on the Total Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Drill mean X Task mean	-2.03	2.40	-5.33	0.00
Drill mean X Translation mean	2.06	1.96	6.65	0.00
Task mean X Translation mean	4.09	2.42	10.67	0.00
L1 total X L2 total	-3.23	1.36	-15.05	0.00

Note.  $\alpha = .05$ .

Next, to follow up the significant interaction effect, nine paired-samples *t* tests were conducted. Table 13 shows the results. Again, Holm's sequential Bonferroni adjustment was used. The mean for the L2 was higher than that for the L1 on each pair of the three activities, in *drill*,  $t(39) = 10.64$ ,  $p = .000$  ( $< .006$ ),  $r = .86$ ; in *task*,  $t(39) = 9.27$ ,  $p = .000$ , ( $< .007$ ),  $r = .83$ ; and in *translation*  $t(39) = 6.82$ ,  $p = .000$ , ( $< .01$ ),  $r = .74$ .

For the scores of the activities using the L1, *task* was significantly higher than *drill* and *translation*,  $t(39) = -5.33$ ,  $p = .000$  ( $< .017$ ),  $r = .65$ ;  $t(39) = 7.72$ ,  $p = .000$  ( $< .008$ ),  $r = .78$ , and *drill* was significantly higher than *translation*,  $t(39) = 3.62$ ,  $p = .001$  ( $< .025$ ),  $r = .50$ .

Also, for the scores of the activities using the L2, *task* was significantly higher than *drill* and *translation*  $t(39) = -3.46$ ,  $p = .001$  ( $< .05$ ),  $r = .49$ ;  $t(39) = 10.11$ ,  $p = .000$  ( $< .006$ ),  $r = .85$ , and *drill* was significantly higher than *translation*,  $t(39) = 6.56$ ,  $p = .000$ , ( $< .013$ ),  $r = .73$ . These results imply that whichever language is used, the *task* activity was more effective than other activities.

Table 13

*The Results of Activity and Language Pair-wise Comparisons on the Total Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	$\eta^2$
L1 Drill_ L1 Task	-2.6	3.09	-5.33	.000	0.65
L1 Drill_ L1 Translation	1.33	2.31	3.62	.001	0.5
L1 Task_ L1 Translation	3.93	3.21	7.72	.000	0.78
L2 Drill_ L2 Task	-1.45	2.65	-3.46	.000	0.49
L2 Drill_ L2 Translation	2.8	2.7	6.56	.000	0.73
L2 Task _ L2 Translation	4.25	2.66	10.11	.000	0.85
L1 Drill_ L2 Drill	-4.1	2.44	-10.64	.000	0.86
L1 Task_ L2 Task	-2.95	2.01	-9.27	.000	0.83
L1 Translation_ L2 Translation	-2.63	2.44	-6.82	.000	0.74

Note.  $\alpha = .05$ .

### Vocabulary scores.

The above analysis shows the results for total scores. Next, the gain scores for vocabulary, sentence, and grammar were examined. Table 14 shows the descriptive statistics of vocabulary scores for the total gain scores on the language factor L1 and L2. The *task* mean was higher than the means for the other two activities for the L2 score as well. Tables 15 and 16 show the results of the multivariate and univariate tests. The *F*-values, *p*-values, and partial eta squared values were identical for all criteria. The results indicated that the *activity* main effect was significant, Wilks's  $\Lambda = .177$ ,  $F(2, 38) = 88.05$ ,  $p < .001$ ,  $\eta^2 = .82$ , and the *language and activity* interaction was also significant, Wilks's  $\Lambda = .80$ ,  $F(2, 38) = 4.65$ ,  $p < .05$ ,  $\eta^2 = .20$ . The univariate test associated with the *language* main effect was significant, Wilks's  $\Lambda = .212$ ,  $F(1, 39) = 145.25$ ,  $p < .001$ ,  $\eta^2 = 0.79$ . Table 17 shows the results of the activity pair-wise comparisons. The mean for *task* ( $M = 6.11$ ,  $SD = 1.20$ ) was significantly higher than the mean for *translation* ( $M = 3.29$ ,  $SD = 0.82$ ),  $t(39) = 11.43$ ,  $p = .000$  ( $< .017$ ),  $r = .88$ . The mean for *drill* ( $M = 5.63$ ,  $SD = 1.17$ ) was significantly higher than the mean for *translation*,  $t(39) = 11.05$ ,  $p = .001$  ( $< .025$ ),  $r = .87$ , but the mean for *task* was not significantly higher than the mean for *drill*,  $t(39) = -1.87$ ,  $p = .07$ ,  $r = .29$ . Considering the descriptive statistics and the ANOVA results, *task* was the most effective among the three activities, and *drill* follows next.

Table 14

*Descriptive Statistics for Vocabulary Gain Scores*

Activity			Activity		
L1 Drill	<i>M</i>	4.60	L2 Drill	<i>M</i>	6.68
	<i>SD</i>	1.35		<i>SD</i>	1.44
L1 Task	<i>M</i>	4.70	L2 Task	<i>M</i>	7.53
	<i>SD</i>	1.87		<i>SD</i>	.99
L1 Translation	<i>M</i>	2.43	L2 Translation	<i>M</i>	4.15
	<i>SD</i>	1.26		<i>SD</i>	1.19

*Note.* *N* = 40.

Table 15

*Multivariate Test Results for the Vocabulary Gain Scores*

Effect		Value	<i>F</i>	<i>p</i>	$\eta^2$
Language	Pillai's	0.79	145.26	0.00	0.79
	Wilks'	0.21	145.26	0.00	0.79
	Hotelling's	3.72	145.26	0.00	0.79
	Roy's	3.72	145.26	0.00	0.79
Activity	Pillai's	0.82	88.05	0.00	0.82
	Wilks'	0.18	88.05	0.00	0.82
	Hotelling's	4.63	88.05	0.00	0.82
	Roy's	4.63	88.05	0.00	0.82
Language * Activity	Pillai's	0.20	4.65	0.02	0.20
	Wilks'	0.80	4.65	0.02	0.20
	Hotelling's	0.24	4.65	0.02	0.20
	Roy's	0.24	4.65	0.02	0.20

*Note.* *df* = 1,  $\alpha$  = .05.

Table 16

*Univariate Test Results for the Vocabulary Gain Scores*

Effect	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta^2$
Language	292.60	1	292.60	145.25	.00	.79
Error (language)	78.56	39	2.01			
Activity	366.10	2	183.05	80.41	.00	.69
Error (activity)	177.57	78	2.28			
Language * Activity	12.63	2	6.32	5.26	.01	.12
Error (Language * Activity)	93.70	78	1.20			



Table 17

*The Results of Activity Pair-wise Comparisons on the Vocabulary Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	$\eta^2$
Drill mean X Task mean	-0.48	1.61	-1.87	.069	0.29
Drill mean X Translation mean	2.35	1.35	11.05	0.00	0.89
Task mean X Translation mean	2.83	1.56	11.43	0.00	0.88
L1 total X L2 total	-2.21	1.16	-12.05	0.00	0.00

*Note.*  $\alpha = .05$ .

In order to follow up the significant language main effect, the means of the L1 and L2 scores were computed, and a paired sampled *t* test was conducted. The mean of the L2 scores ( $M = 6.12$ ,  $SD = 0.73$ ) was significantly higher than the mean of the L1 scores on the three tests ( $M = 3.90$ ,  $SD = 0.97$ ),  $t(39) = 12.05$ ,  $p = .000$  ( $< .05$ ),  $r = .89$ . This means using the L2 was more effective than using the L1. Next, nine paired-samples *t* tests were conducted. Again, Holm's sequential Bonferroni adjustment was used. Table 18 shows the results. The mean of the L2 scores was significantly higher than the mean of the L1 scores on each pair of the three activities, in *drill*,  $t(39) = 8.51$ ,  $p = .000$  ( $< .01$ ); in *task*  $t(39) = 10.03$ ,  $p = .000$  ( $< .006$ ); and in *translation*  $t(39) = 6.02$ ,  $p = .000$  ( $< .017$ ),  $r = .89$ . For the scores of the activities using the L1, *task* is significantly higher than *translation*,  $t(39) = 6.42$ ,  $p = .000$  ( $< .013$ ),  $r = .72$ , but not significantly higher than *drill*,  $t(39) = -.29$ ,  $p = .772$ ,  $r = .05$ . Also, *drill* was significantly higher than *translation*,  $t(39) = 8.69$ ,  $p = .000$  ( $< .008$ ),  $r = .81$ . For the scores of the activities using the L2, *task* was significantly higher than *drill*,  $t(39) = 2.98$ ,  $p = .005$  ( $< .025$ ),  $r = .43$  and *translation*,  $t(39) = 13.79$ ,  $p = .000$  ( $< .006$ ),  $r = .91$ . Also, *drill* was significantly higher than *translation*,  $t(39) = 9.26$ ,  $p = .001$  ( $< .007$ ),  $r = .83$ . These results imply that using the L2 in *task* activity was more effective than other activities.

Table 18

*The Results of Activity and Language Pair-wise Comparisons on the Vocabulary Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	$\eta^2$
L1 Drill_ L1 Task	-0.1	2.17	-0.29	0.772	0.05
L1 Drill_ L1 Translation	2.18	1.58	8.69	0.00	0.81
L1 Task_ L1 Translation	2.28	2.24	6.42	0.00	0.72
L2 Drill_ L2 Task	-0.85	1.81	-2.98	0.005	0.43
L2 Drill_ L2 Translation	2.53	1.72	9.26	0.00	0.83
L2 Task_ L2 Translation	3.38	1.55	13.79	0.00	0.91
L1 Drill_ L2 Drill	-2.08	1.54	-8.51	0.00	0.81
L1 Task_ L2 Task	-2.83	1.78	-10.03	0.00	0.85
L1 Translation_ L2 Translation	-1.73	1.81	-6.02	0.00	0.7

Note.  $\alpha = .05$ .

### Sentence scores.

Table 19 shows the descriptive statistics for the two sentence gain scores on two factors of the L1 and L2. The mean of *task* was higher than the other two activities in the L2 score.

Table 19

*Descriptive Statistics for the Sentence Gain Scores*

Activity			Activity		
L1 Drill	<i>M</i>	0.60	L2 Drill	<i>M</i>	1.03
	<i>SD</i>	0.59		<i>SD</i>	0.70
L1 Task	<i>M</i>	1.45	L2 Task	<i>M</i>	1.95
	<i>SD</i>	0.60		<i>SD</i>	0.22
L1 Translation	<i>M</i>	0.50	L2 Translation	<i>M</i>	1.30
	<i>SD</i>	0.51		<i>SD</i>	0.56

Note.  $N = 40$ .

Tables 20 and 21 show the results of the multivariate and univariate tests. The *F*-values, *p*-values, and partial eta squared values were identical for all effects. The results indicated that the *activity* main effect was significant, Wilks's  $\Lambda = .185$ ,  $F(2, 38) = 83.62$ ,  $p < .001$ ,  $\eta^2 = .81$ . but the *language and activity* interaction was not significant, Wilks's  $\Lambda = .269$ ,  $F(2, 38) = 2.69$ ,  $p = .081$ ,  $\eta^2 = .12$ . The univariate test associated with the *language* main effect was significant, Wilks's  $\Lambda = .408$ ,  $F(1, 39) = 56.63$ ,  $p < .001$ ,  $\eta^2 = 0.59$ .

In order to follow up the significant main activity effects, the means of languages and the three activities were computed and pairwise comparisons were conducted. Holm's sequential Bonferroni adjustment was used to control for Type One errors. Table 22 shows the results. The mean for *task* ( $M = 1.7$ ,  $SD = 0.33$ ) was significantly higher than the mean for *drill* ( $M = 0.81$ ,  $SD = 0.52$ ),  $t(39) = -9.63$ ,  $p = .000$  ( $< .025$ ),  $r = .84$ . The mean for *task* was also significantly higher than the mean for *translation* ( $M = 0.9$ ,  $SD = 0.34$ ),  $t(39) = 11.615$ ,  $p = .000$  ( $< .017$ ),  $r = .83$ , but the mean for *translation* was not significantly higher than the mean for *drill*,  $t(39) = -0.93$ ,  $p = .36$ ,  $r = .88$ . To follow up the significant language main effect, the means of the L1 and L2 scores were computed, and a paired-samples  $t$  test was conducted. Also, the mean of the L2 scores ( $M = 1.43$ ,  $SD = 0.30$ ) was significantly higher than the mean of the L1 scores on the three tests ( $M = 0.85$ ,  $SD = 0.4$ ),  $t(39) = -7.53$ ,  $p < .001$ ,  $r = .77$ , providing the evidence that using the L2 was more effective than using the L1.

Table 20

*Multivariate Test Results for the Sentence Gain Scores*

Effect		Value	$F$	$p$	$\eta^2$
Language	Pillai's Trace	0.59	56.63	0.00	0.59
	Wilks' Lambda	0.41	56.63	0.00	0.59
	Hotelling's Trace	1.45	56.63	0.00	0.59
	Roy's Largest Root	1.45	56.63	0.00	0.59
Activity	Pillai's Trace	0.81	83.62	0.00	0.81
	Wilks' Lambda	0.19	83.62	0.00	0.81
	Hotelling's Trace	4.40	83.62	0.00	0.81
	Roy's Largest Root	4.40	83.62	0.00	0.81
Language * Activity	Pillai's Trace	0.12	2.69	0.08	0.12
	Wilks' Lambda	0.88	2.69	0.08	0.12
	Hotelling's Trace	0.14	2.69	0.08	0.12
	Roy's Largest Root	0.14	2.69	0.08	0.12

Note.  $df = 1$ ,  $\alpha = .05$ .

Table 21

*Univariate Test Results for the Sentence Gain Scores*

Effect	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta^2$
Language	19.84	1	19.84	56.63	0.00	0.59
Error (language)	13.66	39	0.35			
Activity	38.28	2	19.14	64.74	0.00	0.62
Error (activity)	23.06	78	0.30			
Language * Activity	1.58	2	0.79	3.33	0.04	0.08
Error (Language * Activity)	18.43	78	0.24			

Table22

*The Results for the Activity Pair-wise Comparisons on the Sentence Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	$\eta^2$
Drill mean X Task mean	-0.89	0.58	-9.63	0.00	0.84
Drill mean X Translation mean	-0.09	0.60	-0.93	0.36	0.83
Task mean X Translation mean	0.80	0.44	11.62	0.00	0.88
L1 total X L2 total	-0.57	0.48	-7.52	0.00	0.77

*Note.*  $\alpha = .05$ .

**Grammar scores.**

Table 23 shows the descriptive statistics for the two total gain scores on two factors of the L1 and L2. The mean of *task* was the highest of all, but a difference in scores does not seem to exist in the scores between the L1 and L2.

Tables 24 and 25 show the statistical results of the multivariate tests and univariate test. The results indicated that the *activity* main effect was significant, Wilks's  $\Lambda = .62$ ,  $F(2, 38) = 11.89$ ,  $p < .001$ ,  $\eta^2 = 0.38$ . The effect size showed that this factor accounted for 38% of the variance, while the *language and activity* interaction was not significant, Wilks's  $\Lambda = .874$ ,  $F(2, 38) = 2.75$ ,  $p = .08$ ,  $\eta^2 = .13$ . The univariate test associated with the *language* main effect was not significant, Wilks's  $\Lambda = .936$ ,  $F(1, 39) = 2.65$ ,  $p = .11$ ,  $\eta^2 = .064$ .

Table 23

*Descriptive Statistics for the Grammar Gain Scores*

Activity			Activity		
L1 Drill	<i>M</i>	2.38	L2 Drill	<i>M</i>	3.05
	<i>SD</i>	1.31		<i>SD</i>	0.85
L1 Task	<i>M</i>	3.53	L2 Task	<i>M</i>	3.50
	<i>SD</i>	1.15		<i>SD</i>	1.11
L1 Translation	<i>M</i>	2.60	L2 Translation	<i>M</i>	2.68
	<i>SD</i>	1.37		<i>SD</i>	1.05

Note. *N* = 40.

Table 24

*Multivariate Test Results for the Grammar Gain Scores*

Effect		Value	<i>F</i>	<i>p</i>	$\eta^2$
Language	Pillai's Trace	0.06	2.65	0.11	0.06
	Wilks' Lambda	0.94	2.65	0.11	0.06
	Hotelling's Trace	0.07	2.65	0.11	0.06
	Roy's Largest Root	0.07	2.65	0.11	0.06
Activity	Pillai's Trace	0.38	11.89	0.00	0.38
	Wilks' Lambda	0.62	11.89	0.00	0.38
	Hotelling's Trace	0.63	11.89	0.00	0.38
	Roy's Largest Root	0.63	11.89	0.00	0.38
Language * Activity	Pillai's Trace	0.13	2.75	0.08	0.13
	Wilks' Lambda	0.87	2.75	0.08	0.13
	Hotelling's Trace	0.14	2.75	0.08	0.13
	Roy's Largest Root	0.14	2.75	0.08	0.13

Note. *df* = 1,  $\alpha$  = .05.

Table 25

*Univariate Test Results for the Grammar Gain Scores*

Effect	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta^2$
Language	3.50	1	3.50	2.65	0.11	0.064
Error (language)	51.66	39	1.32			
Activity	37.63	2	18.82	13.71	0.00	0.26
Error (activity)	107.03	78	1.37			
Language * Activity	5.73	2	2.87	3.12	0.05	0.074
Error (Language * Activity)	71.60	78	0.92			

In order to follow up the significant main activity effect, the means of the three activities were computed and pairwise comparisons were conducted. Holm's sequential Bonferroni adjustment was used to control for Type One errors. Table 26 shows the results of the pairwise comparisons. The mean for *task* ( $M = 3.51$ ,  $SD = 0.96$ ) was significantly higher than the mean for *drill* ( $M = 2.71$ ,  $SD = 0.73$ ),  $t(39) = -4.19$ ,  $p = .000$  ( $< .025$ ),  $r = .56$ . The mean for *task* was also significantly higher than the mean for *translation* ( $M = 2.64$ ,  $SD = 0.98$ ),  $t(39) = 4.58$ ,  $p = .000$  ( $< .017$ ),  $r = .59$ , but the mean for *drill* was not significantly higher than the mean for *translation*,  $t(39) = .43$ ,  $p = .67$ ,  $r = .07$ . Considering the results including the descriptive statistics results, it was *task* that was the most effective among the three activities for grammar scores.

Table 26

*The Results for the Activity Pair-wise Comparisons on the Grammar Gain Scores*

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	$\eta^2$
Drill mean X Task mean	-0.80	1.21	-4.19	0.00	0.56
Drill mean X Translation mean	0.07	1.10	0.43	0.67	0.07
Task mean X Translation mean	0.88	1.21	4.58	0.00	0.59

Note.  $\alpha = .05$ .

## Discussion

This study has investigated whether students' uptake can lead to their learning and the relationship between students' uptakes and the language as well as activities. Before discussing the results, answers to research questions are stated below.

### Answers to Research Questions

The first research question concerned whether or not students' uptake leads to their learning. According to the results of the one-way repeated measures ANOVA conducted to evaluate the effect of the activities (*drill*, *task*, and *translation*) on the pre-test, post-test, and delayed post-test, the main effects of all tests were significant. The students' scores improved

in all the classes where *drill*, *task*, and *translation* were carried out. Moreover, there was a strong positive relationship between ‘uptake’ written by the participants in the uptake questionnaire and the items written in the uptake questionnaire and also correctly answered on the post-test (see Table 8). Considering these results, the students’ written uptake in the uptake questionnaire was shown to be learned and their uptake leads to learning although their learning may not become acquisition.

The second research question was to investigate which language (L1 or L2) is more effective to facilitate learners’ uptake and learning. The results of a two-way repeated-measures ANOVA revealed which language was effective for vocabulary, sentence, grammar, and total uptake. Although both languages targeted the different grammatical point in each activity, the means calculated for all three activities of the L2 were higher than those of the L1 (see Tables 12, 17 and 22). With the vocabulary uptake, the mean of the L2 scores was significantly higher than the mean of the L1 scores on the three tests ( $p < .001$ ,  $r = .89$ ) shown in Table 17. For sentence uptake, the mean of the L2 scores was significantly higher than the mean of the L1 scores ( $p < .001$ ,  $r = .77$ ) in Table 22, and with the total uptake, the mean of the L2 scores was significantly higher than the mean of the L1 scores ( $p < .001$ ,  $r = .92$ ) shown in Table 12. These results provided the evidence that, regardless of the activities as well as grammatical points conducted in class, using the target language (L2) was more effective than using the shared language with the students (L1), while for grammar uptake, the language main effect was not significant. Thus, there was no statistical difference between the effects of using the L2 and the effects of using the L1. These results support the findings shown in Ohashi (2015), that is, that using the L2 was effective to facilitate vocabulary and sentence uptake, while there was no statistical difference between the effect of using the L2 and the effect of using the L1 in grammar uptake.

The third research question was related to whether or not the effectiveness of three activities (*drill*, *task*, and *translation*) varies to facilitate learners’ uptake and learning.

According to the results of a two-way repeated measures ANOVA (shown in Tables 17, 18, 22, and 26), with the vocabulary uptake, the mean for task was significantly higher than the mean for translation ( $p < .001$ ,  $r = .88$ ). Also, the mean for task tended to be higher than the mean for drill, ( $p = .069$  ( $p < .10$ ),  $r = .29$ ). With sentence uptake, the mean for task was significantly higher than the means for *drill* and *translation* ( $p < .001$ ,  $r = .84$ ;  $p < .001$ ,  $r = .83$ ). With grammar uptake, the mean for task was significantly higher than the mean for *drill* and *translation* ( $p < .001$ ,  $r = .56$ ;  $p < .001$ ,  $r = .59$ ). Moreover, in total uptake as well, the mean for task was significantly higher than the mean for *drill* and *translation* ( $p < .001$ ,  $r = .65$ ;  $p < .001$ ,  $r = .86$ ). Considering these results, it is *task* that was the most effective among the three activities.

The results of this study showed that students' uptake influences their study and it will lead to their learning. Also, using the L2 in class can promote students' vocabulary and sentence uptake, while both the L1 and the L2 have a role for grammar uptake. Regarding the three activities compared in this study (*drill*, *task*, and *translation*), tasks contributed to students' uptake the most.

Language-learning tasks are said to require cognitive processing (Ellis, 2003). Ellis (2003) states that learners need to listen, read a text, and display their understanding. This means four language skills, that is, speaking, listening, reading, and writing, are required to complete a task. Plus, a combination of receptive and productive skills is necessary. In the process of working on the task activity, learners have the opportunities of using the words or grammatical forms that they learned in class, which requires cognitive process. That can be considered to contribute to facilitating their sentence uptake. It is considered that in language-learning tasks, students have more opportunities of using the L2 through the communication with their classmates than in other activities, which resulted in more exposure to the L2 in class. It is clear that between the pre-test and post-test conducted in this study, a lot was learned through all three activities. However, there was a drop in scores between the post-test



and the delayed post-test. Nonetheless, the delayed post-test scores were still significantly higher than the pre-test scores. Thus, for a number of the students, their learning was sustained.

In this study, the results revealed that using the L2 is more effective than using the L1 to facilitate vocabulary and sentence uptake, and each language, the L1 and the L2, has a different role in facilitating grammar uptake. However, one of the identified limitations of the study is that the different grammatical points each language focused on in the three activities could have been the variable affecting the results of the study.

Also, language-learning tasks were more effective than drill or translation to enhance all types of uptake. The results of the study pointed out the effectiveness of using the L2 as well as adopting language-learning task in class, supporting the optimal combination of both using the L2 and having the learners work on the tasks. Language-learning tasks showed positive effect on learners' uptake especially when the L2 was used.

### **Conclusions**

The findings of this study showed that students' uptake leads to learning. It is important for language teachers to choose the effective activities to conduct in class as well as the language to use in class so that students' learning can be facilitated. The more uptakes students are able to procure, the more opportunities of learning they would have.

Also, it should be noted that the influence of L2 input from teachers should not be underestimated. However, this does not mean that teachers are not supposed to use the L1 because mixed utterances appeared to enhance students' grammar uptakes. In this study, it was the language-learning task that most effectively facilitated students' uptake in all types of uptake, and using the L2 was more effective than using the L1 in facilitating students' vocabulary and sentence uptake. These results show that depending on the different types of teaching context, the amount of students' uptake would be different as well. Therefore, it will

be very important and helpful for language teachers to understand which activity and/or which language can facilitate students' learning so as to make their classes more effective. Moreover, the usage of an uptake questionnaire sheet is worth noting. What students write in the uptake questionnaire tends to correspond to what they understand in class. The results gained by the examination using uptake questionnaire are trustworthy as long as accurate procedures are taken when uptake questionnaire sheet is used, such as telling the students not to look at the materials which were used in class. The results of the uptake questionnaire could be used to evaluate how well students understand, suggesting there is a use for uptake data in addition to giving them language tests when measuring the effectiveness of the learning experiences of students.

### **Acknowledgement**

I would like to express my appreciation to the anonymous reviewers for their insightful comments and suggestions. This paper is based on the data of my doctoral dissertation but focused solely on the effectiveness of the activities. I am thankful to Professor Tomoko Kaneko, Professor Yoshimasa Ogaswa, Professor Gordon Robson, and Professor Fujiko Sano for their constructive feedback.

### **References**

- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Hok, R. (1962). The challenge of oral drills. *Language Learning*, 12, 47-55. DOI: 10.1111/j.1467-1770.1962.tb00762.x.
- Kaneko, T. (1991). The role of first language in foreign language classrooms (Unpublished PhD dissertation). Temple University, Philadelphia.
- Larsen-Freeman, D. & Anderson, M. (2011). *Techniques & principles in language teaching*. Oxford: Oxford University Press.
- Lightbown, P. M. (2000). Anniversary article: Classroom SLA research and language teaching.

- Applied Linguistics*, 21, 431-462. DOI:<https://doi.org/10.1093/applin/21.4.431>.
- Macaro, E. (2001). Analyzing student teachers' codeswitching in foreign language classrooms: theories and decision making. *The Modern Language Journal*, 85, 531-548. DOI: 10.1111/0026-7902.00124.
- Mackey, A. (2012). *Input, interaction and corrective feedback in L2 learning*. Oxford: Oxford University Press.
- McNamara, T. (2008). Mapping the scope of theory in TESOL. *TESOL Quarterly*, 42(2), 302-305. DOI: 10.1002/j.1545-7249.2008.tb00124.x.
- Ohashi, Y. (2013). Teacher talk and facilitation of students' spontaneous L2 utterance. *Showa Women's University Language Education and Communication Research*, 8, 1-16.
- Ohashi, Y. (2015). A corpus-based study on the relationship between the language used in junior high school classrooms and learners' uptake. *KATE Journal*, 29, 29-42.
- Paulston, C. B. (1970). Structural pattern drills: A classification. *Foreign Language Annals*, 4 (2), 187-93. DOI: 10.1111/j.1944-9720.1970.tb02033.x.
- Paulston, C. B., & Bruder, M. N. (1976). *Teaching English as a second language: Techniques and procedures*. Cambridge, MA: Winthrop.
- Polio, C. & P. Duff (1994). Teachers' language use in university foreign language classrooms: A qualitative analysis of English and target language alternation, *The Modern Language Journal*, 78(3), 313-316. DOI: 10.1111/j.1540-4781.1994.tb02045.x.
- Slimani, A. (1987). *The teaching / learning relationship: learning opportunities and learning outcomes an Algerian case study* (Unpublished PhD dissertation). University of Lancaster, England.
- Ur. P. (2012). *A course in English language teaching*. Cambridge: Cambridge University press.
- Van Lier, L. (1988). *The classroom and the language learner. Ethnography and second-language classroom research*. Harlow: Longman.

## Appendix A

### A sample of pre-test, post-test, and delayed post-test (drill)

(1 is a part to check vocabulary, 2 is to check sentence, and 3 is to check grammar.)

1. 以下の単語を翻訳しなさい。 [trans: translate each word into English or Japanese.]

temperature (	)	運動 (	)
injection (	)	目的 (	)
sit up late (	)	習慣 (	)

2. 以下の文を穴埋めしなさい。 [trans: complete sentences in English below.]

.....is not very easy for me.  
動物の面倒をみるのはそんなに簡単なことではない。

.....in Rome.  
ローマには訪れるところがたくさんある。

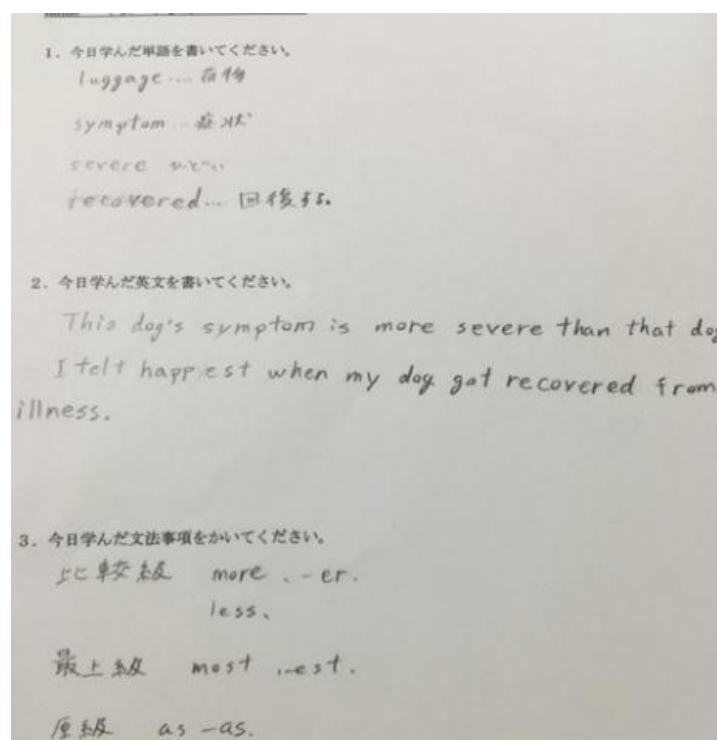
3. 誤った部分に下線を引き正しい形にしなさい。 [trans: underline the wrong parts and correct them.]

1. Use this machine is easy.

2. The work of AHT includes take blood samples, take temperature of animals, and check monitor during operation.

## Appendix B

### A sample of students' answer in the uptake questionnaire



[trans: 1. Write the word that you think you learned in class. 2. Write the sentences or phrased that you think you learned in class. 3. Write about the grammar that you think you learned in class. ]

# **Submission Guidelines**



## ***JACET International Convention Selected Papers, Vol. 4***

### **SUBMISSION GUIDELINES**

Manuscripts for the JACET International Convention Selected Papers (JACET Selected Papers) will only be accepted by online submission. Please read the following guidelines carefully.

#### **Important Dates:**

**Submission Form Open: November 1, 2016**

**Submission Deadline: 11:59 PM Japan Standard Time, January 6, 2017**

#### **Submission Form:**

**<http://www.jacet.org/selected-papers-submission2016/>**

#### **A. Requirements**

1. A paper must be based on a presentation (oral or regular poster presentation) given at the JACET 55th International Convention and the first contributor must be a member of JACET. All other contributors must have also presented the work at the JACET 55th International Convention.
2. A paper based on a plenary lecture may be submitted as an Invited Paper (by invitation only).

#### **B. Editorial Policy**

1. JACET Selected Papers, a refereed, open-access electronic journal, encourages submission of the following:
  - Research Articles on pedagogy and topics of significance to teachers of English
  - Symposium Papers on relevant issues to teachers of English (one per symposium)
  - Practitioner Reports to share findings and insights
2. Manuscripts submitted to JACET Selected Papers must not have been previously published, nor should they be under consideration for publication elsewhere.
3. Manuscripts which do not conform to the guidelines will not be considered for review.
4. Only one paper can be submitted by each contributor.
5. The Editorial Board of JACET Selected Papers reserves the right to make editorial changes in any manuscript accepted for publication to enhance clarity or style. The corresponding author will be consulted if the changes are substantial.
6. Paper offprints will not be provided.

### **C. Guidelines**

1. Manuscripts on A4 paper, including abstract, references, figures, tables, and appendix, should not exceed 30 pages for Research Articles, 20 pages for Symposium Papers, and 15 pages for Practitioner Reports.
2. All manuscripts must be in English.
3. All submissions to JACET Selected Papers must conform to the requirements of the Publication Manual of the American Psychological Association, 6th edition.
  - 3.1 Prepare manuscripts using Microsoft Word (2003 or later).
  - 3.2 Use 12-point Times New Roman font.
  - 3.3 Leave margins of 2.5 cm on all sides of every page (A4 size, 210 mm × 297 mm or 8.27 in × 11.7 in). There are 26 lines to a page.
  - 3.4 Do not justify right margins.
  - 3.5 Do not use running heads.
  - 3.6 Create a paper without the author name(s).
  - 3.7 Include the title, an abstract in English (no more than 200 words), and key words (no more than five words).
  - 3.8 For pagination, use Arabic numerals placed in the upper right-hand corner of each page.
  - 3.9 Delete any textual references that refer to the author(s) and substitute with “\*\*\*\*\*.”

### **D. Submission Procedure**

1. All contributors must complete a submission form on the JACET website, which can be accessed from the Submission Guidelines.
2. Contributors must follow the instructions below.
  - 2.1 Transform the Word file manuscript into PDF format, saving it under the author’s full name as in the following examples: suzukikaoru or smithkerry. Submit the PDF file by clicking “choose file” on the submission form.
  - 2.2 If there are more than four authors, write all authors’ information on a separate file and send to the JACET office by e-mail: Author names, affiliation, membership number, postal code, address, telephone number, and e-mail address. (JACET e-mail address: jacet@zb3.so-net.ne.jp)
  - 2.3 Do not include a cover sheet.

### **E. Contributor’s Responsibility**

1. Contributors are responsible for the content of their manuscripts.
2. Contributors are responsible for obtaining permission to reproduce any material such as figures and tables for which they do not own the copyright, and for ensuring that the appropriate acknowledgements are included in their manuscript.



**F. Copyright**

1. JACET holds the copyright of the articles published in JACET Selected Papers.
2. Anyone, including the author(s), who wishes to reproduce or republish an article, must obtain permission from JACET. Also, it should be clearly stated that JACET holds the copyright.





