

ISSN 1880-2281

# JACET KANSAI JOURNAL

Published by  
Kansai Chapter  
The Japan Association of  
College English  
Teachers

MARCH 2014

NUMBER 16

ISSN 1880-2281

# JACET KANSAI JOURNAL

Published by  
Kansai Chapter  
The Japan Association of  
College English  
Teachers

MARCH 2014

NUMBER 16

**JACET KANSAI JOURNAL**  
THE JOURNAL OF THE KANSAI CHAPTER OF  
THE JAPAN ASSOCIATION OF COLLEGE ENGLISH TEACHERS

**EDITORIAL COMMITTEE 2014**

**CHAIRPERSON**

Yuko Oguri  
*University of Shiga Prefecture*

**MEMBERS**

Kaori Nitta  
*Kinki University*  
Yasuko Sato  
*Otemon Gakuin University*  
Tomoko Yashima  
*Kansai University*

**EDITORIAL OFFICE**

Yasunori Nishina  
*Meiji Gakuin University*

**ACKNOWLEDGEMENTS**

We would like to express our sincere gratitude to the following colleagues who served as reviewers of the papers considered for publication during the preparation of JACET Kansai Journal No. 16:

**LIST OF REVIEWERS** (in alphabetical order)

Nobuhiko Akamatsu ( <i>Doshisha Univ.</i> )	Kei Mihara ( <i>Kinki Univ.</i> )
Sonoko Harada ( <i>Emeritus, Kobe College</i> )	Atsushi Mizumoto ( <i>Kansai Univ.</i> )
Naoya Hase ( <i>Kwansei Gakuin Univ.</i> )	Kazunori Nozawa ( <i>Ritsumeikan Univ.</i> )
Yumi Hato ( <i>Kyoto Institute of Technology</i> )	Takaichi Okuda ( <i>Kansai Univ.</i> )
Keiko Hayashi ( <i>Hiroshima Jogakuin Univ.</i> )	Virginia Peng ( <i>Ritsumeikan Univ.</i> )
Nobuyuki Hino ( <i>Osaka Univ.</i> )	Osato Shiki ( <i>Kwansei Gakuin Univ.</i> )
Maiko Ikeda ( <i>Kansai Univ.</i> )	Tomoko Smith ( <i>Osaka Univ. of Pharmaceutical Sciences</i> )
Hiroyuki Imai ( <i>Kansai University</i> )	Tazu Togo ( <i>Kyoto Notre Dame Univ.</i> )
Keiichi Ishikawa ( <i>Kyoto Women's Univ.</i> )	Masanori Toyota ( <i>Kansai Gaidai Univ.</i> )
Emiko Izumi ( <i>Kyoto Univ. of Education</i> )	Shigeo Uematsu ( <i>Kyoto Sangyo Univ.</i> )
Shuhei Kadota ( <i>Kwansei Gakuin Univ.</i> )	Haruyo Yoshida ( <i>Osaka Kyoiku Univ.</i> )
Toshiyuki Kanamaru ( <i>Kyoto Univ.</i> )	Shinsuke Yoshida ( <i>Kansai Univ.</i> )
Masayuki Kato ( <i>Kobe Univ.</i> )	
Toshiko Koyama ( <i>Osaka Otani Univ.</i> )	

**Copying** All content in this journal is protected by the Japanese copyright law. All rights reserved. The material therein may be reproduced free of charge in any format or medium provided it is reproduced accurately and not used in a misleading context after permission has been granted from the Kansai Chapter of JACET. If any of the items with copyright in this journal are republished or copied, the source of the material must be identified and the copyright status acknowledged.

**Instructions for Contributors** can be found in the appendix and via <http://www.jacet-kansai.org/old/kako.html#kiyou>

掲載論文は、JACET 関西支部、および JACET 関西紀要編集委員会の考え方を反映しているものではありません。

The articles published herein do not reflect the opinions of JACET Kansai Chapter or the JACET Kansai Journal Editorial Committee.

© Kansai Chapter, The Japan Association of College English Teachers, 2014

# JACET Kansai Journal

No. 16 March 2014

---

## **Editorial**

小栗 裕子 iii

## **Invited Article**

根岸 雅史 1  
言語テストの非透明性のインパクト

## **Research Articles**

Ayako HIRANO 13  
The Effects of Output-oriented pre-task vs. Input-oriented pre-task Before a Dictogloss Task

Cheena FUJIOKA 33  
Using the Novelization of the Movie, Dead Poets Society in an EFL Reading Course

Atsuko NISHITANI 50  
A Rasch Analysis of Grammatical Difficulty: What Influences Item Difficulty?

Masumi AZUMA 71  
Figurative interpretations demonstrated by three language groups: English, Chinese and Japanese

## **Research Notes**

藤本 恵子 90  
音韻知識が日本人 EFL 学習者のリスニングに与える影響について

梅原 大輔 / 富永 英夫 103  
日本人英語学習者は主語をどうとらえているか 一量的・質的研究

Kanako YAMAOKA 123  
Qualitative analysis of effects of the use of authentic materials in remedial English classes at a Japanese university

Noriko KIMURA 134  
Factors that affect Japanese EFL learners' motivation to do a learning task

<b>Hiromi MURAKAMI, Yoko YAMADA, Keiko KANAI &amp; Sachiko TAKAGI</b>	154
Designing Materials as a Measure to Facilitate Teacher Development: The Use of an Internet Blog	

## **Appendices**

『JACET Kansai Journal (JACET 関西紀要)』刊行規定	171
『JACET Kansai Journal (JACET 関西紀要)』投稿規定	173

関西支部紀要編集委員会は、*JACET KANSAI JOURNAL No.16*をお届けいたします。今回は過去2年のような特集はありませんが、教師にとっては必須の「テスト」についての論文から始まります。それは、根岸雅史先生（東京外国語大学大学院）の「言語テストの非透明性のインパクト」です。2013年11月に神戸市外国語大学で開催された秋季大会の講演内容で、TOEICと英検を例に挙げながら「テストの透明性」を高めることの大切さを論じておられます。このような大きな英語テストでなくても、我々が日々作成するテストにもこの「透明性」は要求されるのではないのでしょうか。

次に、投稿論文16本のうち査読を経て採択された9本の論文が、提出された順に掲載されています。多くの時間を費やして審査にご尽力くださった査読委員の先生方には、心よりお礼申し上げます。

なお、*JKJ*は限られた予算の中で発行されます。そんな中今回は、例年以上にたくさんの賛助会員の皆様から広告の掲載を通して支部活動のご支援をいただきました。関係各位に厚くお礼申し上げます。また、野口支部長はじめ幹事のみなさんの惜しみないご協力にもお礼申し上げます。編集委員の新田香織先生、佐藤恭子先生、八島智子先生、そして委員会と投稿者および査読者の間で迅速に対応された事務局長の仁科恭徳先生にも感謝の意を表します。

*JACET*が一般社団法人に移行したことにより、投稿規定の著作権は支部から本部に帰属することになりました。また、投稿論文用テンプレートは、できるだけ具体的に説明を加えた英語と日本語版の2つを用意しました。応募される場合は、このどちらかを使用してくださいませう、お願いいたします。*JKJ*16号が日頃の授業改善や研究の一助となることを編集者一同願っております。

*JKJ* 紀要委員会を代表して

小栗裕子

2014年3月8日



Invited Article

## 言語テストの非透明性のインパクト The impact of nontransparency of language tests

根岸雅史  
東京外国語大学大学院

### 1. 言語テストのインパクト研究

テストが学習や指導に与える影響のことを日本語では、「波及効果」という。英語では、backwash effect と言ったり、washback effect と言ったりする。例えば、Fulcher and Davidson (2007: 377)によれば、washback は “Sometimes referred to as ‘backwash’. The effect of a test on learning and teaching. Washback studies focus on practices or behaviour that would not be present if it were not for the test.” というように定義されている。また、その影響を学習や指導といったレベルを超えたものまで含む概念を表すものとしては、impact や consequences といった用語が英語では用いられるようになってきている。

「波及効果研究」における研究対象は、主に「教師」と「学習者」であった。もちろん、テストに直接的に関わるものとしては、「教師」と「学習者」でいいわけであるが、テストのステークスが高くなればなるほど、それが及ぼす影響の範囲は拡大する。Chalhoub-Deville (2009: 119)は、テストのインパクトについて、次のように述べている。

Somewhat simply, impact analysis calls for documentation of the influence of test content, results, and practices on learning, instruction, and the curriculum. Impact could be examined at the micro or macro level. Investigations could be set up to study impact in terms of direct or indirect influences on test takers, the educational community, or society at large. (下線筆者)

こうして考えると、これまでの「波及効果研究」は、主に「ミクロなレベル」の「受験者への直接的な影響」を見てきていると言えるだろう。

### 2. 言語テストはどう語られているか

ここでは、まず、日本における代表的な英語テストの1つである TOEIC を取り上げ、それが社会にどのような影響を及ぼしているかを具体的に見てみる。TOEIC は、日本国内においてビジネス・パーソンを中心に年間 200 万人を超える受験者を誇っているとされる (Retrieved from [http://www.toeic.or.jp/library/toeic\\_data/toeic/pdf/about/transition.pdf](http://www.toeic.or.jp/library/toeic_data/toeic/pdf/about/transition.pdf))。

今日、日本の書店においては、英語学習関連の書籍のスペースは、TOEIC 対策本に埋め尽くされていると言っても過言ではないだろう。では、それらの対策本では、どのような



ことが TOEIC について語られているのであろうか。TOEIC の対策本といったときに、『TOEIC 公式問題集』のように試験の実施団体である「国際ビジネスコミュニケーション協会」が出している書籍もあるが、ほとんどはそうした組織とは関係ない個別の出版社からの書籍である。例えば、私の目を引いた書籍には、『ただのサラリーマンが時間をかけずに半年で TOEIC テストで 325 点から 885 点になれたラクラク勉強法』という異常に長いタイトルの新書がある。これは杉村健一氏といういわゆる「英語オタク」でない「ただのサラリーマン」が書いた対策本で、大筋として英語教育学的に見てもかなり理にかなった学習法が載っていると書いていいだろう。しかし、その中には、「本当の英語力がなくても 900 点台に到達できる」（杉村、2012：21）というような言葉が踊っていたりする。さらに次のようなことも語られている。

...TOEIC は、サービス問題も厳しい問題も基本的には同じ正解 1 問としてカウントされるようなので、難しい問題を正答しても頑張り損らしいです。このため、出るか出ないか分からない所を手広く勉強するよりも、まずは頻出領域から押さえていく勉強をして、易しい問題から正答を狙って行くほうが効率的です。（下線筆者）

杉村（2012：137）

TOEIC の受験を予定している英語学習者たちの多くは、この手の書籍を購入し、対策を立てていると考えられる。

しかしながら、これらの点に関して、TOEIC 公式ページは、実に素っ気ない。

### テストの形式

TOEIC®テストはリスニング（45 分間・100 問）、リーディング（75 分間・100 問）、合計 2 時間で 200 問に答えるマークシート方式の一斉客観テストです。出題形式は毎回同じで、解答はすべて問題用紙とは別の解答用紙に記入します。テストは英文のみで構成されており、英文和訳・和文英訳といった設問はありません。

（Retrieved from <http://www.toEIC.or.jp/toEIC/about/tests.html>）

問題構成に関しても、TOEIC の受験経験があれば、あるいは、『TOEIC 公式問題集』を見れば、だれでもわかるような情報しか提供されていない。

少し具体的に踏み込んでみてみよう。以下は、TOEIC 公式ページにおけるリスニングの応答問題に関する記述である。

#### • 応答問題 30 問

1 つの質問または文章とそれに対する 3 つの答えがそれぞれ 1 度だけ放送される。印刷はされていない。設問に対して最もふさわしい答えを選び解答用紙にマークする。

このリスニングの応答問題に関して、杉村（2012：97）は、次のように語っている。

- 引っ掛けのパターンとしては、
    - 質問文のキーとなる単語を、そのまま間違いの選択肢に使う
    - 質問文のキーとなる単語と発音の似た単語を、間違いの選択肢に使う (degree と agree など)
    - 質問文のキーとなる単語から連想される単語 (package と airmail など) を、間違いの選択肢に使う
    - 時制による引っ掛け
- などがあります。

少なくとも、今日の日本では、このようなテスト開発者自身から発信されたものではない「テストについての語り」が広まり、テスト受験者はその語りをあたかも真実であるかのように信じているという状況が出現していると考えられる。

### 3. 言語テストの不透明性と語り

では、なぜこうした状況が出現しているのであろうか。それは、「言語テストの不透明性 (nontransparency of language tests)」という概念装置によって説明できるだろう。「言語テストの不透明性」という概念は、筆者が Fred Davidson から個人的に聞いた概念である。

言語テスト開発には、次のような様々な段階がある。

#### Stages of test development

1. Deciding what to test
2. Designing test specifications
3. Evaluating, prototyping and piloting
4. Scoring language tests
5. Aligning tests to standards
6. Test administration

Fulcher (2010)

その様々な段階においてなされる判断や行為が、世の中の人々には明らかにされていないのである。テスト・スペックはどのようなものなのか、そして、テストがどのように作成され、どのように採点されるのか等々、その多くは「藪の中」である。このように言語テストの多くの事柄が不透明であるために、テスト作成の当事者以外によって、テストについて様々な推測がなされ、語られるのである。

表1を見てほしい。テスト作成者の持つ情報の多くは、テストの不透明性のために、テスト受験者には直接伝わることはないと考えられる。その両者の間には様々な「仲介者」が存在する。その「仲介者」は、インナー・サークルの「仲介者」とアウター・サークルの「仲介者」から成ると考えられるだろう。インナー・サークルの「仲介者」は、テスト

作成者の外側にいるのであるが、何らかの専門的知見や1次情報へのアクセスの可能性を持った集団である。波及効果研究においてもっばら焦点が当てられてきたのは「教師」であるが、これはインナー・サークルの「仲介者」と考えられる。また、「出版社」の多くも、インナー・サークルの「仲介者」と考えられる。例えば、大学入試では、大学ごとに過去問を載せた書籍や過去問を分析した書籍があるが、そこには出題傾向の分析がなされていたりする。それらは、テスト作成の当事者によるものではないが、ある種の専門性を持った人々によって書かれている。しかし、それにもかかわらず、受験者は全幅の信頼を寄せて読んでいたりする。

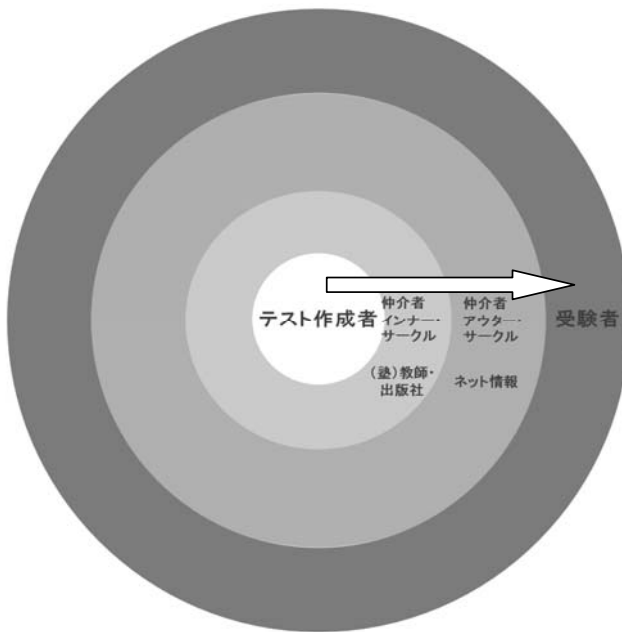


表1. テスト情報はどうか受験者に伝わるか

従来は、こうした集団が受験者にテスト情報をもたらしていたが、今日では、アウター・サークルの「仲介者」の語りが受験者に大きな影響を与えていると考えられる。その語りの代表的なものは、ネット上での語りである。今日では、多くの日本人にとって、パソコンやスマートフォンからのインターネットへのアクセスは日常的になり、情報収集の第1の手段になりつつある。その結果、テストに関する情報も、まずはインターネットから得ようとする傾向が強まっていると思われる。インターネット上の情報には、テスト作成者側が出している公式の情報もあるが、それ以外にもそのテストに関する大量の書き込みがある。上で述べたような「テストの不透明性」のために、公式の情報の中には、受験者が本当に知りたいような情報はほとんど含まれておらず、受験者は自分が知りたい情報に関する「誰か」の情報を信じてしまう傾向にある。しかし、この「誰か」は匿名性が高く、また、その信憑性も様々である。こちらの例としては、日本では「Yahoo!知恵袋」

「2チャンネル」「ブログ」などが挙げられる。こうしたサイトでは、日々様々な受験情報やりとりされている。この種のやりとりの信憑性は、当然のことながら、「インナー・サークルの仲介者」のものよりさらに下がることになる。

こうして実体化する仲介者の語りを「ネッシー・シンドローム」と名付けることにする。ネッシーとは、スコットランドのネス湖で目撃されたとされる、未確認動物の通称であり、正式には「ネス湖の怪獣 (Loch Ness Monster、ロッホ・ネス・モンスター)」とされる。ネッシーは、20 世紀に多くの目撃例が報告されてきたが、目撃証言や写真・映像の多くが、既知動物や船舶、流木、航跡、または波動など自然現象の誤認であるか、あるいは捏造と判定されている (Retrieved from <http://ja.wikipedia.org/wiki/%E3%83%8D%E3%83%83%E3%82%B7%E3%83%BC>)。実際に自分の目で見たことのないネッシーが人々の中で実体化しているプロセスは、テストについての今日の人々の語りとまさに合致する。また、ネッシーを捕らえたとされる写真に見られるネッシーの「波紋」も、テスト作成者からテスト受験者へのインパクトの「波紋」とも、イメージが合致する。

#### 4. 言語テストのビッグ・データ分析

言語テストの波及効果研究は、主に「教師・学習者へのインタビュー」「教師・学習者への質問紙調査」「教室観察」などの手法が用いられていた(e.g. Alderson & Hamp-Lyons; 1996, Watanabe; 1996, Muñoz & Álvarez; 2010)。しかし、ステークスの高い言語テストの場合、その影響はかなりの広範囲に及ぶ。特に、アウター・サークルの仲介者の語りをもたらす影響は、その影響の大きさにもかかわらず、今日まで見過ごされてきている。そこで、本稿では、ビッグ・データに着目し、日本における代表的なテストを例に、それらがどのように語られているのかを見てみたい。

研究対象とした言語テストは、TOEIC と英検 (級を限定しない) である。調査期間は 2013/06/15～2013/07/28 で、解析データ件数は 15,579,018 であった。この中の「TOEIC」または「英検」という言葉を含むサイトを解析したが、商用のサイトのほとんどは、特定の受験対策教材やプログラムの販売を目的としており、データ量も多く、解析結果をゆがめるために、今回の解析からは除いた。この解析データには、ビッグ・データの解析を専門にする Jetrun Technology 社に依頼して行われた。上記期間における「TOEIC」および「英検」の出現サイト数は、それぞれ 1112 と 1009 とほぼ同数であった。本稿で紹介する解析結果は、「ネガティブ・ポジティブ・グラフ」「ワード・マップ」「カテゴリ分析」から成る。

「ネガティブ・ポジティブ・グラフ」は、調査対象となっているもの (この場合は、「TOEIC」と「英検」) が、好意的に語られているのか、否定的に語られているのかを表している。これは Jetrun Technology 社が独自に開発した TrueText という日本語の言語解析エンジンに基づいて自動的に判断している。表 2 と表 3 を見て分かるように、TOEIC と英検は、若干の差はあるものの、過半数の人々に好意的に語られていることが分かる。

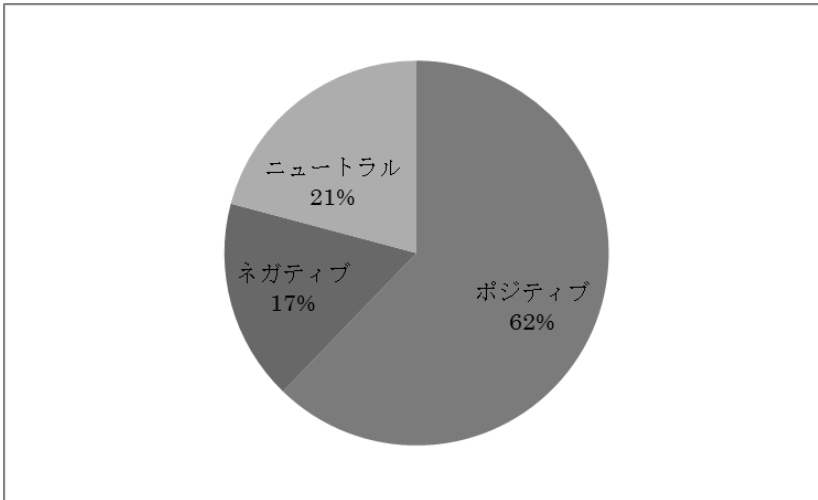


表 2. ネガティブ・ポジティブ・グラフ : TOEIC

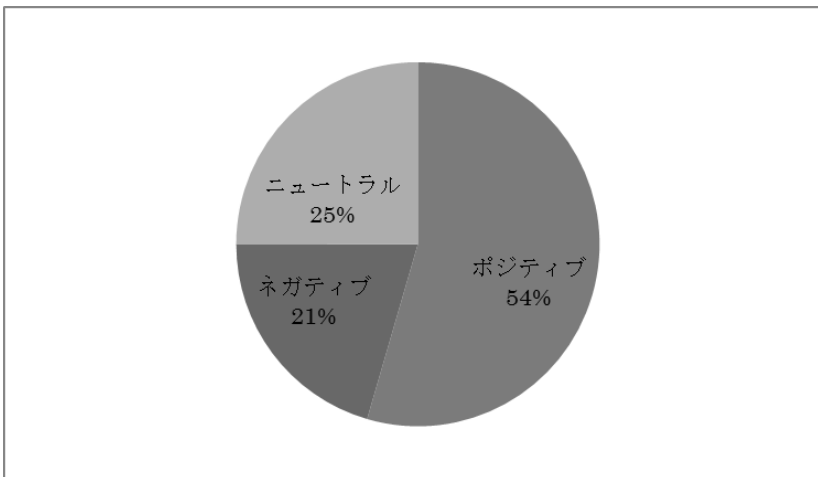
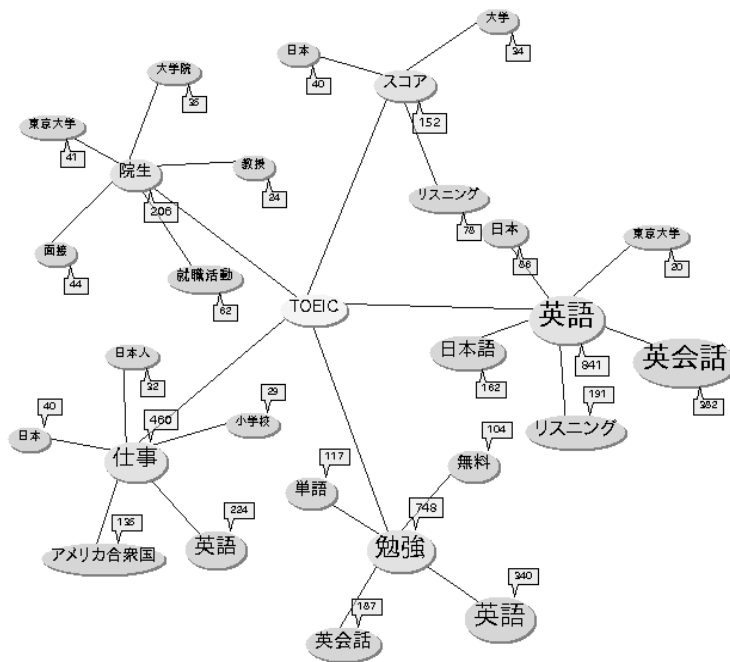
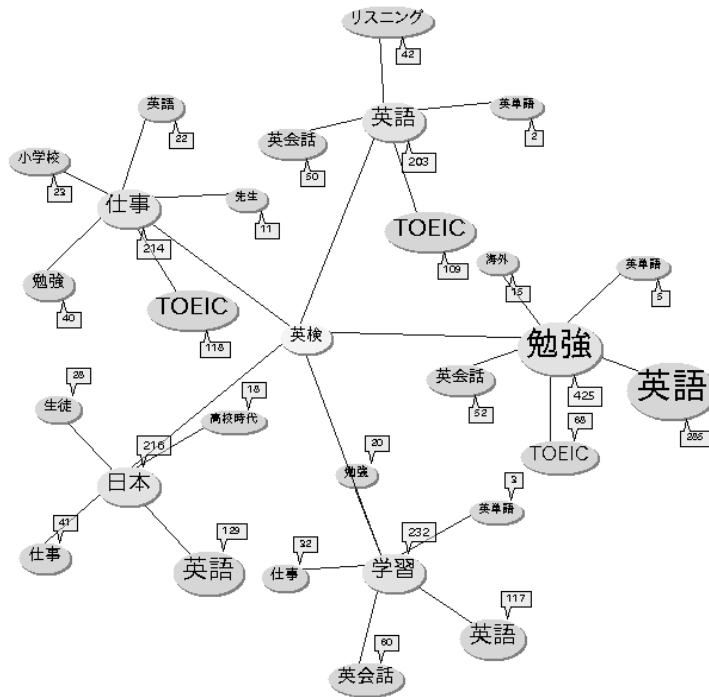


表 3. ネガティブ・ポジティブ・グラフ : 英検

「ワード・マップ」は同一の文に出てくる単語の結びつきを表している。TOEIC で目立つのは（表 4 参照）、「TOEIC→英語(841)」の結びつきの中で強く出ているのが「→英会話(382)」と「→リスニング(191)」である点だ。「TOEIC→勉強 (748)」の結びつきの中でも「→英会話(167)」と出現し、「TOEIC→スコア(152)」の結びつきの中でも「→リスニング(78)」と出現は多い。ネットユーザーに、TOEIC を勉強する上でもスコアを取る上でも「リスニング」や「英会話」が意識されている（場合によっては、苦手意識や問題意識を持っている）ことが分かる。主たる結びつきのうちでもう一つ、「TOEIC→仕事 (460)」と結びついているのも興味深い。これは、日本の TOEIC 受験者が、仕事との結びつきを強く意識していることの表れだろう。

表4. ワード・マップ<sup>o</sup>: TOEIC

これに対して、英検では（表5参照）、大きな結びつきでは「英検→勉強(425)」と「英検→学習(232)」が太い。これは、インターネットユーザーにとっては「英検」というものが（勉強・学習という言葉に象徴される）非常に「学校的なもの」と捉えられている例証である。また、「英語」という言葉がワード・マップのそこかしこに大きな結びつきとして見られる。これは日本人にとって英語といえば英検という連想が強いため、その結果として、日本語での記事・ブログにこれだけの顕れ方をしているという見方もできるかもしれない。

表5. ワード・マップ<sup>o</sup>: 英検

さらに、「カテゴリ分析」に目を向けてみよう。「カテゴリ分析」では、抽出したキーワードに対し、TrueText の独自辞書によるカテゴリを分類し、結果を返却する。TOEIC では（表 6 参照）、上位カテゴリの内 4 位が「ビジネス一般用語」である点が目を引く。仕事やビジネスについてのテキストの中でこれだけ語られているのは TOEFL や英検でも見られない TOEIC だけでの特徴であり 5 位の「英会話学校・英語塾」や 6 位の「受験・受験勉強」より高い数値が出ているのは非常に興味深い。

言語テストの非透明性のインパクト  
The impact of nontransparency of language tests

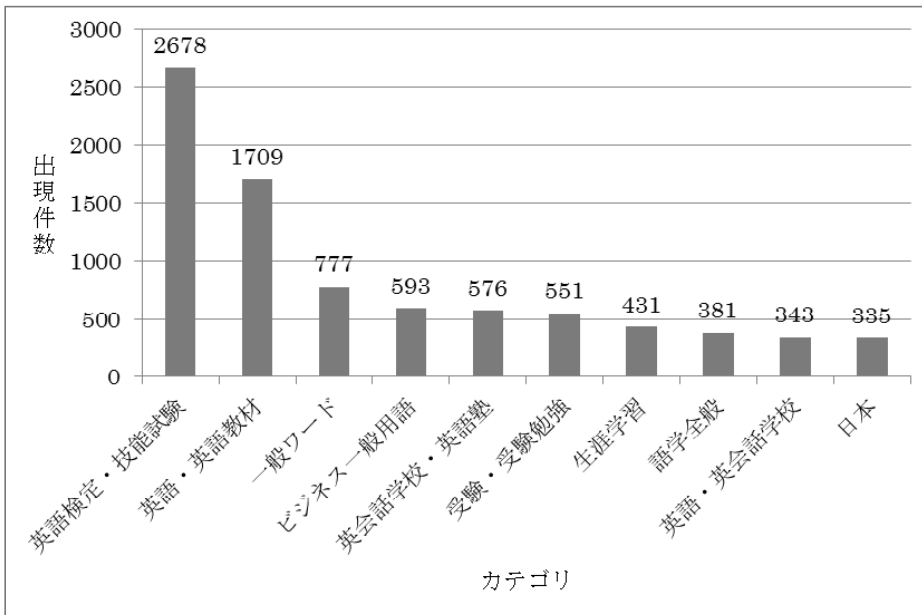


表6. カテゴリ分析：TOEIC

それに対して、英検では（表7参照）、TOEICと対照的に、「英語・英語教材」「受験・受験勉強」「部活動・学校生活」といったカテゴリ属性が顕著で、「生涯学習」「ビジネス一般用語」という文章カテゴリに属するものよりずっと多い。学校や受験に強く紐付けられているというのは、英検の歴史からいっても、英検の当初からの目的からいっても納得がいくだろう。

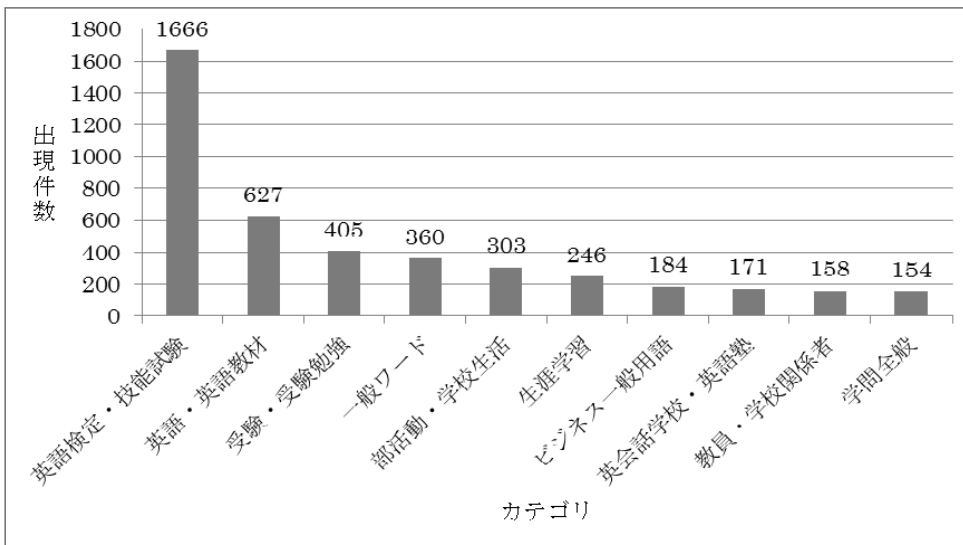


表7. カテゴリ分析：英検



このようなビッグ・データの解析は、ミクロ・レベルの調査では見えなかったものを見せてくれる。しかしながら、この解析に関して、若干の限界もある。それは、一般的な単語はキーワードと認識しないために、言語テストでは重要な概念やトピックであったとしても、それが一般的な単語であれば解析対象となっていない。

こうした弱点を補うために、元データを若干質的に分析しておく。たとえば、「パート (Part)」という単語は、TOEIC の語りの中でよく出てくる。これが出てくるサイトでは、たいていの場合、パートごとの対策はどうすべきかなどについて論じられている。また、TOIEC では、「点」や「スコア」の他に、スコアを示す具体的な数字も頻出している。これらのサイトでは、目標としてのスコアが語られる一方で、「勉強せずに 600 点を取る」という記述のようにそもそもの英語運用能力を上げるのではなく、TOEIC のパターンやルールを理解して効率よく点数を取る方法を伝えている書き込みも存在した。さらに、「スピード」という単語も頻出していることから、素早く解答することが求められており、これに対応しなければならないと受験者が認識していることが分かる。

それに対して、英検は「合格しました」「受かった」「落ちた」というような語りが多いが、これは、英検が TOEIC のようなスコア型のテストとは異なり、合否型であるために当然であろう。さらに、英検では「単語」とか「英単語」がよく出ており、独立した単語問題が各級の 1 番に共通して出題されているために、受験者に強く意識されているように思われる。これは、TOEIC では「リスニング」が強く意識されているのと対照的である。

## 5. テストについての語りの信憑性

では、テストについての語りと実態の関係はどうなっているだろうか。語られている事項が、テストのスペックにそもそも規定されている事柄かどうか、そして、それがテストに顕在化している事柄かどうかによって、およそ 4 つのタイプに分類できるだろう。

### タイプ 1. スペックの規定あり＋テストに見えている事項

このタイプの事例としては、「メディア・チャンネル（筆記試験かパソコンを使うかなど）」「時間」などが挙げられるだろう。これらに関しては、多くの場合スペックで規定されており、外部の人間にも明示的に示されているので、語りの確実性は高いと考えられる。

### タイプ 2. スペックの規定あり×テストに見えていない事項

このタイプの典型的な事例としては、「採点方法」「採点基準 (Accuracy, Appropriacy, Range)」「採点者の資格および訓練」などがある。これらは、外部の人間には見えていないわけであるから、それらについて語られていたとしても、その信憑性は高いとは言えないだろう。

### タイプ 3. スペックの規定なし×テストに見えている事項

テスト・スペックでは、かなり詳細なことが決められているわけであるが、実はすべての事柄が規定されているわけではない。例えば、「選択肢の作成方法」や「テキストの選択」である。選択肢の数やテスト・タイプは決まっても、そこから具体的にどのように選択肢を作っているのかは、個別の作業であり、明文化されていない。また、テキストの選択も同様で、テキスト・タイプやテキストの長さが決まっていたとしても、最終的に特定のテキストを決定するプロセスは、明文化されていない。問題作成者らの隠れた好みや行動のパターンがあるかもしれないのだ。こうした場合は、当然その語りの信憑性は高くはなくなる。

#### タイプ4．スペックの規定なし×テストに見えていない事項

テスト・スペックに規定されておらず、テスト自体を見ても分からない事項もある。やや細かいが、内容理解を問う際の質問文のタイプなどは、スペックにないことも少なくないし、どのようなタイプの問題として出されているのかは、個別のテスト項目を見ても分からない。また、作文においてどのくらいの量を書くとどう評価されるかなども、長めのエッセイ問題の採点においては明確に決まっていないことも多い。こうしたことに関して、外部の人間が何かを語る場合、当然信憑性はかなり低くなると思われる。何しろ当事者が正確に把握していないことを外部の人間が推測することは不可能である。

そもそも、多くのテストにスペックはないという根本的な問題もある。今日実施されているステークスの高い商業ベースの言語テストの多くには、テスト・スペックがあるかもしれないが、高校入試や大学入試には詳細なテスト・スペックを作っているところは一般的ではないだろう。また、仮にテスト・スペックがある場合でも、問題作成者がその存在を知らないとか、見ていない、あるいは、見てはいたが、今では忘れているなど、テスト・スペック通りに必ずしも作っていない。

また、語りの信憑性をさらに下げる要因としては、テスト・スペックに基づいて作っても、外部の人間は、過去の問題から様々な事柄を推定しがちで、自分たちが手にしている情報は、氷山の一角であるという意識がないことが多い。例えば、「テスト構成」「テスト項目の数」「文章の数」「テスト方法」「テスト・タスク」「テキスト・タイプ」「テキストの長さ」「トピック」「リーダビリティ」「文法項目の範囲」「語彙の範囲」などについては、過去の問題はテスト・スペックに基づいて作られているかもしれないが、それがすべてではない。例えば、これまでの「テキスト・タイプ」が「物語」であったとしても、それ以外のテキスト・タイプがテスト・スペックに含まれていて、出題の可能性があるとすると、外部の人間にはそうは見えないだろう。

## 6．結論

本稿では、「言語テストの不透明性」がもたらす言語テストのインパクトについて、アウター・サークルの語りの分析を中心に、論じた。言語テストが不透明であるために、その内容について様々な不正確な語りがなされているにもかかわらず、受験者たちは、その

語りを信じている可能性がある。問題なのは、こうした憶測に基づく語りは、テスト本来の設計理念に反した波及効果をもたらしてしまう可能性があるということだ。だとすれば、テストの透明性を高めることはきわめて重要であると考えられる。

とりわけ説明責任の求められる今日、テスト開発組織はテストの透明性を高めることに及び腰になっているのかもしれない。しかしながら、テスト開発組織は、その社会的責任に基づき、テストの透明性を高める努力をしていかなければならないだろう。なぜならば、誤ったテスト理解は、誤った学習を促進し、そのテストのステークスが高ければ高いほど、大きな社会的損失を生んでしまうからである。

## References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics*, 29, 118-131.
- Fulcher, G. (2010). *Practical language testing*. Hodder Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London, NY: Routledge.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language testing*, 27(1), 33-49.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- 杉村健一. (2012). 『ただのサラリーマンが時間をかけずに半年で TOEIC テストで 325 点から 885 点になれたラクラク勉強法』. アスコム.

## The Effects of Output-oriented pre-task vs. Input-oriented pre-task Before a Dictogloss Task

Ayako Hirano  
Kyoto Sangyo University

### ABSTRACT

The purpose of the current study is to examine the effects of two different cases of tasks on passive voice acquisition and learners' attention during the each task; an output-oriented pre-task and Dictogloss (OT task), and an input-processing based pre-task and Dictogloss (JT task). Dictogloss is one of the Focus on Form (FonF) tasks, consisting of five steps, and the efficacy of the task on learners' grammar acquisition has been reported (Swain, 1995, 2000; Kowal & Swain, 1994, 1997; Lapkin & Swain, 2000). The study was conducted on 46 EFL learners, divided into two groups over two months. The data collection methodology included a grammar test, reconstructed Dictogloss texts, and student pair talk during the reconstruction process in Dictogloss. Data were quantitatively and qualitatively analyzed. The results indicated that OT-task had more favorable influence on learners' passive voice acquisition, and that learner attention to grammatical forms varied across tasks. Furthermore, it was identified that OT-task led students to pay more careful attention to English input during the Dictogloss task. Lastly, the implications of these findings are discussed in relation to Swain's output hypothesis.

**Key words:** output, dictogloss, pre-task, focus on form

### I. INTRODUCTION

There is common agreement among second language acquisition (SLA) researchers and language teachers about the gravity of improving learners' grammatical accuracy. To date, SLA researchers have conducted numerous studies on grammar teaching and emphasized the importance of input (Krashen, 1985; VanPatten, 2007), output (Swain & Lapkin, 1995; Swain, 1995; Storch, 1998; Izumi, 2002; Song & Suh, 2008), interaction (Long, 1983), and noticing (Schmidt, 1990, 1993, 1995). Swain (1995, 2000) claims that output, as well as input, is indispensable in second language learning. Moreover, language teachers have been encouraged to employ classroom activities which require students to focus on form within a meaningful, communicative context (Doughty & Williams, 1998; Williams, 2005).

This study is motivated by my own personal experience with *pushed output* at an interpreter training school and previous grammar teaching experience. In the present study, *pushed output* is defined as a language output produced by stretching learners' interlanguage (IL) to meet communicative goals (Swain, 2000). A typical training activity at the interpreter school is the oral translation task. In this task, interpreter students first listen to a recorded lecture or speech while

taking notes. They then orally translate what they heard into a target language (e.g., A Japanese lecture is translated into English). Afterwards, the teacher presents a model translation. Through the process of oral translation, I always noticed the gap between my interlanguage and the target language (TL). Because the model translation was provided immediately after my own translation attempt, I was able to pay great attention to the words I had found difficult to translate. Since this procedure positively worked to improve my English, I came to believe that the oral output task must be useful for all language learners.

Later, as an English teacher at a vocational language school, I found my students in TOEIC preparation class needed to demonstrate grammatical accuracy. I was adopting some FonF tasks which were used in grammar teaching. Among FonF tasks, Dictogloss seemed suitable for use in my class because previous research has shown the efficacy of the task for grammar acquisition (Swain, 1995, 2000; Kowal & Swain, 1994, 1997; Lapkin & Swain, 2000; Lapkin, Swain, & Smith, 2002; Shak, 2006; Qin, 2008). However, Nabei (1996) reports one drawback: When learners cannot understand the first input, they abandon completing the task. Thus, the comprehensibility of the input determines the success of the Dictogloss task.

Given that learners' initial understanding of the text is essential in Dictogloss, I thought of providing my students with a Japanese translation of the target text. Additionally, I thought having students do the oral translation activity before Dictogloss would be more effective. Consequently, I came to think of two tasks: an output-oriented task, which consists of an oral translation activity before Dictogloss, and an input-processing based task, which consists of Japanese translation support before Dictogloss. As far as I am aware, however, there is little research examining the effects of the two tasks. Therefore, I decided to investigate the effects of both tasks. To this end, I designed and implemented an experiment.

## **II. LITERATURE REVIEW**

### **1. Output and Noticing**

Swain (2000) maintains that output is indispensable for second language learners to "complete grammatical processing needed for accurate production" (p. 99). She emphasizes that input alone is not sufficient for language learning, especially in terms of accuracy. Her claim stems from her research that took place in the context of French immersion programs in Canada. By observing children who had spent years in French immersion programs, she found that the children had a level of competence in the second language (L2) that fell short of native-like abilities. She hypothesizes that what was lacking for them was sufficient opportunities for language use, and then she argues the importance of output.

Kowal and Swain (1997) note that output plays four important roles: It (a) provides the opportunity both for meaningful practice of one's linguistic resources and for developing automaticity in their use; (b) provides the opportunity to test hypotheses- in other words, to try out means of expression and see if they work; (c) prompts responses from speakers of the target language that can provide learners with information about the comprehensibility of their utterances;

and (d) forces the learner to move from semantic processing to syntactic processing. The three functions of output stated by Swain are the noticing triggering role, testing the learner's hypothesis / interlanguage role, and reflecting on the learner's own output role (Kowal & Swain, 1994, 1997; Lapkin & Swain, 2000; Lapkin, Swain & Smith, 2002).

Schmidt (1990) emphasizes that noticing is the necessary and sufficient condition for converting input to intake. He also notes that this requirement of noticing holds true for all aspects of language such as lexicon, phonology, grammatical form, and pragmatics. Consequently, the key factor is how language teachers can effectively draw learners' attention to language forms. Previous research has shown that Focus on Form techniques favorably work to draw learners' attention to the target language items, and the effectiveness of these techniques has been proven (Izumi, 2009).

## **2. Dictogloss**

Focus on Form (FonF), firstly characterized by Long (1991), refers to a range of pedagogical interventions that seek to attract and direct learners' attention to specific formal aspects of the language code in the context of meaningful language use. The FonF approach seeks to briefly shift learners' attention to linguistic code features as problems occur in the context of an otherwise meaning-focused lesson (Ortega, 1999; Long & Robinson, 1998 ).

The present study employs a FonF task called *Dictogloss* or *Grammar Dictation*, which was originally introduced by Wajnryb (1990). She claims that the procedure encourages learners to reflect on their own output. The original Dictogloss procedure consists of five steps: (a) A short, dense text is read (twice) to the learners at normal speed; (b) while it is being read, the learners jot down familiar words and phrases; (c) then, the learners work together in small groups to reconstruct the text from their shared resources; (d) each group of students produces its own reconstructed version, aiming at grammatical accuracy and textual cohesion; and (e) the class analyzed and compares the final versions (Wajnryb, 1990, pp. 5 – 6). In her explanation of Dictogloss, Wajnryb argues that “through both the task of reconstruction and the following error analysis, students refine their understanding of the language they have used” (p. 5). To date, a number of studies have proven the favorable effect of Dictogloss for developing learners' grammatical accuracy (Swain, 1995, 2000; Kowal & Swain, 1994, 1997; Lapkin & Swain, 2000; Lapkin, Swain, & Smith, 2002; Shak, 2006; Qin, 2008).

## **3. Pre-task**

The idea of employing pre-task activities has become prevalent in grammar teaching (Ortega, 1999; Park, 2010). According to Richards (2002), the goals of pre-task activities are (a) to provide language support that can be used in completing a task, (b) to clarify the nature of the task so that students can give less attention to procedural aspects of the task and hence monitor the linguistic accuracy of their performance while carrying out the task. He proposes some pre-task activities such as pre-teaching certain linguistic forms, reducing the cognitive complexity of the task, and giving time to plan the task (pp. 161-162).

For a pre-task activity, which gives time to plan the task, Richards (2002) proposes supplying schemata, vocabulary, and language forms that students can call upon while completing the task. This means that providing vocabulary and translation in the L1 for the model text that will be presented in the L2 in the main task can serve as a pre-task that functions to allow planning time. Also, because this pre-task supports students' comprehension, it is considered an input-processing based task. Furthermore, a pre-task that provides translation in the L1 of the model text can deal with a drawback on Dictogloss that was reported by Nabei (1996), as I mentioned earlier. She states that the difficulty of comprehension could hamper the success of the task.

In terms of another pre-task activity aimed at reducing the cognitive complexity of the task, Richards (2002) suggests providing students with a chance for prior rehearsal of a task. As an example, he proposes a simplified version of a pre-task similar to the one the learners will carry out. In this regard, when the task goal is to reconstruct a model text in the L2, the pre-task can require an oral translation of the model text from the L1 to L2. In this way, the pre-task reduces the cognitive complexity of the main task. Also, because this pre-task requires "pushed output" from the students, it is considered an output-oriented task.

Inspired by previous research on the impact of pre-tasks on IL development (Ortega, 1999; Richards, 2002; Park, 2010), I thought of a pre-task that provides students with text meaning in their L1 before Dictogloss. In addition, I thought of another pre-task that requires students to orally produce their own translation in the L2 after listening to the meaning of the text in the L1. It is predicted that these two kinds of tasks will trigger students' noticing two times. Consequently, the present study examines two types of cases. The first is an output-oriented task: an oral translation activity before Dictogloss (OT task); and the other is an input-processing based task: Japanese translation support before Dictogloss (JT task). Two specific research questions are addressed in this study.

- (1) Do the OT task and the JT task promote language learners' acquisition of English passive voice constructions?
- (2) Does the reconstruction process in Dictogloss contribute to learners' negotiation of a target form? More specifically, what do the students notice and discuss during the reconstruction process?

The research question contains one term that need to be defined prior to the study: *acquisition*. Several researchers distinguish between two types of *acquisition*: *acquisition* as the internalization of new forms, and *acquisition* as an increase in control over forms that have already been internalized (Ellis, 1997; Bialystok & Sharwood Smith, 1985; de Bot, 1996; Lyster, 2004). In the present study, I apply the latter definition, an increase in control over forms that have already been internalized.

### III. METHOD

#### 1. Setting and Participants

The data were collected in three intact, credit-bearing TOEIC classes, which I taught at a vocational school in Japan. The participants were 58 EFL learners (50 women and 8 men) aged 18



to 26. Their nationalities were Japanese (46), Chinese (9), Korean (2), and Thai (1). The international students' Japanese proficiency was regarded as advanced, as all had passed Level 1 or 2 of the Japanese Language Proficiency Test (JLPT) before enrolling at the school. All participants' English proficiency was considered intermediate, based on their TOEIC scores taken in July, ranging from 405 to 795 (average 633). They were divided into two groups, an oral translation activity group (OTG) and a Japanese translation support group (JTG). The OTG consisted of one first-year class and one second-year class, and the JTJG, one first-year student class. The data of those who were absent from more than two treatment sessions or who did not take the pre- or post-tests were discarded. In addition, the data of the triad were removed. Thus, the final analysis considered 46 students, 27 (5 international students) in the OTG, and 19 (6 international students) in the JTJG.

## 2. Research Schedule

The study was carried out in a 90-minutes class from October to December 2011, over a period of about two months. The study consisted of two practice sessions, a pre-test session, four treatment sessions, and a post-test session. A research schedule for this study can be seen in Table 1. At sessions 1 and 2, the students underwent two practice sessions and completed practice Dictogloss 1 and 2. During the two practice sessions, the students were provided instruction on how to operate an IC recorder. In session 3, they were given a pre-test consisting of a grammar test and a pre-Dictogloss test.

Subsequently, from sessions 4 through 7, students were given treatment and carried out Dictogloss 1, 2, 3, and 4. During the four treatment sessions, each dyad recorded their interaction with an IC recorder. After the treatment sessions, in session 8, the students were given a post-test, which was the same as the grammar test given in session 3.

**Table 1. Research schedule**

Session	Date	IC recording	Test	Activity
1	10/18, 10/26	○		Practice Dictogloss 1
2	10/25, 10/28	○		Practice Dictogloss 2
3	10/27, 11/2		pre-test	
4	11/8, 11/9	○		Treatment Dictogloss 1
5	11/15, 11/16	○		Treatment Dictogloss 2
6	11/24, 11/25	○		Treatment Dictogloss 3
7	11/29, 11/30	○		Treatment Dictogloss 4
8	12/1, 12/2		post-test	

## 3. Target Form

The target forms in this study were English passive voice constructions (e.g., *I was introduced to a manager*) and perfect form (e.g., *I have lost my key. / I had lost my key*). The reason for selecting these forms derives from the results of a term-end test in the previous year. The test



was taken by 210 students enrolled in TOEIC classes at the vocational school where I was teaching. The results indicated passive voice and perfect form were their weak points in terms of grammatical forms. In the present study, however, the coefficient of reliability of perfect form in the grammar test results was relatively low (pre-test,  $\alpha = .35$ ; post-test,  $\alpha = .71$ ). Therefore, the results of perfect form in the grammar test are excluded in the study.

#### 4. Material

This study utilized three kinds of materials: (a) grammar test, (b) pre-Dictogloss text, and (c) four Dictogloss texts.

##### 4.1 Grammar Test

The grammar test was administered as pre- and post-tests to assess the participants' productive ability of the passive voice. The sentences were taken from *Grammar in Use Intermediate* (Murphy, 2009) and *600 Essential Words for the TOEIC* (Lougheed, 2008). The test consisted of 20 independent questions: 10 passive voice questions and 10 perfect form questions. At the end of each question, the dictionary form of a verb was given. The students were directed to fill in the blank by changing the given verb form into one that fit grammatically within the sentence. Blanks in the passive voice questions were expected to be filled in with the target form, be + the past participle verb form.

10 perfect form questions out of 20 questions were excluded when the test was scored. Students received two points for each correct passive voice form; therefore, the maximum score was 20. The answers were classified into three types: (a) correct *be* + past participle, (b) correct *be* + past participle with wrong tense, and (c) wrong form. Type (a) answers were scored as 2, type (b) as 1, and type (c) as 0. As for type (a), no deductions were made for minor mistakes such as a misspelling (e.g., *If the bills [are not payed] in five days, the company will seek damages.*) or a third-person-singular error (e.g., *Office branches [is located] throughout the metropolitan area*).

There was a one-month interval between the pre- and post-tests. The same form and same order of the questions was used for the both tests. The reliability of the pre- and post-tests were acceptable (pre-test,  $\alpha = .72$ ; post-test,  $\alpha = .72$ ).

##### 4.2 Pre-Dictogloss Text

A pre-Dictogloss text was used as a pre-Dictogloss test employed before the treatment to assess each student's ability to conduct Dictogloss. The text was taken from *Crown* (Sanseido, 2007), an English textbook for second-year high school students. The pre-Dictogloss text comprised six sentences and the topic was Ayers Rock. The students were provided a sheet with a word list that would appear in the model English, and then instructed to individually reconstruct the model text after listening to it.

##### 4.3 Dictogloss Text

The four Dictogloss texts used in the treatment sessions were taken from *Grammar Dictation* (Wajnryb, 1990), which consists of three levels, pre-intermediate, intermediate, and advanced (p.

20). Four texts at an intermediate level were chosen for the study. The topics of the texts were natural phenomena, technology, job satisfaction, and housing and the law (See Appendix 1). The texts comprised five or six sentences. A CD was prepared as the listening material for the pre-Dictogloss and Dictogloss texts. In the recording, a female American English teacher read all the passages at a natural speed.

Qualities of students' reconstructed texts for the pre-Dictogloss test and Dictogloss 1, 2, 3, and 4 were assessed according to the same criteria. An accurate use of the passive voice (*be* + past participle) and perfect form (*has/have/had* + past participle) was scored as 1, and wrong forms were scored as 0. No deductions were made for a misspelling, third-person-singular errors, or tense errors in either passive voice or the perfect. The maximum score of the pre-Dictogloss text was 2 for the passive voice, and 2 for the perfect. The maximum scores of Dictogloss 1, 2, 3 and 4 were 4, 2, 1 and 3, respectively, for passive voice, and 1, 2, 2, and 4, respectively, for the perfect.

I removed the reconstructed texts written by triads and dyads containing learners who were not part of the study. Subsequently, 46 pre-Dictogloss texts and 68 reconstructed texts covering Dictogloss 1, 2, 3, and 4 were analyzed.

## 5. Procedure

The treatment session consisted of six steps. Figure 1 shows the overall procedure. At Step 1 in Figure 1, all students were given a sheet with a word list. The word list shows content words that appear in the Dictogloss text.

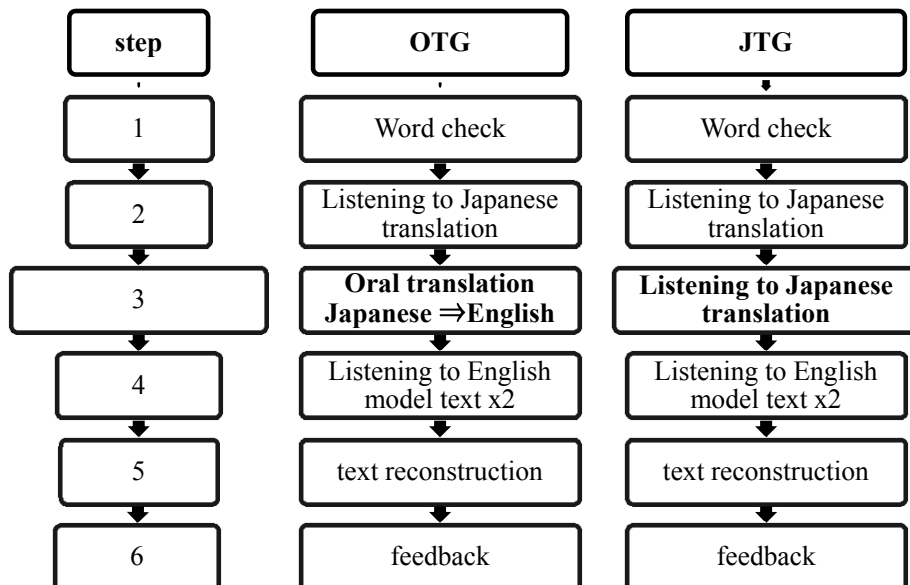


Figure 1. Two groups' steps of the treatment

After the students studied the words in class, they listened to a Japanese translation of the

model text at Step 2, as shown in Figure 1. Then, the two groups engaged in different activities. The OTG listened to the Japanese translation one more time and orally translated it from Japanese to English. On the other hand, the JTG listened to the Japanese translation again but were not asked to provide an oral translation. From here on, the two groups performed exactly the same activities. Students listened to an English text twice and then collaboratively reconstructed the text in pairs. During the reconstruction process at Step 5, the pair interactions were recorded. Afterwards, they checked their answers by comparing the model text provided in class by the teacher. The same treatment procedures were repeated in the following three treatment sessions. I arranged for every student to write the reconstructed text of the group in the two practice and four treatment sessions. For example, in the first and third treatment sessions, students sitting on the left side were assigned to write the text, and in the second and forth, ones sitting on the right side were assigned to do so. The treatment took between 30 and 40 minutes and was conducted once a week.

## 6. Language Related Episodes (LREs)

The recording of pair-interaction at the reconstruction stage in Dictogloss (Step 5 illustrated in Figure 1) was transcribed using Storch's (2002) transcription conventions (See Appendix 2). Swain and Lapkin (1995) define a language related episode (LRE) as follows:

any segment of the protocol in which a learner either spoke about a language problem he/she encountered while writing and solved it either correctly or incorrectly, or simply solved it (again, either correctly or incorrectly) without having explicitly identified it as a problem (p. 378).

Following their definition, I examined the transcripts and identified any learner utterances that indicated the learners encountered or recognized a linguistic problem. I then traced the verbal exchanges until I recognized a learner utterance that suggested they either solved their problem (successfully or unsuccessfully) or abandoned trying. The stretch of the exchanges from the identification of the problem to the end of the discussion is considered one LRE.

After coding all transcripts, I examined the target problem of LREs. There were three main problems: passive voice problem (PV- LRE), perfect form problem (PF- LRE), and other problem (O-LRE). O-LREs included issues related to determiners, word order, plural or singular, verb form except for passive voice or perfect form, tense, lexis, prepositions and linking words.

Example 1 shows a PV-LRE. In utterance 96, S1 confirmed that the learners should use passive voice. Then, in utterance 97, S2 agreed. They resolved the problem, thus ending the episode.

### Example 1

94 : S1: こっちは覚えてる。 / Some people are affected. /

(= I remember this part. / Some people are affected. /)

95 : S2: / some.../

96 : S1: / are affected- /、 やったよね？ここ、受け身やったよね？

( = / are affected - / Am I right? **This sentence was passive voice, wasn't it?**)

97 : S2: うんうん。 そうやった。

( = Yes, yes. That's right.)

(All the Japanese utterances are translated by the author)

Further, as I examined the dialogue, I found students referring to Japanese translation provided during the pre-task stage. Although their reference to Japanese translation is not exactly a linguistic problem, it appeared important evidence. This is because their reference indicates they paid attention to language in order to carry out the reconstruction task. Thus, I coded the students' reference to Japanese translation as a JT-LRE. Example 2 illustrates a JT-LRE. In utterance 172, S4 confirmed a Japanese translation. S3 told S4 Japanese translation and then they composed an English sentence, thus ending the episode.

#### Example 2

171: S3: sense of directions... ((a correct word is dissatisfaction))

172: S4: 日本語の意味、何だっけ？

( = **What was the Japanese translation?**)

173: S3: / 不満の...不満感 /、 やから、 多分/ sense of direction / かな、 って思っけんけど。

( = **/dissatisfaction.....dissatisfaction / , so, maybe / sense of direction /** , I thought.)

174: S4: / sense of.../かあ。

( = / sense of.../, I see.)

(All the Japanese utterances are translated by the author)

LREs may contain more than one LRE as in Example 3. This example shows a JT-LRE is embedded in a PF-LRE. These were independently coded and counted as one JT-LRE and one PF-LRE. In utterance 96, S3 asked whether the learners should use the past tense or the perfect; as a result, this episode was categorized into a PF-LRE. Further, in utterance 98, S3 uttered Japanese translation to make sure which form they should use. This is the trigger for a JT-LRE. In utterance 99, S4 said they should use “have” because the sentence includes a phrase “7 years.” In utterance 100, S5 reconstructed the sentence, thus ending the episode.

#### Example 3

94: S5: they...[主張する]

( = / they.../ /claim /.)

95: S6: [主張する]

( = /claim /.)

96: S5: / **claimed? they have claimed?** /

( = / **claimed? they have claimed?**)

97: S6: うん、 / they /

( = Yes, /they /.)

98: S5: 主張した？主張しました、って言ってたっけ？ ただの主張しました、やったら、 /  
claimed / で...いいと思うけど。

( = claimed? claimed? What did the model say? If the model said just “claimed ((past tense)),” I think we can use / claimed /.)

99: S6: 7年やったら、 / have / ちゃうかな？いるかな？

( = ((a phrase)) seven years requires / have /, I suppose. Should we use it ((have)) )?

100: S5: / have / いる？ 7年もの間、あ、でもそれは、7年もの間、頼んでいたと主張した。

( = Do you think we need / have / ? Seven years, for seven years, well.... but...it..., they claimed they have asked ... for seven years.)

(All the Japanese utterances are translated by the author)

The LRE categorization was checked for inter-coder reliability. Another experienced English teacher from the same vocational school coded 10% of the data. She received detailed instructions before coding. The inter-coder reliability was good ( $r = .97$ ). All disagreements were resolved through discussion.

#### IV. RESULTS

##### 1. Effectiveness in promoting EFL learners' acquisition of English passive voice constructions

In order to answer this question, (a) grammar test results and (b) the results of reconstructed texts that were employed as a pre-Dictogloss test and Dictogloss 1, 2, 3, and 4 were examined. First, grammar test results are scrutinized. The descriptive statistics for the grammar test are shown in Table 2. In order to confirm the homogeneity of the two groups' passive voice proficiency levels before the treatment, an independent  $t$ -test was conducted on the results of the pre-test. The  $t$ -test showed there was no significant difference at the .05 probability level between the OTG and JTG,  $t = -.37, p = .71$ .

**Table 2. Descriptive Statistics for the Grammar Test (maximum score: 20)**

	Pretest	Posttest
	Mean (SD)	Mean (SD)
OTG ( $n=27$ )	8.8 (4.6)	11.4 (5.1)
JTG ( $n=19$ )	9.4 (5.4)	9.5 (4.6)

Subsequently, an independent  $t$ -test was conducted on the each of the groups' gaps between the pre- and post-tests. Table 3 indicates the results of the  $t$ -test. This  $t$ -test compares the two groups' differences between the pre- and post-tests (Haebara, Ichikawa, & Shimoyama, 2001; Yoshida, 2006; Yamada, Sugisawa, & Murai, 2008). According to the  $t$ -test, there was a significant

difference,  $t = 2.34$ ,  $df = 43$ ,  $p < .001$ , and the effect size was between medium and large,  $d = 0.7$ . The results illustrate that the OT task more effectively promoted language learners' acquisition of passive voice, in comparison to the JT task.

**Table 3. The t-test on the Two Groups' Gaps between the Pre- and Post-tests**

	Mean ( <i>SD</i> )	<i>t</i> -value	<i>p</i> -value	Effect size ( <i>d</i> )
OTG ( $n = 27$ )	2.56 (2.4)	2.33	.001***	0.7(medium-large)
JTG ( $n = 19$ )	0.16 (4.6)			

\*\*\* $p < .001$

Second, I examined the results of reconstructed texts that were employed as the pre-Dictogloss test, and Dictogloss 1, 2, 3, and 4. Because the maximum scores for the pre-Dictogloss text and Dictogloss 1, 2, 3, and 4 texts were different, each text was scrutinized on the basis of what percentage was scored. On the pre-Dictogloss test, the passive voice and perfect form each had a maximum score of 2. In the passive voice results for the pre-Dictogloss test, the OTG scored 69 %, and the JTG, 63 %. These results indicate there was no significant difference between the OTG and JTG on passive voice. In the perfect form results for the pre-Dictogloss test, the OTG scored 59 %, and the JTG, 71 %. The results show that the JTG scored better than the OTG on perfect form. The results of the pre-Dictogloss test are shown in Table 4.

**Table 4. The Results of Reconstructed Texts for Pre-Dictogloss Test**

Passive Voice	OTG ( $n = 27$ )	Score ratio* <i>SD</i>	37/54 (69%) 0.69
	JTG ( $n = 19$ )	Score ratio <i>SD</i>	24/38 (63%) 1.3
Perfect Form	OTG ( $n = 27$ )	Score ratio <i>SD</i>	32/54 (59%) 1.2
	JTG ( $n = 19$ )	Score ratio <i>SD</i>	27/38 (71%) 1.4

*Note.* *SD*: standard deviation; ratio\*= score / maximum

Subsequently, the results of the Dictogloss 1, 2, 3, and 4 were examined. In each treatment, the number of students who attended the class varied. In addition, the maximum scores for passive voice and perfect form for Dictogloss 1, 2, 3, and 4 changed. Hence, the maximum scores of each treatment for the OTG and JTG were determined by multiplying the number of dyads who attended the treatment session by the number of questions for passive voice and perfect form, respectively. Then, each group's text was scrutinized on the basis of what percentage was scored. After the maximum scores covering Dictogloss 1, 2, 3 and 4 were calculated, each group's overall percentage

of correct answers covering Dictogloss1, 2, 3 and 4 were tallied for passive voice and perfect form respectively. Table 5 illustrates the results of reconstructed texts.

**Table 5. The Results of Reconstructed Texts for Dictogloss 1,2,3 and 4**

			Dictogloss1	Dictogloss2	Dictogloss3	Dictogloss4	Total
Passive Voice	O T G	Score	37/44	18/18	8/10	19/30	82/102
		ratio*	(84%)	(100%)	(80%)	(63%)	(80%)
		SD	0.98	0.00	0.60	0.83	
		n**	11	9	10	10	40
	J T G	Score	14/24	11/12	4/8	20/24	49/68
		(ratio)	(58%)	(92%)	(50%)	(83%)	(72%)
		SD	0.75	0.37	0.50	0.71	
		n	6	6	8	8	28
Perfect Form	O T G	Score	9/11	17/18	18/20	38/40	82/89
		(ratio)	(82%)	(94%)	(90%)	(95%)	(92%)
		SD	0.39	0.31	0.46	0.4	
		n	11	9	10	10	40
	J T G	Score	4/6	8/12	13/16	24/32	49/66
		(ratio)	(67%)	(67%)	(81%)	(75%)	(72%)
		SD	0.47	0.87	0.48	1	
		n	6	6	8	8	28

*Note.* ratio\*= percentage of correct answers; n\*\*= number of dyads

As for the results of passive voice, the OTG scored 80 %, and the JTG, 72 %. Regarding perfect form, the OTG scored 92 %, and the JTG, 72 %. These results indicate the OTG performed better than the JTG on both passive voice and the perfect.

These results obtained from the grammar test and reconstructed texts indicate that the OT task has a tendency to promote learners' acquisition of English passive voice in comparison to the JT task.

## 2. What the students notice and discuss during the reconstruction process

In order to answer this question, the percentages of the four types of LREs produced by the OTG and JTG were examined. Among 420 LREs identified among the OTG's pair discussions, 67 (16%) were discussion on perfect form while 25 (6%) were on passive voice. They also made 63 LREs (15 %) in reference to Japanese translation models provided in the pre-task. In other words, the OTG focused more on perfect form than the other target form, passive voice, during the reconstruction process in Dictogloss.

On the other hand, among the 360 LREs that the JTG produced, 106 (30 %) LREs were related to Japanese translation models provided during the pre-task stage. Approximately similar numbers of LREs (33 or 11 % and 27 or 9 %) were produced in reference to the target forms, perfect form and passive voice, respectively.

Other LREs (O-LREs) include issues related to determiners, word order, plural or singular,

verb form except for passive voice or perfect form, tense, lexis, and prepositions and linking words, as mentioned earlier. O-LREs were 63 percent of the total produced by the OTG and 46 percent of those produced by the JTG. The results are juxtaposed in Table 6.

**Table 6. Ratio (%) and Sum of Four LREs in Each Group**

	Passive voice (PV-LRE)	Perfect form (PF-LRE)	Japanese translation (JT-LRE)	Other (O-LRE)	Total
	Ratio (Sum)	Ratio (Sum)	Ratio (Sum)	Ratio (Sum)	Ratio (Sum)
OTG ( <i>n</i> *=40)	6% (25)	16% (67)	15% (63)	63% (265)	100% (420)
JTG ( <i>n</i> *=28)	9% (27)	11% (33)	34% (106)	46% (140)	100% (306)

*Note.* *n*\*= total number of dyad's interactions covering Dictogloss1, 2, 3 and 4

Careful examination of O-LREs revealed that some students in OTG alone made interesting utterances. As in Example 4 I will present later, the OTG participants appeared to make a linguistic comparison between what they thought the language should be and what they thought they heard in the English model in the Dictogloss text. Since learners' verbal protocols during the reconstruction process are evidence of learners' attention, I identified such a segment as a "comparison episode." A comparison episode starts when a student indicates he/she felt discrepancies between what he/she thought the form was (during oral translation in the pre-task) and what he/she heard in the English model of the Dictogloss. The comparison episodes are embedded in LREs dealing with a linguistic problem such as verb form, prepositions, and determiners.

Example 4 illustrates a comparison episode. In utterance 26, S8 says she thought "why?," which represents that she noticed a discrepancy between her IL and the model English by comparing them.

Example 4 (O-LRE: verb form)

23 : S7: でも、 / bark / ?

( = However, was it / bark / ? )

24 : S8: そうやねん。 でも、ここは / bark / しか言わへんかってん。

( = Yeah, it was ((weird)), but the model said only / bark / . )

25 : S7: あっ、 そうなん？

( = Oh, really? )

26 : S8: うん、 なんでやろって思って、 わからんかって。

( = Yes. **I thought "Why ?,"** but I didn't have any idea. )

(All the Japanese utterances are translated by the author)

Example 5 presents another comparison episode. In utterance 196, S14 explained that during the second listening pass, she carefully paid attention to places where she had missed words on the



first attempt. However, she thought the model English said still / at /. This is indicative that she found a difference between her IL and the incoming English on the first attempt by making a comparison. Furthermore, it demonstrates how she carefully paid attention to the incoming speech on the second attempt in order to check her hypothesis or IL.

Example 5 (O-LRE: preposition)

189 : S13: 聞き取れた？

( = Were you able to catch the word? )

190 : S14: え、なんか、全然聞こえんくて、 / a / と、 / at / って聞こえて。

(= Well, I couldn't catch it at all, just... / a / or / at / ...)

191 : S13: / at? /

192 : S14: そんなん、そんなわけ、ないやんか？ / 加えて？ / おかしいやん。

( = That can't be it, don't you think so ? / Additionally? / It sounds weird. )

193 : S13: わからん。

( = I have no idea. )

194 : S14: え？と思って。2 回、聞いたやん？

( = I thought, "What?" We listened twice, right? )

195 : S13: うん。

( = Yes. )

196 : S14: 1 回目で聞こえへんかったから、2 回目、ちょっと注意しとってんけど、どう聞いても / at / って聞こえて。

( = **I missed catching it on the first attempt, so the second time I carefully listened to it ((to identify the word ))**, but it still sounded / at / . )

(All the Japanese utterances are translated by the author)

These results show that (a) the reconstruction process of the Dictogloss task did not encourage the OTG to discuss target forms as much as expected, (b) during the reconstruction process, what the learners talked about the most were language problems other than target forms, such as plural or singular, prepositions and determiners, (c) the OTG talked about perfect form more frequently than passive voice, (d) the JTG tried to recall the meaning of Japanese translation during the reconstruction process, and (e) during the reconstruction process, five OTG learners compared their IL and the model English in Dictogloss, and discussed the discrepancy between their IL and the model English.

## V. DISCUSSION

The findings for the two research questions reveal that the OT task is as an effective pre-task for *acquisition* of passive voice when it is conducted before Dictogloss in comparison to the JT task. As I mentioned earlier, *acquisition* in the present study is defined as an increase in control over forms that have already been internalized (Ellis, 1997; Bialystok & Sharwood Smith, 1985; de Bot,

1996; Lyster, 2004).

The findings in this study support the output hypothesis (Swain, 1995, 2000; Kowal & Swain, 1994, 1997; Lapkin & Swain, 2000; Swain & Lapkin, 1995, 1998, 2001) and noticing hypothesis (Rutherford & Sharwood Smith, 1985; Schmidt, 1990, 1993, 1995). As I presented in the result section, the OTG gained better command of English passive voice after the OT task, an oral translation pre-task and Dictogloss. The OT task, an output-oriented task, provided the OTG with opportunities for noticing language features several times. According to the output hypothesis, one of the functions of output is the noticing/ triggering role, which facilitates the second language acquisition.

The noticing hypothesis states that learners' attention to the specific language form is necessary for learning; therefore it was expected that frequent PV-LREs would be observed if the OTG participants had learned passive voice. The grammar test results indicated the OTG's better performance on passive voice. Yet, the OTG participants did not produce as frequent PV-LREs as expected. These contradictory findings could be explained when we consider the task procedures and the data collection process in this study. The OTG participants were actually given two output opportunities; thus two noticing opportunities. Being an output task, the OT task in the pre-task stage could have required the students to pay attention to language before they actually heard the English model in Dictogloss. Their attention to linguistic forms during this pre-task stage could have enhanced their understanding of the forms at the main task stage, Dictogloss.

As I presented in Example 4 in the result section, some OTG participants made comparison between what they thought the form should be (based on their own English translation attempt based on Japanese cues) and what they heard in the model passage during Dictogloss. If the form is relatively easy and salient to process, they could have successfully learned the form during the oral translation task stage before Dictogloss. Thus, they did not have to discuss the form with the partner during the reconstruction stage at which LREs were collected. There is a possibility that the OTG resolved many passive voice problems at the oral translation pre-task stage and students did not have to produce PV-LREs. This data collection process might have resulted in the contradictory findings, infrequent PV-LREs but the better outcome in the grammar test by the OTG. Of course, a future research is necessary to confirm this hypothesis by collecting data during the oral translation pre-task.

Although the OTG made infrequent PV-LREs, it did not mean that they generated infrequent LREs in general; in fact, they produced as many as 67 PF-LREs (16%) in contrast to PV-LREs (25 or 6%). The results may be closely related to the Japanese linguistic feature, more specifically, which language form is salient from the standpoint of Japanese. As I have mentioned earlier, passive voice could have been salient to the students because the same feature exists in the Japanese language. Previous studies showed (Leow, 1997; Lyster, 2004; Park, 2010) that salient language form or feature is more likely to attract a foreign language learner's attention than less salient one; thus, the OTG learners could have paid attention to passive voice in the earlier stage of the whole task and produced infrequent LREs. On the other hand, perfect form does not exist in Japanese

grammar and is a difficult tense concept for Japanese EFL learners. The OTG members similarly had the opportunities to notice linguistic problems in using perfect form at the oral translation stage and to notice the appropriate use of perfect form in the English model in Dictogloss. Unlike passive voice, the students discussed the perfect form problems during the reconstruction stage of Dictogloss. The different degrees of the OTG learners' discussion on the two target forms could have stemmed from the level of difficulties. It means the salience of the linguistic features from the view of Japanese.

Comparison episodes in which a student compared his/her own attempt of translation with the incoming model speech were found only in the data of the OTG dyads. These episodes show that the reconstruction process in the OT task provided these OTG members with opportunities to reflect on their own attempt and notice the difference between their IL and the model English. Student 8 in Example 4 mentioned that she thought "Why?" after listening to English (line 26). Her utterance illustrates that Student 8 reflected on her translation attempt and noticed a discrepancy between her IL and model English by comparing them. In other words, learners' discussion during the reconstruction process in Dictogloss is influenced by the linguistic resources that learners have. Because the OTG students were given opportunities to think of English translation based on Japanese cues, the OTG paid careful attention to English input during Dictogloss. Student 14 in Example 5 clearly claimed that she listened to the English passage even more carefully at the second time (line 196). The OT task provided the OTG students with L1 cues and incomplete L2 as linguistic resources before they listened to L2 input in Dictogloss.

On the other hand, LREs produced by the JTG students revealed that Japanese cues provided during the pre-task made them rely heavily on meaning comprehension; subsequently they tried to translate the L1 meaning into English during the reconstruction process. Because the JTG participants were given only L1 meaning as resource, it is likely that they did not make best use of L2 input provided in Dictogloss.

The entire set of findings suggests that the OT task can favorably enhance the effects of Dictogloss, the main task of the present study. Dictogloss itself promotes learners' noticing of language forms and incoming English in the Dictogloss text. Considering the characteristics of Dictogloss, the OT task, an output-oriented pre-task, seems to increase the efficacy of Dictogloss.

## VI. CONCLUSION

This study explored the impacts of two cases of tasks, OT task and JT task, on acquisition of passive voice and learners' attention. The findings suggest that compared with the JT task, the OT task has a tendency to promote learners' acquisition of English passive voice. In addition, an output-oriented pre-task and an input-processing based pre-task function differently in terms of manipulating learners' attention during the task. In this regard, I believe that my study sheds light on the importance of a pre-task. A well-thought-out pre-output task can promote the efficacy of the main task.

This research contains several limitations. The first limitation was the participants' L1. Even

though Japanese translation was used in the task, Japanese was not the L1 for some participants. There was no choice but to include non-Japanese participants in order to keep appropriate sample sizes. The non-Japanese participants' Japanese proficiency was proven to be relatively high, but Japanese was their L2, not L1. Therefore, this issue should be addressed in future research. Another issue concerns learner variables such as personality or learning style. In terms of LREs, previous research has examined the effects of personality, for instance, dominant, passive, or strong (Storch, 2001, 2002). Park (2010) reports that if one member of a dyad is form-oriented while the other is meaning-oriented, and the former dominates the latter, the meaning-oriented learner might produce more LREs than he/she normally would. Future studies should eliminate these personal variables as much as possible. In addition, future studies should put a longer interval between pre- and post-tests in order to eliminate the possibility of a practice-effect variable.

I hope that the findings and implications of the current study will be beneficial to English teachers as well as EFL learners who are searching for an effective activity for acquiring English passive voice.

## ACKNOWLEDGEMENT

I would like to express my gratitude to anonymous reviewers for their useful comments.

## NOTE

This article is based on the author's unpublished master thesis submitted to the Graduate School of Foreign Language Education and Research, Kansai University, Osaka, Japan.

## REFERENCES

- Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind. *Applied Linguistics*, 6, 101-117.
- de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46, 529-555.
- Doughty, C. & Williams, J. (1998). Issues and terminology. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 1-11). Cambridge: Cambridge University Press.
- Ellis, R. (1997). Investigating form-focused instruction. In R. Ellis, (Eds.), *Form-focused instruction and second language learning* (pp. 1-46). Malden, MA: Blackwell.
- Haebara, A., Ichikawa, S., & Shimoyama, H. (2001). *Shinnrigaku kenkyu Nyumon Cyosa: Jikken kara Jissen made*. [Introductory research of psychology: From experiment to practice]. Tokyo: Tokyodaigakushuppan.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition*, 24, 541-577.
- Izumi, S. (2009). *Focus on form wo toriireta atarashii eigo kyouiku*. [New English education with focus on form]. Tokyo: Taisyuukansyoten.

- Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness*, 3, 73-93.
- Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote it in the immersion classroom? In Johnson, K. & Swain, M (Eds.), *Immersion education: International perspectives* (pp. 284-309). Cambridge: Cambridge University Press.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Lapkin, S. & Swain, M. (2000). Task outcomes: A focus on immersion students' use of pronominal verbs in their writing. *Canadian Journal of Applied Linguistics*, 3, 7-22.
- Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal*, 86, 485-507.
- Leow, R.P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47, 467-505.
- Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied linguistics*, 4, 126-141.
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K.de Bot, R. Ginsberg, & Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.
- Long, M. & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press,
- Lougheed, L. (2008). *600 Essential words for the TOEIC*. New York: Barron's Educational Series, Inc.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399-432.
- Murphy, R. (2009). *Grammar in use intermediate*. Cambridge, USA: Cambridge University Press.
- Nabei, T. (1996). Dictogloss: Is it an effective language learning task? *Working Papers in Educational Linguistics*, 12, 59-74.
- Ortega, L. (1999). Planning on focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109-148.
- Park, S. (2010). The influence of pre-task instructions and pre-task planning on focus on form during Korean EFL task-based interaction. *Language Teaching Research*, 14, 9-26.
- Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, 12, 61-82.
- Richards, J. (2002). Addressing the grammar gap in task work. In J.C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 153-166). Cambridge: Cambridge University Press.
- Rutherford, W. & Sharwood Smith, M. (1985). Consciousness-raising and universal grammar. *Applied Linguistics*, 6, 274-82.
- Sanseido. (2007). *Crown*. Tokyo: Author.

- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Reviews of Applied Linguistics*, 13, 206-226.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In Schmidt, R. (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Second Language Teaching and Curriculum Center, University of Hawai.
- Shak, J. (2006). Children using dictogloss to focus on form. *Reflections on English Language Teaching*, 5, 47-62.
- Song, MJ., & Suh, BR. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, 36, 295-312.
- Storch, N. (1998). Comparing second language learners' attention to form across tasks. *Language Awareness*, 7, 176-191.
- Storch, N. (2001). How collaborative is pair work? ESL tertiary students composing in pairs. *Language Teaching Research*, 5, 29-53.
- Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52, 119-158.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and Practice in Applied Linguistics: Studies in honour of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Eds.), *Sociocultural Theory and Second Language Learning* (pp. 97-114). Oxford: Oxford University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371-391.
- VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.) *Theories in second language acquisition: An introduction* (pp. 115-135). Mahwah, NJ: Lawrence Erlbaum.
- Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.
- Williams, J. (2005). Form-focused instruction. In E. Hinkel (Eds.) *Handbook of research in second language teaching and learning* (pp. 671 -691). Mahwah, NJ: Lawrence Erlbaum.
- Yamada, T., Sugisawa, T., & Murai, J. (2008). *R niyori yasashii toukeigaku [Simple statistics by use of R]*. Tokyo: Ohmsha.
- Yoshida, T. (2006). Kennkyuuhouhou nitsuiteno gakusyuu to kyouiku no arikatanitsuite omoukoto, arekore in *Shinrigaku kenkyuuhou no atarashii katachi [New approaches to research in psychology]* (pp. 244-270). Tokyo: Seishinsyobou.

### Appendix 1: A Text Used in Dictogloss 2

1. The Earth is only a tiny part of an endless universe.
2. Many secrets in the universe are waiting to be uncovered.
3. Over the last twenty years, enormous amounts of money have been invested in space exploration.
4. This policy has been criticized by some people since the beginning of the space race.
5. Certainly it is impossible to halt the race.
6. However, this issue should be fully debated.

### Appendix 2: Transcription conventions

( )	nonverbal sound, e.g., (laugh)
(= )	English translation of Japanese expressions
...(multiple periods)	short pause, between 0.5 and 3 seconds
xxx	words/ phrases difficult to decipher
italics	word/s pronounced with emphasis
[	beginning of simultaneous/overlapping talk (end of overlapping talk is indicated by )]
!	an exclamation mark denotes a sharp rise at the end of a word or phrase
?	a question mark denotes a sharp rise at the end of a word or phrase
“ ”	quotation marks denote that the participant is reading the given text (e.g., instructions, text given in the editing task)
Wor- w-o-r	only part of a word is pronounced; i.e., utterances are cut off or unfinished the speaker is spelling out the word
(( ))	comments made by the researcher to describe other phenomena, e.g., ((writing))

## **Using the Novelization of the Movie, *Dead Poets Society* in an EFL Reading Course**

Cheena Fujioka

University of Marketing and Distribution Sciences

### **ABSTRACT**

This article describes the use of the novelization and tasks which were developed for a reading EFL course at the university level. For this purpose, the novelization of the American movie, *Dead Poets Society* (Kleinbaum, 1989), was utilized for the academic year. Along with viewing the movie, various types of graphic organizers were used for assigned tasks in order to assist the students' comprehension of the content of the story and the characters and to encourage empathic reactions and critical thinking skills of the students. This authentic literary text was used throughout the year to illustrate a sense of achievement as well as to arouse a sense of community among the students and the teacher through sharing their personal emotional reactions to the plot, characters, and the poems which appeared in the story. The class survey that was conducted at the end of the year resulted in positive feedback for the method of using a novelization with the movie for enhancing empathy and critical thinking skills among the students.

**Key words:** novelization of a movie, content-based literature instruction, poetry, graphic organizers

### **I. INTRODUCTION**

Traditionally in Japan, English had mostly been taught with Western classical or modern literature at the university level. Yet, up to the 1990s, English education shifted to a more practical, communication-oriented one in an effort to respond to communication needs in the real world. However, in recent years, student motivation for learning English seems to be declining in many EFL classrooms in Japan. Japanese educators in higher education remain challenged to develop materials and methods which strive to motivate their students. In such a difficult time, reviving the use of literature (in content-based EFL) can offer motivating, effective materials for teaching Japanese students, along with the implementation of extensive reading that has been recognized and is expanding nation-wide. Despite some arguments against its pedagogical value, the gradual restoration in teaching L2 with literary texts has been evident in more recent years (Carter, 2007; Hismanoglu, 2005; Maley, 2001). This is due to the characteristics of literature providing L2 learners with "valuable authentic material," "cultural enrichment," "language enrichment," and "personal involvement" (Collie & Slater, 1990, pp. 3-6), as long as the literary text is well-selected, relevant and is based on L2 learners' level of English



proficiency. Hence, in this paper, literature-based instruction using the novelization of an American movie, *Dead Poets Society* (Kleinbaum, 1989), is examined as an effective teaching medium, hoping to raise L2 learners' empathy and critical thinking skills as well as overall English proficiency in a university EFL classroom.

## II. LITERATURE REVIEW

### 2.1 Three models of teaching literature

According to Carter & Long, Lazar (as cited in Clandfield, n.d., p. 2), there are the following three models that can be used to teach a literary text in the language classroom:

#### *The cultural model*

views a literary text as a product. This means that it is treated as a source of information about the target culture. It is the most traditional approach, often used in university courses on literature. The cultural model will examine the social, political and historical background to a text, literary movements and genres. There is no specific language work done on a text. This approach tends to be a quite teacher-centered. (p. 2)

#### *The language model*

aims to be more learner-centered. As learners proceed through a text, they pay attention to the way language is used. They come to grips with the meaning and increase their general awareness of English. Within this model of studying literature, the teacher can choose to focus on general grammar and vocabulary (in the same way that these are presented in course books for example) or use stylistic analysis. Stylistic analysis involves the close study of the linguistic features of the text to enable students to make meaningful interpretations of the text—it aims to help learners read and study literature more competently. (p. 2)

#### *The personal growth model*

is also a process-based approach and tries to be more learner-centered. This model encourages learners to draw on their own opinions, feelings and personal experiences. It aims for interactions between the text and the reader in English, helping make the language more memorable. Learners are encouraged to “make the text their own.” This model recognizes the immense power that literature can have to move people and attempts to use that in the classroom. (p. 2)

Regarding the reading course that is going to be described below, it seems to fit the personal growth approach the most as it focuses on connecting L2 learners and the text emotionally. In addition to this, there seems to be a little element of the language model, while paying some attention to the complex language and literary expressions such as metaphor in the text.

## **2.2 Why incorporate a movie into a novel?**

In this course, a novelization is utilized as the course material. According to Longman's *Advanced American Dictionary* (2000), a novelization is defined as "a story that was originally written as a movie or television program before being written as a book" (p. 991). Brown (2005), however, warns educators of the quality of novelization, compared to novels that are later made into movies. She claims that novelization generally has limited quality as a literary work, even though they usually follow the plot and add detail. Nevertheless, reading an intriguing story along with viewing the original movie is assumed to be an enjoyable and motivating educational experience for L2 learners, particularly when the movie is well made and highly acclaimed like the one used in this course. As for incorporating visual media such as movies into novels, Kasper (2000) notes that these movies are particularly significant for lower level L2 learners as "they present graphic, visual illustrations of key critical thinking concepts, thus helping to consolidate and reinforce learning by making the subject matter more concrete" (p. 111). Furthermore, they challenge L2 learners "to match images elicited from the printed texts with visual representations on the screen, further engaging them in critical thinking tasks" (Kasper & Singer as cited in Kasper, 2000, p. 111). With the use of the movie, it is hoped that lower level students in particular can keep up with the novel and would feel less anxious about the complexity of the language in it.

## **2.3 Why use graphic organizers?**

An important educational tool used by the author was the graphic organizers such as the chapter organizer (see Appendix A). They were continuously used in conjunction with the reading assignments throughout the course. In the U. S., graphic organizers are often seen and used in education at all grade levels including college and its textbooks. According to Lane (1999), "the term graphic organizers refers to a body of graphic representations of information" and they may take the forms of "graphs, diagrams, or illustrations; however, graphic organizers are commonly left blank for students to fill in" (p. 19). One typical example of a graphic organizer is a Venn diagram, which consists of two overlapping circles in the middle to show the differences and similarities (overlapping-middle part) between two people or topics. McKnight (2010) notes that they are significant and effective pedagogical tools since they can organize content and ideas and also facilitate learners' understanding of the information they have newly acquired. Furthermore, by presenting material through the visual and spatial modes and reinforcing what they are taught in class, using these organizers helps students internalize their learning (McKnight, 2010). Dillon and Johnson et al. (as cited in Lane, 1999) regard graphic organizers as useful devices in terms of enhancing cognitive competence of students by assisting them to comprehend, absorb, and reveal their ideas. In other words, through the use of graphic organizers, teachers can assist their students to develop their critical thinking skills as well. Based on the author's extensive teaching experiences in reading, using graphic organizers appears to be an effective teaching strategy for students at all proficiency levels. Lane (1999) points out, by developing students' critical thinking skills, "the

general proficiency levels, performance, and motivation of students can be improved” (p. 19). When introducing these graphic organizers to students, it is necessary that teachers first demonstrate how they work by showing an example or initiate one with students on the board, letting them finish it on their own or in pairs/groups.

### III. COURSE DESCRIPTION

#### 3.1 Students

The students were sophomores majoring in sociology and psychology at a private university in Osaka, Japan. Two sections of a reading course were involved in two continuous semesters (spring and fall) of the academic year 2010. The class met once a week for 90 minutes. In one of the classes, 42 sociology majors enrolled in the spring semester and 29 in the fall semester. In the other, 44 psychology majors enrolled in the spring semester and 30 in the fall semester. These student numbers include several students who dropped out from each class during the semesters. The students were evaluated based on: attendance (20%); participation/classroom behavior (20%); pop quizzes/skit (20%); homework assignments (20%); final exam (20%).

#### 3.2 Course material and its purpose

For this course, the unabridged novelization of the American movie, *Dead Poets Society* (Kleinbaum, 1989), was used throughout the academic year. The story is set in a rigid, college preparatory, all-male boarding school in Vermont, U. S. A. in the late-1950s, where a new English teacher, Mr. Keating, totally changes his students’ lives with his message of *Carpe Diem* (“seize the day” in Latin, which actually means “live life to the fullest”). The novel consists of 15 chapters and is 166-page long. According to Yohan (n.d.), this title is suitable for L2 learners who have 470 on the TOEIC or the pre-second grade of the EIKEN test (in practical English proficiency) in Japan.

While reading the story, a portion of the original film, *Dead Poets Society* (Schulman & Weir, 1989), was shown twice each semester (totally four times to show the entire movie in the academic year) to help students comprehend the content and to develop empathy. The Japanese title of the movie is “Ima wo Ikiru,” which literally means “live this moment.” The movie starring Robin Williams as the charismatic English teacher, Mr. Keating, won the Academy award for the best script with several more nominations such as the best film and best actor in 1989.

The reasons for selecting this particular novelization are: (1) since the characters are close to the students’ age, they are more likely to empathize with the characters, setting, and plot; (2) it is full of words/dialogs (including the excerpts from the poems written by some of the most influential poets such as Walt Whitman and Robert Frost) that are touching and thought-provoking; (3) it has strong, positive messages to readers; and (4) the original movie earned the best script for that year’s Academy Awards. Therefore, though the movie was released more than two decades ago, the story is still expected to appeal to students today due to its universal themes: *Carpe Diem* and friendship.

The following lessons aimed: (1) to enhance students’ empathy, critical thinking skills, and

overall reading skills, while appreciating authentic expressions and dialogs as well as poetry used in the story; (2) to help them think deeply of the significance of Mr. Keating's lesson, *Carpe Diem*, and of what learning really is about; and (3) to appreciate and find empathy with the characters who are students in the story. Due to the limited English proficiency of the students and the complexity of the text and poems, the instructional language was both English and Japanese. However, Japanese was always used to explain complicated things such as the poems, setting, characters, and cultural background of the story. Students were encouraged to write in English in their worksheets/graphic organizers as much as possible except a few organizers which allowed them to write in Japanese.

### 3.3 Teaching procedure

#### 3.3.1 Course design

The lessons using the continuous two semesters were conducted in the following manner:

1. Before each class meeting, *assign students to read one/two chapters* of the novel that would be covered in class, and to fill in the table of the *chapter organizer* (see Appendix A) which consists of the characters, setting, and plot for each chapter. In order to check the reading assignments, give them three pop quizzes on the content of the material using wh-questions each semester.
2. With emotion, *read the story aloud to the students* and also ask for some volunteers or call on students to read it aloud. Although it takes time to do so from each lesson, reading aloud should be included for low or intermediate level L2 learners, in particular. This is because reading the story aloud to students is crucial as it incorporates listening, an essential component of language acquisition (Seda, Liguori, & Seda, 1999). Richard-Amoto (1996) has pointed out that it is often beneficial for students to hear the portion of the story read aloud by the teacher or others. She contends that students need to hear the intonation, the pauses, the rhythm, and the pronunciation of the words. Interestingly, Krashen (1997) claims the significance of being read aloud: "Teachers should also be adept at reading aloud to students; even older students can get 'hooked on books' when they hear part of the story read to them" (p. 26).
3. Every time that a few pages of the story are read aloud, first *check students' reading comprehension* by asking some factual questions regarding the setting, characters, and plot. Then, ask a few inferential questions to check their interpretations of the content. Additionally, discuss the content, characters, and cultural background with the class, while welcoming their personal reactions (as well as those of the teacher) to them in order to deepen their understanding and to encourage empathy.
4. After finishing each chapter, call on students and let them *share what they wrote in the chapter organizer*, which was assigned as homework in the previous week. Then, write a sample of the plot on the blackboard so that they can compare to theirs.
5. Begin each class with *having a student summarize orally what they read in the previous week*,

so that they can recall and reinforce what they read and learned, and get ready for the new development of the story. Add some key points in the summary if the student fails to cover them.

6. Hand out the Japanese translations of all the poems which appeared in the story. To analyze and discuss poetry in class over the two semesters, select nine excerpts of the poems which are particularly inspiring for students such as Robert Herrick's "To the Virgins, Make Much of Time", Walt Whitman's "O Me! O Life!" and Alfred Lord Tennyson's "Ulysses". These excerpts are crucial throughout the story as they "are used almost like biblical scriptures to encourage, to motivate, to comfort, to teach and to give directions on coming to terms with oneself and others" (Grunert, 2008, p. 4). Due to the main character, Mr. Keating's passion and enthusiasm for teaching poetry as well as the plot constantly dealing with the problems and growing pains of adolescents, students are more likely to become interested in the power of poetry (Grunert, 2008). Have each student read them aloud and *interpret the content in pairs/group* before discussing as the whole class.
7. Before submitting the homework assignments described below, *have them exchange their works*, share them in pairs, and write their comments on the worksheets. It is educationally beneficial for them to see different perspectives of their peers and to share theirs with others since they often get themselves stimulated by doing so.
8. *Show a portion of the movie* after reading that part of the novel so that they can view the entire film in one year (two viewings each semester). Show the movie with the English subtitles so that students can recognize the identical dialogs as the novel, spoken by the native English speakers and enhance their listening skills as well.
9. Toward the end of the fall semester, divide them into several groups of 5-6 people and let them *role-play a scene/scenes* where they like to make a skit in English from the novel for about 10 minutes. Voting for the MVP for the best group in class might be motivating. When announcing the class award in the following week, read aloud some comments from peers.
10. When finishing reading the story and watching the last quarter of the movie, have the *final discussion* regarding the movie, characters, and their feedback as the whole class as long as time is permitted.

### 3.3.2 Tasks

For enhancing students' comprehension of the story and empathy toward the characters and/or the story, the following tasks were developed by the instructor. Due to the limited time of the class, these tasks using different graphic organizers except #9 were assigned as out-of-class work in the following order:

1. *Chapter organizer* (see Appendix A)

Aims: to enhance comprehension, skimming, scanning, summarizing, writing

Students organize each chapter by filling in the characters, setting, and the plot in the table. In

order to demonstrate how to do the chapter organizer, together as a class, the instructor writes the information from the first chapter on the board. Assign them to do the rest of the chapters as the weekly assignment so that they can prepare for each lesson.

2. *My favorite teacher*

Aims: to develop empathy, critical thinking skills, writing

Students choose one unforgettable teacher whom they have encountered in their lives. Each student writes the following information about that teacher in the organizer: where and when the student met this teacher; what subject he/she taught; and what made him/her so special.

3. *Character chart*

Aims: to enhance comprehension, scanning, summarizing

Students organize the information of all the main 16 characters in the table. Next to each character's name, enter the page on which he/she appears first in the story and some brief information on whom he/she is in the story including his/her job and characteristics.

4. *Character relationship chart*

Aims: to enhance comprehension, scanning, critical thinking skills

As a bonus-point-task, students draw the chart of the character relationship using all the main 16 characters and labeling each character to show how each character is related to one another.

5. *Character emotional development* (see Appendix B)

Aims: to enhance comprehension, scanning, writing, and empathy

Students select two characters among Mr. Keating, Neil, Todd, Knox, Charlie, Cameron, Meeks, and Pitts that they are interested in. Draw two line graphs in different colors showing the two characters' emotional changes in each chapter (cover chapter 1-8 for spring semester and cover chapter 9-15 for fall semester). The vertical axis shows the characters' emotional changes while the horizontal axis indicates the chapters. Put two graphs in one chart to show the emotional differences in two characters, and write down the reasons briefly for each chapter in the graphs. Students may use a computer to draw the graphs.

6. *Impressive quotes from the novel*

Aims: to enhance comprehension, writing, critical thinking skills, and empathy

Since this story is full of words/lines that are touching, have them choose their favorite quotes once each semester. The organizer is divided into five columns: page number, who said, their favorite quotes (each student quotes them), how this made them feel, and buddy's comment (from a classmate who read it).

7. *Casting Dead Poets Society*

Aims: to develop critical thinking skills and to enhance empathy, personalization

Students pretend to be the executive producer of the Japanese movie/drama, *Dead Poets Society*. Each student picks suitable actors/actresses for the characters and writes down the specific reasons for the choices in the table. The soundtrack music to be used in the movie/drama should also be chosen.

8. *Poetry Response* (see Appendix C) (partially adapted from McCloskey, M. L., & Stack, L. 1996, p. 135)

Aims: to develop analytical skills, critical thinking skills and to enhance empathy

Students select their favorite poem used in this story and interpret it in Japanese as it would be too complex for them to discuss poetry in English. Each student writes her/his personal response in the following order in the organizer: the page number where the poem used in the story; the name of the poet; the title of the poem (if available); the impressive lines from the poem; what the student liked about this poem; what she/he thought about: the words, the subject and the way the words sound; what the poem reminds her/him of; and what the poem makes her/him think about.

9. *Comparing and contrasting the movie and the novel*

Aims: to enhance listening, empathy, and to develop critical thinking skills

While watching the movie, let them write the elements that were close to the original story on the left side of the table and those that were different on the right side of the table. They decide how many stars they would give to the movie and write some comments on the movie in the table. This can later lead to the whole class discussion sharing students' perspectives of the movie.

10. *Skits/role-playing (group project including the script)*

Aims: to enhance comprehension, speaking, empathy, personalization

Give them at least one month before the skits: each group is required to get together to practice outside the class and to submit the script on the day of their performance to the instructor. Students may add their original lines in their skits imagining the characters' feelings in the particular scene if they like. Encourage them to use the narrator or some theater props, and to wear some costumes in order to make their skit more effective. Let them write down each student's role in making the skit such as the director and the typist in the script as well as the names of the characters they're acting so that each student shares the responsibility for the group work.

## IV. COURSE OUTCOMES

### 4.1 Questionnaire and the results of the questionnaire

In order to investigate the educational effectiveness of this course, on the final day of the course, a Likert-type self-report questionnaire with one open-ended question in Japanese developed by the instructor (see Table 1) was given to the students who had taken this course for the academic year. The 16 items of the questionnaire were conducted with a five-level modulated response scale ranging from "strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5)". Students chose the number (1-5) which indicates their opinion for each question. Items 1-3 investigated the students' motivation for reading a novelization as well as the effectiveness of viewing the movie. Item 4 revealed whether the students could empathize with any

characters in the story. Item 5 focused on whether the students felt they improved their critical thinking skills and imagination. Items 6-11 focused on the degree to which the students felt they improved their reading, listening, vocabulary, grammar, and knowledge of American culture. Items 12-14 were about the impact of literature: reading a novel and poetry on the students. Items 15-16 investigated the benefits of doing a skit (role-play) in groups. The last item, 17, was an open-ended question regarding their personal feedback about this course. In the two classes, a total of 47 students participated in the questionnaire, and every student (except one who failed to answer item 13) answered all 16 multiple-choice items. As for the last item of the open-ended question, more than half of the students (26 out of 47 students) in the two classes responded. Microsoft Excel was used to organize and analyze the data.

The following table shows the results of the class survey on this course. The 16 items of the questionnaire were divided into 6 elements by the categories of the question items as follows: motivation for reading a novelization (3 items), effect on empathy (1 item), effect on critical thinking skills (1 item), effect on L2 and culture (6 items), effect of literature (3 items), and effect of doing a skit (2 items).



**Table1. Distribution of Item Responses**

Item Category and Number	Percentage of Students Selecting Option					<i>M</i>	<i>SD</i>
	(1)	(2)	(3)	(4)	(5)		
<b><i>Motivation for reading a novelization</i></b>							
1. Reading a novelization of a movie was interesting for me.	0.0%	0.0%	17.0%	34.0%	48.9%	4.31	0.75
2. I felt a sense of achievement after reading the original literary text throughout the year.	2.1%	4.2%	14.8%	29.7%	48.9%	4.19	0.99
3. Watching the movie was helpful for me to understand the story.	0.0%	0.0%	0.0%	19.1%	80.5%	4.80	0.39
<b><i>Effect on Empathy</i></b>							
4. I could empathize with a character/characters while reading the story.	0.0%	4.2%	12.7%	40.4%	42.5%	4.21	0.83
<b><i>Effect on Critical Thinking Skills</i></b>							
5. Through this course, I think my critical thinking skills and imagination have improved.	0.0%	6.3%	21.2%	51.0%	21.2%	3.87	0.82
<b><i>Effect on L2 and Culture</i></b>							
6. Through this course, I think my reading skills have improved.	0.0%	17.0%	25.5%	46.8%	10.6%	3.51	0.90
7. Through this course, I think my listening skills have improved.	4.2%	25.5%	46.8%	21.2%	2.1%	2.91	0.85
8. I could understand the movie with the English subtitles and/or get the dialogue.	0.0%	17.0%	21.2%	42.5%	19.1%	3.63	0.98
9. Through this course, I think my vocabulary including colloquial expressions and slang has increased.	4.2%	4.2%	12.7%	63.8%	14.8%	3.80	0.90
10. Through this course, I think my English grammar has improved.	6.3%	23.4%	51.0%	14.8%	4.2%	2.87	0.89
11. Through this course, I think my knowledge of American culture has increased.	0.0%	10.6%	12.7%	55.3%	21.2%	3.87	0.87
<b><i>Effect of Literature</i></b>							
12. Through this course, I became more interested in reading poetry in English.	4.2%	31.9%	23.4%	25.5%	14.8%	3.14	1.16
13. Reading a novel can make a significant impact on my life, or make my life change.	0.0%	17.0%	23.4%	40.4%	19.1%	3.61	0.99
14. Reading poetry can make a significant impact on my life, or make my life change.	4.3%	19.5%	19.5%	36.9%	19.5%	3.47	1.14
<b><i>Effect of Doing a Skit</i></b>							
15. Doing a skit was helpful for me to understand the content and the characters of the story.	8.5%	6.3%	19.1%	34.0%	31.9%	3.74	1.22
16. Making a skit in a group and role-playing a scene/scenes from the novel were helpful for me to improve English skills.	6.3%	10.6%	23.4%	40.4%	19.1%	3.55	1.11

*Note.* The response options were: (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, and (5) strongly agree.

The first category of the question items (Items 1-3) indicates students' motivation for reading a novelization as well as its benefits. Reading a novelization of a movie was viewed as interesting (Item 1) by 82.9% of the students who answered agree (4) and strongly agree (5). The sense of accomplishment after reading an unabridged literary text throughout the year (Item 2) received 78.6% of the students' agreement. Watching the movie helped the students understand the story (Item 3) gained the most agreements 99.6% from them with the highest mean of 4.80 of all. The overall mean of these question items was 4.43, which shows using a novelization with the movie motivated the students to learn L2 in an EFL classroom.

The second category of the question item (Item 4) shows the lessons' effect on fostering empathy. 82.9% of the students empathized with a character/characters in the story well. The result implies that the lessons using this particular novelization was effective for developing empathy in the students. Selecting a literary material that students can empathize with is crucial so that they will be motivated to learn L2.

The third category of the question items was about the effect on developing critical thinking skills and imagination (Item 5). Importantly, 72.2% of the students agreed that they enhanced their critical thinking skills and imagination through this course, which was one of the main objectives for this course.

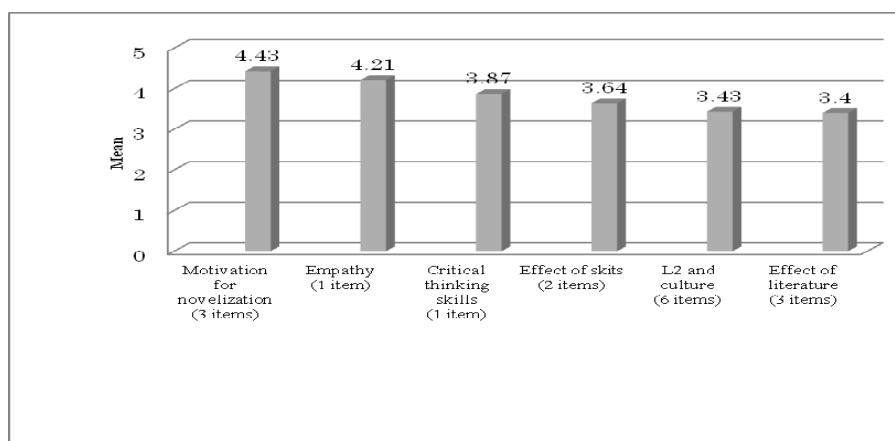
The fourth category of the question items (Items 6-11) indicates the impact on their L2 learning as well as their knowledge of American culture. First, the improvement on their listening skills and grammar (Items 7 and 10) received low agreements of 23.3% and 19% with the means of 2.91 and 2.87. On the other hand, the improvement on their reading skills (Item 6) was better recognized; 57.4% of the students' agreed. Remarkably, the increase of their vocabulary, including colloquial expressions and slang (Item 9), gained 78.6% favorable counts with a mean of 3.80. As Gareis, Allard, and Saindon (2009) have pointed out, authentic literature can provide great opportunities for vocabulary development. According to their study, through reading novels possessing words related to memorable characters and plots, L2 learners find vocabulary development easier than through regular textbooks or other texts. The course was not particularly designed for improving their English proficiency. It centers on helping students understand the content and on their personal growth as human beings. Therefore, these outcomes were somewhat expected except for the positive impact on their vocabulary improvement. Interestingly, even though the novel was fairly complex, 61.6 % of the students understood them and/or the dialogues when they watched the movie with the English subtitles (Item 8), which was better than it was expected. This would be due to watching the movie after the students studied the content. There were many scenes where the actors were speaking almost the identical dialogues as written in the novel. Additionally, the increase of the knowledge of American culture (Item 11) received 76.5% of the students' agreement. The overall mean of this section was 3.43, which implies the lessons were somewhat effective, but not a great deal in terms of improving English proficiency.

The impact of literature on the students is shown in the fifth category of the question items

(Items 12-14). As for increasing their interest in reading poetry in English (Item 12), 40.3% of the students responded favorably, which indicates the difficulty of reading poetry in L2. However, more than half of the students were aware of the power of literature; 59.5% and 56.4 % of them responded positively that reading a novel and poetry could make a difference in their lives (Items 13 and 14) respectively. The total mean of this category was 3.40, which implies a moderate impact of using literature on the students.

The last category of the question items is on the effect of doing a skit (Items 15 and 16). The effectiveness of doing a skit in groups on understanding the content of the story and the characters (Item 15) and on improving their English skills (Item 16) counted 65.9% and 59.5% of positive feedback respectively. The total mean, 3.64, shows that doing a skit among groups was somewhat beneficial for the students.

As the last item, 17, was an open-ended question regarding their personal feedback to this course, it is not summarized here, but described below in the next section. Additionally, Figure 1 shows the overall results of the effects of the lessons in each category based on the class survey.



**Figure 1. Overall results of the effects of the lessons**

## 4.2 Results of the open-ended question

For the last item 17, the students were asked to write down their personal feedback regarding the course and/or the novel. 26 students out of 47 responded: The positive feedback was 20, the neutral was 5, and the negative one was 1. Here are some actual comments from the students who responded positively translated from Japanese into English by the instructor:

- Even though the novel was rather difficult, I was into the story as we went on reading it.
- I am grateful that you chose this novel as the text. There were many scenes where I could empathize with the characters, and I could learn various things from this story.
- I looked for the actors on the Internet when we finished reading the story, since

I became so interested in the story.

- I felt a great sense of achievement throughout the year. I really appreciate it.
- Thanks to the skit, I could cooperate with my classmates that I usually didn't talk to, and I could understand the story even more.

#### 4.3 Student responses to graphic organizers

As for the tasks using various kinds of graphic organizers, each organizer seemed to be very effective to help the students to organize what they just read and learned, to comprehend the content of the story and the characters, to develop empathy, and to foster their critical thinking skills and imagination. Due to the limited time to cover the whole novel in one academic year (26 lessons totally), these tasks needed to be assigned as homework to each student. *Chapter organizer* (see Appendix A) was helpful for the students to keep track of the setting and the plot while reading the whole story. This novel's main character is a charismatic teacher, so at the beginning of the lessons, it was productive to let the students share their unforgettable teacher that they had encountered in their lives through *My favorite teacher*. Since there are many different characters, organizing each one's characteristics in *Character chart* seemed to be helpful for the students. In addition to this task, the students optionally could draw a *Character relationship chart (sokanzu)* on their own to deepen their understanding of the characters' relationships to one another. Many of them submitted their charts, which revealed their creative abilities in making their own graphic organizers and understanding the relationships among the characters. *Character emotional development* (see Appendix B) is a task which was done throughout the year; the students kept track of the two characters' emotional changes in each chapter that they chose by drawing the line graphs. This continuous task seemed to make them empathize with the two characters more and comprehend them better. This story is full of touching lines said mostly by Mr. Keating such as "Lads, there is a great need in all of us to be accepted, but you must trust what is unique or different about yourself, even if it is odd or unpopular" (Kleinbaum, 1989, p. 87). The students shared their impressive quotes at the end of each semester by filling in *Impressive quotes from the novel*. This was a great opportunity for them to look back on what they read, to savor the language, and to enhance their critical thinking skills. To share their favorite quotes in pairs seemed to encourage their empathic reactions even more. *Casting Dead Poets Society* was the task that stimulated their imagination and the students enjoyed doing it and sharing the work with the peers and the class. *Poetry response* (see Appendix C), one of the last tasks, was the one that allowed the students to answer in Japanese, while the other tasks needed to be completed in English or in English as much as possible. Even though most students had not read poetry in English before, they definitely appeared to enjoy reading the selected excerpts of the poems from the novel as shown in their work. Therefore, amazingly they could deeply interpret the poems that they liked and show their empathy and appreciation for reading poetry through this course. *Comparing and contrasting the movie and the novel* while watching or right after watching it, encouraged them to watch the movie critically.

Actually watching the characters in the movie appeared to involve them more emotionally in the story and deepen their overall understanding. Lastly, *Skits* seemed to work well to let them empathize with the characters by actually acting their roles and deepen their understanding at the end of the year, as supported by the results of the class survey. This also gave them an invaluable opportunity to cooperate with one another in groups and to feel a sense of accomplishment.

## V. CONCLUSION

This paper has presented content-based literature instruction using the novelization of the American movie, *Dead Poets Society* (Kleinbaum, 1989), as a motivating teaching medium for a university reading course. The overall student performance including assigned tasks, attitude toward the lessons, and class survey illustrated that the use of the novelization along with the movie in teaching reading is effective. Tasks using various graphic organizers and role-play skits were used in this course to integrate the four skills of English, and to increase the students' critical thinking skills and build their capacity for empathy. Most interestingly, by using a novel that has a plot and characters that the students can relate to throughout the year, a strong sense of community seemed to have been created in the language classroom. This led the students to be engaged even more emotionally in the story until the last despite the complexity of the language of the text (including the poems). The use of the excerpts of the poems written by renowned poets also appeared to be effective to inspire the students and draw them into the story. Amazingly, several students expressed a little sadness over completing the story on the last day commenting such as, "I will miss Mr. Keating."

On the other hand, as the focus of the instruction was on the comprehension of and appreciation for the content of the materials, not on measuring improvement of English proficiency by these lessons, the students were not tested to measure the effect of these lessons to improve their English proficiency. In order to measure the effectiveness of such a content-based literature instruction on students' English proficiency, a separate research study would be required.

Today, selecting a motivating material for less motivated university students is challenging in a reading classroom. As shown in this paper, carefully-chosen authentic literature can be effective teaching material for them. If a movie of the novel is available, it will assist students to comprehend the text and encourage empathic reactions. If the movie does provide these advantages, it will also reduce their anxiety and frustration to understand the second language and culture, and will make the whole learning process more enjoyable. As a result, it is likely that their reading combined with successfully including the movie will together give them more intrinsic motivation. Additionally, Stephen Krashen (1997) gives educators a wise advice on selecting a reading material:

In selecting what students will read, I think that teachers should consider their own interests as well. The teachers' enthusiasm (or extreme distaste) for a book will add a charge to the teaching, and make it much more interesting. The goal is to find texts that are interesting for both students and teachers. (p. 23)

Finally, as Christopher Brumfit states, by teaching literature, educators can introduce L2 learners to “such a serious view of our world” and initiate them “in the process of defining themselves through contact with others’ experience” (as cited in Carter, 2007, p. 3) and thus, educators’ task is basically to create the environment which makes this contact in the best ways they can (Carter, 2007). Hopefully, the power of literature will inspire not only students but also educators to keep striving and to live their own life to the fullest.

## NOTES

1. This paper is a modified version of the oral presentation given at the JACET Kansai Chapter conference at Osaka University on June 16<sup>th</sup>, 2012.
2. The question items of the class survey were translated from Japanese for Table 1.

## REFERENCES

- Brown, J. E. (2005). Film in the classroom: The non-print connection. *The ALAN Review*, Winter 2005, 67-72. Retrieved from <http://scholar.lib.vt.edu/ejournals/ALAN/v32n9/brown.pdf>
- Carter, R. (2007). Literature and language teaching 1986-2006: A review. *International Journal of Applied Linguistics*, 17, 3-13.
- Clandfield, L. (n.d.). *Teaching materials: using literature in the EFL/ESL classroom*. Retrieved from [http://www. Onestopenglish.com/support/methodology/teaching-materials/](http://www.Onestopenglish.com/support/methodology/teaching-materials/)
- Collie, J. & Slater, S. (1990). *Literature in the language classroom: A resource book of Ideas and activities*. Cambridge: CUP.
- Gareis, E., Allard, M., & Saindon, J. (2009). The novel as textbooks. *TESL Canada Journal*, 26 (2), 136-147. Retrieved from <http://www.teslcanadajournal.ca/index.php/tesl/article/view/419/249>
- Grunert, R. (2008). *Seize the day: The “cape diem” scene from Dead poets society as an anticipatory set for beginning a teaching unit on poetry*. Norderstedt, Germany: GRIN Verlag.
- Hismanoglu, M. (2005). Teaching English through literature. *Journal of Language and Linguistics Studies*, 1 (1), 53-65. Retrieved from <http://www.jlls.org/Issues/Volume1/No.1/murathismanoglu.pdf>
- Kasper, L. F. (2000). The short story as a bridge to content in the lower level ESL course. In L. F. Kasper (Ed.), *Content-based college ESL instruction* (pp. 107-121). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Kleinbaum, N. H. (1989). *Dead poets society*. New York: Hyperion.
- Krashen, S. D. (1997). *Foreign language education: The easy way*. CA: Language Education Associates.
- Lane, K. (1999). Graphic organizers for active learning: Goal for reading and writing. *The Language Teacher*, 23 (5), 19-21.

- Longman's Advanced American Dictionary (2000). Essex, England: Pearson Education Limited.
- Maley, A. (2001). Literature in the language classroom. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 180-185). New York, NY: Cambridge University Press.
- McCloskey, M. L., & Stack, L. (1996). *Voices in literature: Gold*. Boston: Heinle & Heinle Publishers.
- McKnight, K. S. (2010). Why are graphic organizers such important tools for teaching and learning? In *The teacher's big book of graphic organizers* (pp. 1-3). San Francisco, CA: Jossey-Bass.
- Richard-Amato, P. A. (1996). *Making it happen: Interaction in the second language classroom* (2<sup>nd</sup> ed.). New York: Longman.
- Schulman, T. (Writer) & Weir, P. (Director). (1989). *Dead poets society* [Motion picture]. United States: Touchstone Pictures.
- Seda, M. M., Liguori, O. Z., & Seda, C. M. (1999). Bridging literacy and social studies: Engaging prior knowledge through children's books. *TESL Journal*, 8(3), 34-40.
- Yohan (n.d.). *Eigo gakusyusha muke osusume paperback*. [The paperbacks recommended for English learners]. Tokyo: Author.

## Appendix A

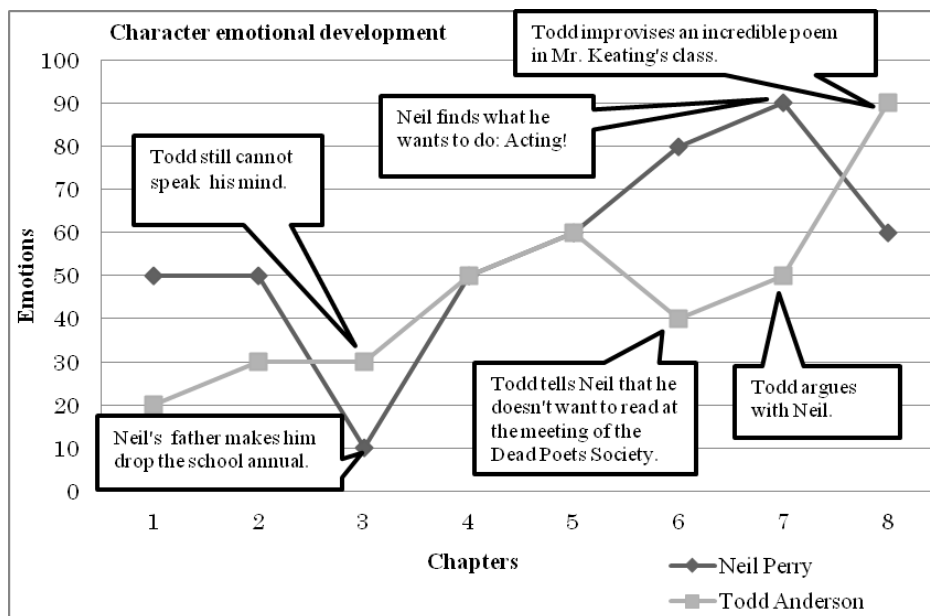
### Task 1: Chapter Organizer

#### Chapter Organizer: *Dead Poets Society*

Chapter	Characters	Setting	Plot
1	Gale Nolan, John Keating, Todd Anderson, Knox Overstreet, Charlie Dalton, Richard Cameron, George Hopkins, Neil Perry, parents, Alexander Carmichael, Jr.	- Chapel of Welton Academy, Vermont, U. S. A., 1959 (fall)	Headmaster Nolan makes a speech at the entrance ceremony. Later, he talks to a transfer student, Todd Anderson and then Neil Perry before their parents take off.
2	Mr. McAllister, Dr. Hager, Nolan, Neil, Todd, Cameron, Knox, Charlie, Meeks, Pitts, Mrs. Nolan	- Nolan's inner office - Dorm of W. A.	Nolan assigns the juniors extracurricular activities. As they head toward their dorms, Neil approaches Todd and introduces himself to his new roommate. Later, Neil's buddies come to his room and meet Todd for the first time.
3	Mr. Perry, Neil, Todd, Meeks, Charlie, Knox	- Neil & Todd's dorm room	Mr. Perry, Neil's father, suddenly shows up at his room and tells Neil to drop the school annual in front of his buddies. Neil gets upset, but he can't resist his father.

## Appendix B

### Task 5: Character Emotional Development



## Appendix C

### Task 8: Poetry Response

#### Poetry Response Form

Page # the poem used in the story:	Name of the poet:	Title of the poem (if available):
P. 41	Walt Whitman	O Me! O Life!
Impressive lines from the poem:		
"Answer. That you are here---that life exists and identity; That the powerful play goes on, and you may contribute a verse."		
<ul style="list-style-type: none"> <li>What did you like about this poem?</li> </ul>		
I like the strong message from the poet . . .		
<ul style="list-style-type: none"> <li>What did you think about: the words, the subject and the way the words sound?</li> </ul>		
<ul style="list-style-type: none"> <li>What does the poem remind you of?</li> </ul>		
<ul style="list-style-type: none"> <li>What does the poem make you think about?</li> </ul>		



## **A Rasch Analysis of Grammatical Difficulty: What Influences Item Difficulty?**

Atsuko Nishitani  
Kyoto Sangyo University

### **ABSTRACT**

This study investigated the difficulty order of 38 grammar structures obtained from an analysis of multiple-choice items using a Rasch analysis. The participants were 872 Japanese university students, whose TOEIC scores ranged from 200 to 875. The difficulty order of the 38 structures was displayed according to their Rasch difficulty estimates. This study also demonstrated the difficulty of writing items testing the knowledge of the same grammar point that show similar Rasch difficulty estimates. Even though the vocabulary and the sentence positions were carefully controlled and the two items looked parallel to teachers, they often displayed very different difficulty estimates. A questionnaire was administered concerning such items, and the students' responses suggested that they seemed to look at the items differently than teachers and what they notice and how they interpret what they notice strongly influences item difficulty. Teachers or test-writers should be aware that it is difficult to write items that produce similar difficulty estimates and their own intuition or experience might not be the best guide for writing effective grammar test items.

**Key words:** Rasch analysis, grammar, multiple-choice items, difficulty order

### **I. INTRODUCTION**

Communicative language teaching has been popular for approximately three decades, and the role of grammar teaching in the L2 communicative curriculum has been questioned (Purpura, 2004). However, researchers such as Canale and Swain (1980), Larsen-Freeman (1982), Canale (1983), Bachman (1990), and Bachman and Palmer (1996) have included grammatical knowledge as an important component of communicative competence in their models. Developing and measuring learners' grammatical knowledge should not be abandoned because "knowledge is always fundamental to developing a skill" (Haladyna, 1999, p. 7), and "[w]hat learners can do with language is to a very considerable extent dependent on what language they know" (Ellis, 2008, p. 18).

Although tests of learners' grammatical knowledge are widely administered, the relative difficulty of different grammatical structures has not been investigated or established yet (DeKeyser, 2005; Ellis, 2006). As DeKeyser (2005) pointed out, few researchers have compared the difficulty of a wide range of grammatical structures. After the morpheme studies in the 1970s and 1980s,

most of the studies on grammar difficulty were mainly focused on single grammar points, such as negatives, interrogatives, and relative clauses, and few researchers have investigated which negative form is acquired before which interrogative form and vice versa. Although Ioup (1983) and Ellis (2006, 2008) compared a number of different grammatical structures, Ioup only looked at six types of subordinate clauses, and Ellis focused on the difference between implicit and explicit knowledge rather than the difficulty of each structure per se, and the interrelationships among different grammatical structures were not discussed. Processability theory (Pienemann, 1998, 2003) predicts the difficulty order of different grammatical structures, but it is based on ESL learner data and not on EFL learner data. Also previous researchers who tried to verify the assumption of processability theory mostly used speech data. Baten (2010) analyzed data gathered with fill-in-the-blank items, but only examined German cases. In other words, the interrelationships among different grammatical structures have been discussed relatively little (see Nishitani, 2012 for a literature review), and thus, language testers tend to write grammar items based on their intuition or general and vague perception of difficulty of different grammatical structures, which is obviously unsatisfactory (Ellis, 2001).

In sum, the difficulty order of a wide variety of grammatical structures needs to be established in order for language testers to design efficient and reliable grammar tests. An empirically validated order can allow testers to avoid relying on their own intuition or experience of testing when choosing grammatical structures for a grammar test.

## **II. PURPOSES OF THE STUDY**

The first purpose of this study is to investigate the difficulty order of different English grammatical structures using multiple-choice items, which are frequently used in high-stakes tests in Japan. The dichotomous Rasch model is used to determine the difficulty of each grammatical structure and to construct an empirical hierarchy of difficulty.

The second purpose of this study is to investigate the issues that occur when attempting to write two items testing the same grammar point that have similar difficulty estimates.

More specifically, the research questions that guide this study are as follows.

1. What is a hierarchy of grammatical difficulty when tested in a multiple-choice format?
2. To what extent is it possible to create two items measuring the same grammatical point that have similar difficulty estimates?

## **III. METHODS**

### **1. Participants**

A total of 872 Japanese university students, 442 male and 430 female students, participated in this study. Of the 872 students, 461 were first-year students, 332 were second-year students, 52 were third-year students, and 27 were fourth-year students. Three hundred and four students were

majoring in English or other faculties that have a curriculum stressing English. Their TOEIC scores ( $n = 121$ ) ranged from 200 to 875 ( $M = 509.18$ ,  $SD = 148.47$ ), and their TOEFL ITP scores ( $n = 172$ ) ranged from 313 to 543 ( $M = 418.28$ ,  $SD = 45.05$ ), which suggests that the participants' English proficiency was mixed, ranging from false beginners to advanced proficiency learners.

The grammar tests used in this study were administered to intact classes at six universities in Osaka, Kyoto, Kobe, and Nagoya, where the teachers who agreed to cooperate with this study were teaching. The participants were asked for permission to use the data for research purposes. All of the participants gave their permission to do so. The six universities were large, co-educational, private universities, whose *hensachi* (deviation value) ranges from 47 to 60.

## 2. Instrumentation

In order to investigate the difficulty of different grammatical structures, five tests were administered. Each test had two or three parallel forms that were made up of items testing the same grammatical structures; however, different lexis was used in the sentences testing the same structure. The assumption was that the two or three sentences would have the same or similar Rasch difficulty estimates. Cobb's (2011) Vocabprofile was run to check the lexical composition of the tests. The results indicated that 88.3% to 90.8% of the vocabulary used was within the first high frequency 2,000 words of English, which Japanese students are supposed to learn before graduating from high school. The items that did not work on Test 1 were revised and tested on Test 2. The items that did not work on Test 2 were revised and tested on Test 3. Test 4 had new items, and the items that did not work on Test 4 were revised and tested on Test 5. All of the tests used a multiple-choice format with one correct answer and three distractors. All items were scored dichotomously. Table 1 presents a brief description of the tests used in this study.

**Table 1. Description of the tests**

Test	Number of forms	Number of structures	First 2000 words
1	2	21	89.4%
2	2	15	89.0%
3	3	6	89.0%
4	3	17	90.8%
5	2	11	88.3%

## 3. Procedures

All the tests were timed in the sense that the participants were told how much time was allotted for each test. The time allotted for each item was set at 30 seconds, which is often said to be the maximum response time for grammar items on the TOEIC test. This amount of time should have been sufficient for most of the participants to answer the items, particularly for those that included easier vocabulary than the TOEIC test.

A pilot study was conducted prior to this study. After 278 grammar items in a multiple-choice format taken by 1,409 university students were analyzed using a Rasch analysis, pairs of items that were testing the same grammar point in a similar way (i.e., having a blank in the same position in the sentence) and that had similar Rasch item difficulty estimates were extracted. The average of the two Rasch difficulty estimates for each pair were calculated, and the difficulty order of 21 grammatical structures was tentatively established. The items were revised to predominantly include the high frequency vocabulary of English so that they would act as a purer measure of the participants' grammatical knowledge and their ability to identify the correct answer would not be influenced by their lexical knowledge.

Test 1 was administered at three universities in Kyoto and Nagoya. Each form consisted of 30 items: the 21 structures mentioned above, and nine items from the other form (i.e., nine items from Form A were included on Form B, and vice versa), which were used as anchor items. In other words, out of 42 items written for Test 1, 18 items were administered to all the participants. A total of 265 participants took either Form A ( $n = 129$ ) or Form B ( $n = 124$ ) of the test. Twelve participants, who happened to be in two classes where the test was administered, took both forms. The participants were given 10-15 minutes to complete the test. The data were combined using the 18 anchor items to place the items from the different test forms onto a common scale using the Rasch model. Note that the 12 participants who took Form A and Form B answered the anchor items twice. If they answered differently on the two test forms, those answers were excluded from the analysis.

The Rasch analysis was conducted using the WINSTEPS computer software, version 3.71.0 (Linacre, 2009). First, the infit mean-square statistics for each item were examined to determine whether they were in the range of the mean  $\pm$  twice the standard deviation of the mean square statistic. According to McNamara (1996), infit statistics are the most informative, given that they are focused the fit of the most typical observations, and "for  $n$  sizes of 30 or more, the [acceptable] range is the mean  $\pm$  twice the standard deviation of the mean square statistic" (p. 181).

Next, the difficulty estimates for each pair of items designed to measure the same grammar point were examined to determine whether they had similar estimates. According to Linacre (2011), "when we want to say 'Item A is definitely more difficult than Item B'... their measures need to be more than 3 S.E.s [standard errors] different." In this study, however, a 2 S.E. difference was used as the criterion to ensure that the two items had similar difficulty estimates. The pairs that did not meet the 2 S.E. criterion (i.e., 15 pairs) were revised and tested once again in Test 2.

Test 2 was administered in two classes at a university in Osaka. Each form was made up of 15 items, and of the 30 items, six items were also on Test 1, so they were used as anchor items. The participants took both forms of the test with a two-week interval separating the administration of the first and second test form. To avoid the possibility of an order effect, one class was given Form A first, and the other class was given Form B first. Because some students were absent either the first

week or the second week, the number of students who took both forms, only Form A, and only Form B was 48, 2, and 5, respectively. The test takers were given 5-8 minutes to finish the test. The test results were analyzed with the dichotomous Rasch model, and six pairs were revised and retested in Test 3.

Test 3 was administered at one university in Kyoto. Each form was made up of six items, and of the 18 items, three items were the same items as on Test 2, nine items were revised, and six were newly written. The participants ( $n = 51$ ) took all three test forms at one-week intervals and were given 2-3 minutes to finish the test. To avoid an order effect, the order of the forms (Forms A, B, and C) was randomized. Thus 16 students took the tests in the order of ABC, 9 students in the order of BCA, 12 students in the order of CAB, 8 students in the order of CBA, and 6 students in the order of BAC. Out of the 51 participants, 48 had taken Test 1 also; thus, the data gathered from those students were used as anchors when combining the data from Tests 1, 2, and 3. After obtaining the results from Test 3, the combined data of Tests 1, 2, and 3 were analyzed using the dichotomous Rasch model. They all had difficulty estimates within the 2 S.E. criterion; thus, further revisions were unnecessary.

Test 4 was administered at three universities in Osaka, Kobe, and Kyoto. Each form was made up of 30 items: an additional 17 structures selected from the pilot study and 13 anchor items from the previous tests. The participants ( $n = 242$ ) were given 10-15 minutes to finish the test. The number of participants who took Forms A, B, and C was 82, 79, and 81, respectively. The data from the three forms were combined using the anchor items. Again, after obtaining the data from the Test 4 administration, a Rasch analysis was conducted on the combined data of Tests 1, 2, 3, and 4, and 11 pairs were revised and retested in Test 5.

Test 5 was administered at three universities in Osaka and Kobe. Each form was made up of 22 items: 11 revised items and 11 anchor items from the previous tests. The participants ( $n = 289$ ) were given 7-11 minutes to complete the test. The number of participants who took Form D and Form E was 131 and 158, respectively. The data were combined using the anchor items. A Rasch analysis was conducted not only on the data from Test 5, but also on the combined data of Tests 1, 2, 3, 4, and 5 using one-step equating. Two items with similar Rasch difficulty estimates were selected for each grammatical structure, and the mean estimates of each pair were calculated to examine the difficulty order of 38 grammatical structures.

Then a questionnaire was administered at a university in Kyoto in order to ask the participants ( $n = 38$ ) for explanations as to why some incorrect responses could be chosen over the correct responses.

## **IV. RESULTS AND DISCUSSION**

### **1. Difficulty order of 38 structures**

Research question 1 is “What is a hierarchy of grammatical difficulty when tested in a multiple-choice format?” Table 2 shows the means of difficulty estimates of 38 pairs of structures. As the Wright map (Figure 1) and Table 2 show, the subjunctive was far more difficult than the other grammatical structures, followed by the relative pronoun *which* (as opposed to *where*), and an adjective after a be-verb and an adverb. On the other hand, present perfect was the easiest, followed by past tense. Note that this table should be read with care. Because the mean is the average estimate of two items and not the difficulty estimate of the example item listed in the table, the item in one column is not necessarily more difficult than the one below. For example, Item 19, *You need to be (active) when you participate in a discussion*, is listed above Item 38, *Peter felt (happy) when he heard he passed the test*. However, their difficulty estimates were .63 and .80, respectively, which is opposite to what the table indicates. It is interesting to note that adjective items were spread out in terms of difficulty, and an adjective immediately after a *be*-verb was easier than the one after a *be*-verb and an adverb. Items of parts of speech seem to be influenced by sentence position. Further details are reported in the next section.

**Table 2. Mean Difficulty Estimates of 38 Structures (from Most to Least Difficult)**

Item	Structure	Example	Mean
21	Subjunctive	Tom has suggested that his classmates (be) invited to the party.	2.775
30	Relative pronoun <i>which</i> (vs. <i>where</i> )	My cousin will move to New York, (which) is one of the biggest cities in the world.	1.520
18	Adjective: be adverb ( )	Most of the people who lost weight found that the raw food diet was especially (beneficial).	1.370
37	Past perfect	By the end of last week, Linda (had written) 10 essays on Japanese culture.	.980
20	Present tense	This book (tells) how I have worked with difficult people who complain about everything.	.870
19	Adjective: be ( )	You need to be (active) when you participate in a discussion	.605
38	Adjective: sensory verb ( )	Peter felt (happy) when he heard he passed the test.	.590
34	S-V agreement	The products you can order from the store (are) listed in the catalogue.	.570
10	Adverb: S ( ) V	The manager (periodically) blamed the economy for his section's poor performance.	.410
32	Future tense	The next meeting has been scheduled for next Monday, and a new project (will be) then.	.395
7	Conjunction: ( ) SV, SV. (only one conjunction in the choices)	(Because) she lost her English textbook, she had to buy another one.	.385

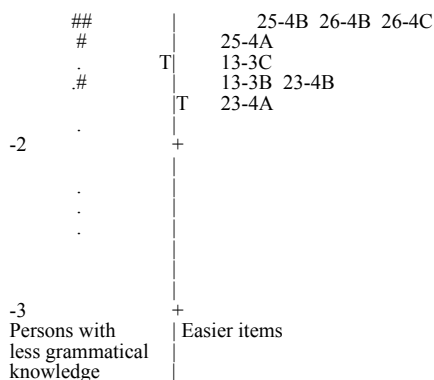
11	Passive + Present tense	The highway will be closed for maintenance this weekend. We are sorry for any inconvenience that (is caused) by the closing.	.365
12	Conjunction (when there are two or more conjunctions in the choices)	Ms. Lee changed the sales policy (while) she was the director of the department.	.360
33	Causative -ed	It is necessary to have your speech (checked) by a native speaker.	.355
27	Noun: SV ( )	The owner has checked (locations) for franchise stores.	.285
6	Adverb: be ( ) adjective	Eating only apples to lose weight seems to be (seriously) wrong.	.225
14	Adverb: be ( ) past participle	The babies' hands have been (cleanly) washed by their mothers.	.220
17	Past progressive (with another past tense verb in the sentence)	When we saw Jason in the cafeteria, he (was buying) sandwiches for lunch.	.110
22	Subject pronoun	In order for Ms. Jackson to win the game, (she) has to practice harder.	.005
31	Preposition (vs. conjunction)	(Despite) his age, my grandfather tried to climb Mt. Fuji.	-.025
35	Present participle: noun ( )	People (waiting) for the next bus should buy a ticket first.	-.105
1	Noun: article ( ) preposition	The company plans to build a factory for the (production) of a new model.	-.115
3	Past participle: be ( )	A new energy policy seems to have been (agreed) on by EU leaders.	-.120
4	Possessive pronoun	Although Internet shopping is easy and convenient, there still are people who are worried about using (their) credit cards online.	-.195
36	Past participle: noun ( )	In this area, there are a lot of buildings (designed) for older people.	-.260
29	Relative pronoun that	My sister sold the car (that) ran on both electricity and gasoline.	-.335
24	Adverb: SVO ( )	You should understand the textbooks (completely).	-.380
5	Adjective: ( ) noun	Ms. Taylor is known as a strict teacher and also as an (attractive) person in the city.	-.385
8	Base form: to ( )	In order to (act) the role of Romeo, he will have to go on a diet.	-.455
2	Reflexive pronoun: by ( )	Ms. Ford said she used to take care of five dogs by (herself) before getting married.	-.605
9	Infinitive: VO ( )	He told his assistant (to attend) the party because he didn't want to make a speech there.	-.690

15	Noun: verb adjective ( )	Our new computer will offer high (performance) when the new operating system is installed.	-0.950
16	Gerund	My brother has improved his English by (listening) to rock music.	-0.950
28	Verb: S ( ) O	According to the report, India (creates) the cheapest car in the world.	-1.040
26	Relative pronoun <i>who</i>	Students (who) want to pass the test should study at home too.	-1.230
25	Object pronoun	If students wish, the final grade will be emailed to (them) in a week.	-1.260
13	Past tense (with keyword such as <i>last year</i> )	The baseball players (took) a bus to Tokyo last month.	-1.585
23	Present perfect (with <i>since</i> )	Jennifer (has lived) in Osaka since she moved from L.A. in 1998.	-1.710

*Note.* Correct answers are shown in parentheses.

Persons with more grammatical knowledge		More difficult items	
4	#####	+	
	.		
	####	T	
	.		
3	.	+	
	###		21-2B
	#		21-2A
	#####		
	##		
	#	S	
2	##	+	
	#####		
	###	T	
	#####		30-4A 30-4C
	#####		18-1B 18-2A
	#		
	#####		37-5D
1	#####	+	
	#####	M/S	20-2A 20-5D 37-4B
	##		38-4A
	#####		10-1A 19-2A 19-3C 34-4C
	###		11-1A 32-4B 34-4A
	#####		12-1A 12-1B 14-5E 27-5D 33-4B 33-5E 38-4C 6-5E 7-5D
	###		7-5E 10-2B 11-1B 27-4A 32-4A
	####		1-2A 14-1B 17-2B 17-3C 35-4C 6-1B
0	####	+M	22-4A 22-5E 31-4A 31-4C
	#		3-1A 3-1B 36-4A 4-1A
	#####	S	29-4B 35-4B 4-1B 5-1A
	#		1-1B 24-4B 24-5D 29-4C 36-4B 8-1B
	##		2-1A 5-2B 8-1A 9-1B
	#####		2-1B
	.		
	#	S	15-2A 16-5E 9-2A
-1	##	+	16-5D 28-5D 28-5E
	.		15-3C





**Figure 1. Wright map for the 38 pairs of items. Each "#" is 6 persons. Each "." is 1 to 5 persons. M = Mean; S = 1 SD; T = 2 SD.**

## 2. Difficulty order within the same grammatical category

### *Adjectives*

There are four adjective items. Table 3 shows their mean estimates and mean S.E.s. Using the 3 S.E. criterion, their difficulty levels can be expressed as item 18 > items 19 and 38 > item 5. Items 18 and 19 indicate that an adverb after a *be*-verb makes the item far more difficult than placing an adjective immediately after a *be*-verb. This could be because adverbs are not explained explicitly in junior and senior high school textbooks, and thus less familiarity with adverbs might confuse some students and make answering item 18 difficult. On the other hand, placing an adjective immediately before a noun was the easiest.

**Table 3. Difficulty Order of Adjective Items**

Item	Sentence Position	Mean Estimate	Mean S.E.
18	be adverb ( )	1.370	.175
19	be ( )	.605	.170
38	sensory verb ( )	.590	.270
5	( ) noun	-.385	.160

### *Adverbs*

There are four adverb items, three of which were sentence-internal adverbs and were found to be more difficult than a sentence-final adverb. Table 4 shows their mean estimates and mean S.E.s. The difference between items 4 and 24 is slightly smaller than 3 S.E.s, but bigger than 2 S.E.s. If we use the 2 S.E. criterion, the difficulty levels of these four items can be expressed as items 10, 6, and 4 > item 24, which means that sentence-internal > sentence-final.

**Table 4. Difficulty Order of Adverb Items**

Item	Sentence Position	Mean Estimate	Mean S.E.
------	-------------------	---------------	-----------

10	S ( ) V	.410	.135
6	be ( ) adjective	.225	.195
4	be ( ) past participle	.220	.195
24	S V O ( )	-.380	.250

### **Nouns**

There are three noun items, and Table 5 shows their mean estimates and mean S.E.s. Using the 3 S.E. criterion, their difficulty levels can be expressed as items 27 and 1 > item 15. It is interesting to note that both items 27 and 15 require test-takers to place a noun in an object position, but the presence of an adjective makes the item much easier. In answering item 15, students might not perceive the noun as occupying an object position but rather as something occurring after an adjective.

It is also interesting that both items 15 and 5 (discussed in the adjective section) had the same structure, the combination of an adjective and a noun, but placing a noun after an adjective (item 15) was easier than placing an adjective before a noun (item 5). Their difference in estimates is slightly within 3 S.E.s but more than 2 S.E.s. In other words, item 15 was considerably easier than item 5.

**Table 5. Difficulty Order of Noun Items**

Item	Sentence Position	Mean Estimate	Mean S.E.
27	S V ( )	.285	.225
1	article ( )	-.115	.160
15	S V adjective ( )	-.950	.240

### **Tense**

There are six tense items. Table 6 shows that the past perfect and the present tense items were difficult and the past tense and the present perfect items were easy when tested in a multiple-choice format. It is unsurprising that the past perfect was the most difficult, as it is conceptually difficult for Japanese students, who often have difficulty distinguishing it from the simple past tense. The present tense, which is the first tense taught in junior high school, was difficult for the Japanese students in this study. In my own experience, many students use the present progressive when the simple present tense is appropriate. With item 20, the students might have had difficulty choosing the present tense because there were no adverbials such as *today* and *every day* in the sentences. In addition, item 20 had the present perfect tense in its subordinate clause, and this might have confused the students and influenced item difficulty. Similarly, item 32, which assessed knowledge of future tense, did not include adverbials such as *tomorrow* or *next month*, and had the present perfect tense in the other clause in the sentence. This might have increased the difficulty of this item. On the contrary, the items assessing knowledge of past tense (item 13) and present perfect (item 23) were relatively easy. The present perfect is difficult for Japanese students, and they often avoid using it. However, in this test both past tense and present perfect items had keywords such as *last*

and *since* in the sentences, which might have made the items much easier. In sum, the presence or absence of keywords such as *last*, *next*, and *since* appears to influence the difficulty of tense items.

**Table 6. Difficulty Order of Tense Items**

Item	Tense	Mean Estimate	Mean S.E.
37	Past perfect	.980	.230
20	Present	.870	.170
32	Future	.395	.200
17	Past progressive	.110	.220
13	Past	-1.585	.405
23	Present perfect	-1.710	.415

### ***Relative pronouns***

As Table 7 shows, there were only three relative pronoun items. Using the 3 S.E. criterion, their difficulty can be expressed as item 30 > items 29 and 26. The difference between items 29 and 26 was more than 2 S.E.s but less than 3 S.E.s. It is unsurprising that item 30 was the most difficult. When choosing a relative pronoun, Japanese students usually look at the noun immediately before the blank, because they are explicitly taught to do so in their junior and senior high school English classes. Because item 30 had a city name such as *New York* as an antecedent, many students chose *where* instead of *which*. Item 26 was the easiest, as the participants simply chose *who* when the antecedent was a noun referring to people. The relative pronoun *that* (Item 29) was more difficult than *who*, for two possible reasons: (1) Japanese students tend to learn that *which* is the relative pronoun to use when an antecedent refers to a thing, and *that* is an alternative; and (2) Japanese students often confuse *what* with *that*. In short, the difficulty order of these three relative pronoun items seems the same as the order that would be predicted by Japanese teachers of English.

**Table 7. Difficulty Order of Relative Pronoun Items**

Item	Relative pronoun	Mean Estimate	Mean S.E.
30	<i>which</i> (vs. <i>where</i> )	1.520	.160
29	<i>that</i>	-.335	.330
26	<i>who</i>	-1.230	.425

### ***Pronouns***

There were four pronoun items (Table 8). Using the 3 S.E. criterion, item 22 was more difficult than item 25. If we use the 2 S.E. criterion, the difficulty of the four items can be expressed as items 22 and 4 > item 2 > item 25. It is surprising that the subject pronoun item was more difficult than the object pronoun item. It might be because the students misunderstood it as a relative pronoun item. The blank was in the middle of the sentence, preceded by an infinitive phrase, and thus the presence of a noun before a blank could have confused the students. It is also surprising that the reflexive pronoun item was easy. However, this could be because the students

only knew it as a set phrase *by oneself*. If a reflexive pronoun were required immediately after a verb, the item might have been more difficult.

**Table 8. Difficulty Order of Pronoun Items**

Item	Pronoun	Mean Estimate	Mean S.E.
22	Subject pronoun	.005	.230
4	Possessive pronoun	-.195	.180
2	Reflexive pronoun	-.605	.175
25	Object pronoun	-1.260	.355

### ***Infinitives and gerunds***

There were two infinitive items and one gerund item (Table 9). Because they were within a 3 S.E. difference and there was only one pattern of gerund items, it is inappropriate to suggest that gerunds are easier than infinitives. It was surprising that item 8 was the most difficult of the three because I expected that placing a basic form of verb after *in order to* would be something students could do easily. However, 19% of the students who answered item “*In order to (act) the role of Romeo, he will have to go on a diet.*” and 21% of the students who answered item “*In order to (write) a good essay for your English class, you will need to follow some rules.*” chose a gerund for reasons that are not clear. Therefore, I administered a questionnaire and asked 38 university students attending the same school as many of the participants in this study if they could explain the reason(s) for this incorrect response. Their explanations provided some insight into the choice of the gerund:

- I thought that *to* is always followed by a noun.
- I thought that *to* was a preposition and *-ing* was appropriate.
- After the blank are a noun and a comma, and a subject and a verb follow. So I thought that it was a participle construction and *-ing* was correct.

**Table 9. Difficulty Order of Infinitive and Gerund Items**

Item	Infinitive / Gerund	Mean Estimate	Mean S.E.
8	Infinitive: to ( )	-.455	.210
9	Infinitive: V O ( )	-.690	.245
16	Gerund: preposition ( )	-.950	.240

### ***Prepositions and conjunctions***

There were three preposition and conjunction items, all of which required the students to distinguish between the two grammar points (Table 10). They were all within a 2 S.E. difference; thus, it can be concluded that prepositions and conjunctions can be of similar difficulty regardless of the number of the prepositions or conjunctions in the choices.

**Table 10. Difficulty Order of Preposition and Conjunction Items**

Item	Preposition/Conjunction	Mean Estimate	Mean S.E.
7	Conjunction: 1 conjunction in choices	.385	.195
12	Conjunction: 2 or more conjunctions in choices	.360	.130
31	Preposition: 1 preposition in choices	-.025	.295

## 2. Problematic items in the multiple-choice tests

There were two problematic items used in the multiple-choice tests, present tense (item 20) and an adjective after a sensory verb (item 38), which had two possible answers. The example items were as follows:

20-2A: This book ( ) how I have worked with difficult people who complain about everything.

(a) told (b) tells (c) telling (d) had told

38-4A: Peter felt ( ) when he heard he passed the test.

(a) happily (b) happiness (c) less happily (d) happy

The expected answers for 20-2A and 38-4A were (b) *tells* and (d) *happy*, respectively, but (a) *told* and (b) *happiness* were also possible. I noticed that item 20 had two possible answers before administering Test 2, but I left them as they were because I assumed that Japanese students who have received formal instruction in classroom would apply the knowledge of tense agreement and would not choose the past tense when the subordinate clause uses the present (perfect) tense. In fact more students chose (b) *tells* ( $n = 174$ ) than (a) *told* ( $n = 70$ ) for item 20-2A, for example, and their Rasch average ability measures were 1.91 and .96, respectively, which suggests that students with more grammatical knowledge chose the present tense.

I did not notice that two answers were possible for item 38 until after administering Test 4, the test that contained the items. I also examined the results of the Rasch analysis for this item and found that more students chose (d) *happy* ( $n = 43$ ) than (b) *happiness* ( $n = 18$ ) for item 38-4A, for example, and their average ability measures were 1.38 and .43, respectively. Again it was found that students with more grammatical knowledge chose the expected answer, an adjective. However, these items could have had different difficulty estimates if they had had only one possible answer.

## 3. Writing two items testing the same grammar point

Research question 2 was “To what extent is it possible to create two items measuring the same grammatical point that have similar difficulty estimates?” I first thought that items testing the same grammar point would show similar difficulty estimates. However, when conducting the pilot study, I discovered that the vocabulary used and the position of the blank in the sentence influenced

the difficulty considerably. Therefore, when writing the items for this study, I used high frequency vocabulary and had two parallel forms for the same structure in the same sentence position but using different lexis, so that the two items would be more likely to have similar difficulty estimates. However, in Test 1, which consisted of 21 grammatical structures, only six pairs of items showed similar difficulty estimates. The 15 pairs that did not meet the 2 S.E. criterion were revised so that they were more similar and they were tested once again in Test 2. Still, six pairs did not show similar difficulty estimates, and they were revised again to be even more similar and retested in Test 3. The revised items had difficulty estimates within the 2 S.E. criterion. However, note that Test 3 had three parallel forms for each structure, and the three items measuring each structure did not necessarily show similar estimates. Table 11 shows the items for six structures that finally met the 2 S. E. criterion after being revised twice and tested in Test 3. Note that the difficulty estimates in the table were from a Rasch analysis of the combined data of Tests 1 to 3. Notice how the difficulty estimates changed based on slight revisions to the items.

**Table 11. Difficulty Estimates of the Items Measuring Six Structures Revised Twice Before Meeting the 2 S.E. Criterion**

Structure	Test form	Item	Difficulty estimate
Infinitive: VO ( )	1A	Ms. Johnson wanted the students (to work) hard and finish the project by next Tuesday.	-1.14
	2A	Ms. Johnson told the students (to study) hard because they didn't get a good score on the last test.	-1.02
	1B	He told his assistant (to attend) the party because he didn't want to make a speech there.	-.78
	2B	He asked his assistant (to attend) the party because he didn't want to make a speech there.	.30
	3B	He asked his friend (to go) to the party because he didn't want to make a speech there.	-.55
	3C	Mr. Kondo wanted his children (to clean) the room because they didn't have space to sit down.	-3.12
Past tense	1A	Gasoline prices (increased) last month when the war broke out in oil producing countries.	-1.33
	2A	Because of the war last year, the oil companies (increased) the price of gasoline by 20 yen.	.59
	3A	The oil companies (increased) the price of gasoline by 20 yen last year.	-.86
	1B	According to news reports yesterday, a drunk driver (caused) an accident that resulted in five deaths.	-.90
	3B	A drunk driver (caused) an accident in Kyoto yesterday.	-1.92
	3C	The baseball players (took) a bus to Tokyo last month.	-1.31
Adverb: be ( ) pp	1A	The hotel owner told the local newspaper that all the rooms will be (beautifully) decorated by an Italian designer.	1.07
	2A	The rooms of the new hotel will be (beautifully) decorated by a famous designer from Italy.	1.55

	3A	The rooms of the new hotel have been (beautifully) decorated by a famous designer from Italy.	.49
	1B	The image of Japanese cars has been (greatly) damaged by wrong reports in the news media.	-.14
	2B	The image of Japanese cars has been (greatly) damaged by some wrong reports in the news.	.01
	3B	The image of Japanese cars has been (greatly) damaged by a news reporter from the U.S.	-.71
	3C	The walls of the old house have been (quickly) painted by a college student from Paris.	-1.92
Noun: verb adj ( )	1A	Our new desktop computers promise high (performance) while using less power than old models.	-.82
	2A	Our new computer will offer high (performance) when the new operating system is installed.	-1.00
	1B	The company has enjoyed the great (popularity) of its portable music player and online digital media store.	.65
	2B	The company has enjoyed great (popularity) since its portable music player was introduced.	1.85
	3B	The company will enjoy great (success) when its new television is introduced.	1.32
	3C	Tom will get a fast (promotion) when this new project is finished.	-1.58
Past progressive	1A	When I spoke with Amy on campus, she (was having) difficulty finding a job in design.	1.53
	3A	When I met Amy on campus, she (was drinking) coffee with a friend.	-.55
	1B	Jason (was buying) sandwiches when we saw him in the cafeteria.	-.11
	2B	When we saw Jason in the cafeteria, he (was buying) sandwiches for lunch.	-.09
	3C	When I called Michael at home, he (was cooking) dinner for his children.	-1.09
Adjective: be ( )	1A	Children who play outside need to be (careful) and should not talk to strangers.	-.54
	2A	Children who play outside need to be (careful). They should not talk to strangers.	.22
	3A	Children need to be (careful) when they play outside.	.22
	1B	The bed was too soft, so we had to sleep on the floor in order to be (comfortable).	-.97
	2B	The hotel room where we stayed used to be (comfortable). They also served a good breakfast.	-.93
	3B	Students need to be (comfortable) when they make a speech.	-.71
	3C	You need to be (active) when you participate in a discussion.	.11

*Note.* Correct answers are shown in parentheses.

### ***Infinitives***

Test form 1B had *He told his assistant (to attend) the party because he didn't want to make a speech there*. In Test form 2B, *told* was simply changed to *asked*, and then the difficulty estimates increased from  $-.78$  to  $.30$ , which was more than a 2 S.E. difference. This suggests that *asked* is more difficult in some sense than *told*. In Test form 3B, *assistant* was changed to *friend*, and then the difficulty estimates decreased from  $.30$  to  $-.55$ , which is again more than a 2 S.E. difference. This is understandable because *friend* is a higher frequency and more familiar word to the students, although the word *assistant* is used as a loan word in Japan.

The difference between 3B and 3C is also surprising. These two items appeared to be quite parallel when I wrote them, but their difficulty estimates were  $-.55$  and  $-3.12$ , respectively, which is more than a 3 S.E. difference. Is it because of the familiarity of the situation? Is the situation or meaning of *cleaning a room* more familiar than *going to a party and making a speech*, and thus influenced the difficulty? Neither *going to a party* nor *making a speech* seems difficult though.

Therefore, I also asked about these items to the 38 students mentioned above. Once again, their explanations provided a measure of insight into the responses:

- We are more familiar with *want to* than *ask to*.
- I thought *that* was omitted after *ask* and *his friend* was a subject in the *that*-clause. So I thought a verb should be chosen.
- I thought that it is strange to have two *tos* before and after *go* as in *to go to* so I didn't think *to go* was correct.
- I thought *to* in the *to the party* was the *to* to make *ask to* so I thought *to go* was incorrect.
- I didn't understand the meaning of 3B. I didn't understand the relationship between *a party* and *a speech*.

### **Past tense**

Comparing the past tense items on Forms 1A, 2A, and 3A, the difficulty estimates changed from  $-1.33$  to  $.59$  and to  $-.86$ . The difference between both 1A and 2A, and 2A and 3A were more than 3 S.E.s. In other words, 2A was significantly more difficult than 1A or 3A. This might be because a keyword such as *last year* appeared before the blank. 1A had a keyword immediately after the blank, and 3A had one at the end of the sentence. It might have been easier for the students to identify such keywords after the blank. This possibility also applies to the difference between 1B and 3B. Their difference was more than 2 S.E.s, and it might be largely because 3B was much shorter and simpler than 1B, but it is also important to note the position of the keyword *yesterday*: 1B had one before the blank, and 3B after the blank.

### **Internal adverbs (between a be-verb and a past participle)**

The difficulty estimates of the items on Forms 1A, 2A, and 3A were  $1.07$ ,  $1.55$ , and  $.49$ , respectively, and the difference between 1A and 2A was more than 2 S.E.s, and the difference between 2A and 3A was more than 3 S.E.s. Although 2A was shorter and simpler than 1A, 2A was



more difficult. The difference between 2A and 3A was only the tense used (i.e., the future form *will be* was changed to present perfect *have been* in 3A), and although present perfect is usually more difficult for Japanese students, 3A was significantly easier. I asked if the 38 students had an explanation for this, but I received only two possible reasons:

- In 3A, it is easy to see the present perfect verb and because a blank was in the middle of the present perfect form, it is easy to choose an adverb.
- I automatically think that a noun or an adjective should follow *will be*.

Comparing the items on Forms 3A, 3B, and 3C, which looked quite parallel, their difficulty estimates were surprisingly different: .49, -.71, and -1.92, respectively. The difference between 3A and 3B was more than 3 S.E.s, and the difference between 3B and 3C was more than 2 S.E.s, and the difference between 3A and 3C was more than 3 S.E.s. Why was 3A significantly more difficult than the other two? Is it because of the words *decorate* and *designer*? Japanese college student should know both words, but it is possible that the students did not recognize the spelling of *designer* although it is commonly used as a loan word in Japanese. On the other hand, it is understandable that 3C was the easiest because it had the easiest vocabulary and meaning. I asked if the 38 students could explain the difficulty difference of these items as well. Their responses are as follows:

- We rarely see or use *beautifully* but we often see and use *quickly*. We tend to choose words that are familiar.
- The meaning of 3C was easiest to understand.
- When reading 3B, I translated *damaged* as the noun *damage* and thus thought that an adjective should be in the blank.

### ***Nouns (after a verb and an adjective)***

The difficulty estimates of the items on Forms 1B, 2B, and 3B were .65, 1.85, and 1.32, respectively. The difference between 1B and 2B was more than 3 S.E.s, and the difference between 1B and 3B was more than 2 S.E.s (and barely missed 3 S.E.s). In other words, 1B was significantly easier than the other two items, even though the sentence was longer. This might have occurred because 2B had a *since*-clause, and 3B had a *when*-clause, which might have caused the sentences to look more complicated and confused the students. To confirm my assumption, I asked the 38 students about this as well, and I received several insightful comments:

- 1B was the easiest because there is an article *the* before the blank, which made it obvious that we should choose a noun.
- 1B was the easiest because *of* is after the blank. This made it obvious that we should choose a noun.
- 2B and 3B consisted of two *S + Vs* and looked difficult.

- 2B and 3B had a passive verb at the end, which made it difficult to understand the meaning of the sentences.

The difference between 3B and 3C was also significant. The difficulty estimate of 3B was 1.32 and that of 3C was -1.58, which was more than a 3 S.E. difference. Again the items on Test 3 were created to look quite parallel, and both *success* and *promotion* should be familiar to the students. The suffix *-tion* in *promotion* might have made it easier for the students to recognize it as a noun. Explanations provided by the 38 students, which partially confirm my assumption, are as follows:

- 3C has an article *a* before the blank, which makes it easier to choose a noun.
- *Promotion* is easily understood as a noun because of *-tion*, but it is difficult to know the noun form of *succeed*.
- When memorizing, we use the same Japanese translation for *success* and *succeed* and thus it is difficult to know which one is a noun.
- I thought that *succeeding* was a noun form of *succeed*.
- 3B has a verb *enjoy* so I thought that *-ing* was the correct answer.

### ***Past progressive tense***

The difficulty estimates of the items on Forms 1A and 3A were 1.53 and -.55, respectively, and the difference was more than 3 S.E.s. This could be because *having difficulty* was a much more difficult concept than *drinking coffee*. Also some students might have learned that the verb *have* cannot take the progressive. In fact, slightly more than 50% of the students who answered this item chose either *has* or *has had*. This is why I changed it to *drinking coffee* when revising the item for Test 3.

### ***Adjectives (after a be-verb)***

The difficulty estimates of the items on Forms 1A, 2A, and 3A were -.54, .22, and .22, respectively, and the differences between 1A and 2A, and between 1A and 3A were more than 2 S.E.s. Comparing 1A and 2A, 2A consisted of two short, simple sentences, but 2A was found more difficult. Comparing 1A and 3A, 3A was shorter but more difficult. I asked the 38 students for possible explanations, but they were unable to provide plausible answers. I received only two plausible comments.

- In 2A, the blank appears at the end of the sentence, so I thought that an adverb that modifies the be-verb should be used.
- 3A has *when* after the blank, which made it difficult to decide what word to select.

Comparing the items on Forms 3A, 3B, and 3C, whose estimates were .22, -.71, and .11, respectively, the differences between 3A and 3B and between 3B and 3C were more than 2 S.E.s; thus, 3B was much easier than 3A and 3C. It is possible that the word *comfortable* was easier than

*careful* and *active* despite the lack of any clear reason why that should be so. The concepts in the three sentences do not seem difficult either. I cannot think of any possible explanation for this. Therefore, I asked the 38 students; their comments partially confirmed my assumption:

- Among the choices of 3B, *comfortable* was most well known, so many students chose it based on their familiarity with the word.
- Because *comfortable* has the suffix *-able*, it is easily recognized as an adjective.
- Among the choices for 3C, I did not know which one was an adjective.

## V. CONCLUSION

This study established a tentative difficulty order of 38 grammatical structures when tested in a multiple-choice format. It also demonstrated the difficulty involved in writing multiple-choice grammar items testing the knowledge of the same grammar point that would produce similar Rasch difficulty estimates. Slightly different wording sometimes changed the difficulty estimate significantly. Even though the vocabulary used and the sentence positions were carefully controlled, the items often displayed quite different difficulty estimates. As discussed above, the students seemed to look at items differently than most teachers, and what they notice and how they interpret what they notice strongly influenced item difficulty. In the meantime, teachers or test-writers should be aware that it is difficult to write items that produce similar difficulty estimates. Especially when making college entrance examinations in Japan, teachers are usually required to make multiple forms, and thus they should be careful to produce forms of similar difficulty; otherwise, one form can be considerably more or less difficult than another and thus be the cause of unfairness to some students.

## VI. LIMITATIONS

The first limitation concerns the testing format. Multiple-choice tests are utilized in this study, as they are predominantly used to measure grammatical knowledge on entrance examinations and placement tests in Japan. Thus, the results of this study might not apply to different testing formats, particularly tests that elicit on-line data (e.g., speaking tests), or tests that measure productive knowledge of grammar.

Second, a limited number of grammatical structures are focused on in this study. To investigate the difficulty order, 38 structures were examined, and two items were selected as representatives of each structure. Assessing 38 structures, or two items per structure, is insufficient because (a) there are many more structures in the English language and their difficulty remains unknown, and (b) if the difficulty of one structure is determined, more than two items testing the structure should show similar difficulty.

Third, the difficulty of some items was unexplainable. Although I asked 38 university students, I still do not have a plausible explanation for why *Children need to be (careful) when they*

*play outside.* is more difficult than *Children who play outside need to be (careful) and should not talk to strangers.*, for example. Interviewing more students could provide useful information as to why one is more difficult than the other.

The fourth limitation concerns the items included in the multiple-choice tests. Two items had two possible answers. Although more students and students with more grammatical knowledge seemed to have chosen the expected answers, their difficulty estimates could have been different if there had been only one possible answer. Although these items were proofread by a native speaker of English, these problematic items were not detected beforehand. In order to prevent such problems, more teachers should proofread test items before administering them.

Finally, multiple-choice test items are affected by many factors, especially the distractors. When students are more familiar with the words used as distractors than the correct answer, they may choose the distractor. When they mistakenly notice the collocational links between words used in the sentence and a distractor, they may choose the distractor. For example, in *The company will enjoy great (success) when its new television is introduced.*, those who noticed that *enjoy* takes a gerund form might have chosen the distractor *succeeding*, which was an unintended result. Again careful writing of distractors is necessary when writing multiple-choice items. These limitations should be addressed by future researchers.

## ACKNOWLEDGEMENT

I would like to thank the anonymous reviewers for their valuable comments. This study was partially supported by KAKENHI (Grant-in-Aid for Scientific Research (C) 22520596).

## NOTE

This study was orally presented at JACET Kansai Chapter 2012 Fall Conference held at Kyoto Sangyo University on November 24<sup>th</sup>, 2012.

## REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.
- Baten, K. (2011). Processability theory and German case acquisition. *Language Learning*, 61, 1-51. doi: 10.1111/j.1467-9922.2010.00615.x.
- Canale, M. (1983). On some dimensions of language proficiency. In J. Oller (Ed.), *Issues in Language Testing Research* (pp. 333-342). Rowley, MA: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.

- Cobb, T. (2011). Web Vocabprofile. Available at <http://www.lex tutor.ca/vp/eng>.
- DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55, 1-25.
- Ellis, R. (2001). Some thoughts on testing grammar: an SLA perspective. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 251-263). Cambridge, England: Cambridge University Press.
- Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, 27, 431-463.
- Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, 18, 4-22.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). Mahwah, NJ: Erlbaum.
- Ioup, G. (1983). Acquiring complex sentences in ESL. In K. M. Bailey, M. H. Long, & S. Peck (Eds.), *Second language acquisition studies* (pp. 41-55). Rowley, MA: Newbury House.
- Larsen-Freeman, D. (1982). The 'what' of second language acquisition. In M. Hines & W. Rutherford (Eds.), *On TESOL '81* (pp. 107-128). Washington, DC: TESOL.
- Linacre, J. M. (2009) WINSTEPS: Rasch-model computer programs (Version 3.71.0). Chicago, IL: [winsteps.com](http://winsteps.com).
- Linacre, J. M. (2011). Winsteps Help for Rasch Analysis. Retrieved from <http://www.winsteps.com/winman>
- McNamara, T. F. (1996). *Measuring second language performance*. London, England: Longman.
- Nishitani, A. (2012). Investigating an order of grammatical difficulty: A review of literature. *Acta Humanistica Et Scientifica Universitatis Sangyo Kyotiensis*, 45, 59-79.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
- Pienemann, M. (2003). Language processing capacity. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 679-714). Oxford, England: Blackwell.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge, England: Cambridge University Press.

## **Figurative Interpretations Demonstrated by Three Language Groups: English, Chinese and Japanese<sup>1)</sup>**

Masumi Azuma

Kobe Design University (Emeritus Professor)

### **ABSTRACT**

This paper describes a study that investigates the cognitive aspects of the interpretations of figurative expressions among different native language groups (English, Chinese and Japanese). The term ‘figurative expressions’ refers to ‘hiyuteki hyogen’ (比喩的表現) in Japanese, the expressions of which in this study include metaphorical and metonymic expressions originating in Japanese and in English. The study aims to discover prominent cognitive aspects of figurative interpretations using an author-designed test. The test items feature both universal concepts and non-universal or specific concepts, i.e., culture-bound features originating in English and Japanese. The features and their effects on the understanding and interpretations of figurative expressions are considered; the causes and effects of cognitive prominence (e.g., analogy, schematic knowledge and mental images) used for interpretations, together with the influences of the knowledge generated from the participants’ mother tongues on figurative interpretations, are analyzed and discussed using the test results and interviews with participants. The analogical reasoning or logical thinking, image schematic functions, and cultural elements, including the knowledge generated from the participants’ mother tongue, are evidenced. The study aspires to contribute to language teaching, for example, avoidance or alleviation of miscommunication in the understanding and use of figurative expressions in different languages.

Key words: metaphor, analogy, culture, universality/specificity

### **1. INTRODUCTION**

Every language has its own unique features of figurative expressions. Some features may be rooted in the characteristics of the language itself or in cultural traits. This study investigates the figurative/metaphorical interpretations demonstrated by three different mother-tongue groups, with a view to identifying problems with figurative/metaphorical interpretations possibly caused by the schematic, cultural knowledge generated from a mother tongue and/or other cognitive factors and to alleviating such misunderstanding and miscommunication.

Regarding understanding and interpreting, both of which are inseparable in mind, thought and performance, this study takes the following view: understanding figurative meanings refers to perceiving or comprehending the expressions figuratively, whereas interpreting figurative expressions concerns construing, clarifying and/or explaining the meanings figuratively.

### **II. BACKGROUND**

Linguistics and its disciplinary areas have long been contributing to language education; however, it could be beneficial for more attention to be paid to the theory and mechanism of metaphor and its application, meaning those who are engaged in language education may deeply consider the cognition of language and the relation of language to brain, mind and body (Gibbs, 1980, 1994, 1999; Gibbs & Colston, 2012; Lakoff

& Johnson, 1980, 1987, 1999, 2003; Lakoff & Turner, 1989; Langacker, 1986, 1991, to name just a few).

With regard to the situation of figurative studies in Japanese, in the Meiji Era (the middle of the 19<sup>th</sup> century), literary figures revived the study of rhetoric and took significant notice of figures of speech. Since then, rhetoric has long been studied in depth, but it is only recently that figures of speech (especially, metaphor, metonymy, personification and onomatopoeia, etc.) have attracted people's attention regarding the aspects of language and cognition. Since the 1970s, in line with the development of cognitive science, in Japan, too, the number of studies in cognitive linguistics has increased to keep pace with research in the Western world. This development has led Japanese researchers to take an interest in the application of metaphor study to language education but not as much interest as in the study of metaphor per se in cognitive linguistics (Kusumi, 2007; Nabeshima, 2011; Taniguchi, 2006; Yamanashi, 1988, to name just a few).

The small number of studies or practices to take advantage of the theory of metaphor in foreign language (FL) education in Japan is due to the language environments: learners start learning an FL (mostly English) from a basic level during late childhood to early adolescence and move through an intermediate level to an advanced level at a near adult age. In general, figurative learning in an FL is thought to be an issue that should be dealt with at the most advanced level. In a situation like the one in Japan, it is quite natural for teachers to emphasize basic content when teaching English; therefore, only a few qualitative and quantitative studies have sought to investigate EFL learners' performance regarding the use of figurative English expressions. Due to the limited number of studies in Japan, this study tends to resort to resources in the fields of applied and cognitive linguistics in the Western world. Although limited, a strong implication comes from Kuroda (2007); Kuroda & Nozawa (online, 2012) imply the importance of cognitive knowledge ('frame') for understanding metaphors, and Taniguchi (2011) also inspires research into metaphor.

The Western world abounds with research, for example, on metaphor awareness (Deignan, Gabrys & Solska, 1997 and Boers, 2000); and on metaphorical competence (Charteris-Black, 2002; Johnson & Rosano, 1993; Littlemore, 2001; Littlemore & Low, 2006; Low, 1988). Of such research, the most relevant to this study is Charteris-Black (2002), who investigated the comparison of figurative phraseologies of English and Malay with Malay native speakers who had learned English as a second language (ESL). His study concluded that figurative units that have an equivalent linguistic form and equivalent conceptual basis are the easiest to understand, the most difficult being those that have an equivalent linguistic form but a different conceptual basis (*ibid.*).

This study also involves language and culture; therefore, it would be better to look at the research on metaphor and culture, for example, metaphor in culture (Kövecses, 2005), and in English (subtitled *Meaning and Culture*) (Wierzbicka, 2006), and then to consider the association between analogy and strategy (McCarthy, 2001; Oxford, 1990, 1996). Kövecses (2005) argues conceptual metaphors vary culturally and within cultures, and that the causes on which universality and variation in metaphor depend include embodiment (p. 293). He claims that the cognitive processes that human beings use are universal but that their applications are not. Wierzbicka (2006) covers a wide range of meanings of English words from other languages that involve specific cultural connotations; however, there is no reference to Japanese expressions.

Oxford (1990, 1996) classifies the systems of learning strategies in general, among which those most relevant to this study are analyzing and reasoning in cognitive strategies and guessing intelligently in compensation strategies.

As for metaphor interpretation processes and metaphorical ability, Littlemore (2008, pp. 201-204) points



out associative fluency, analogical reasoning and image formation as the processes involved in metaphor interpretation and word concreteness, with contextual clues and cognitive style as factors likely to affect one's ability to use metaphoric extension strategies (the term used in Littlemore, 2008 and elsewhere, is 'metaphoric', while the term in the author's study is 'metaphorical'). The definition of metaphoric competence (ibid., p. 201) is one's ability to perceive and create a metaphoric relationship between different concepts.

The author's past study, which investigated aspects of the metaphorical competence of Japanese EFL students, evidenced a strong correlation between EFL students' knowledge of English vocabulary and their metaphorical competence in English (Azuma, 2005). It was also noticed in the same study that some investigation into the cultural influence on figurative interpretations generated from mother-tongue knowledge and into the involvement of logical and schematic elements employed in interpretation is necessary.

### III. METHODOLOGY

As Winner (1988, p. 47) cautioned, there is 'no "pure" measure of metaphor comprehension'. It is a delicate problem, and so it is difficult to make a test for figurative measurements. However, to examine figurative interpretations, measurement instruments are necessary.

The methods used in this study were collecting numerical data and finding some cognitive aspects by (paper-based) testing, and confirming the contents of interpretations by interviewing participants. After several experiments and a pilot test, Metaphor Cognition Tests (MCTs) were designed. The tests comprised a multiple-choice type test and a descriptive type test. The multiple-choice type test consisted of 15 items, with 4 options each item: 4 items shared concepts in Japanese and English; 6 items featured English concepts; 5 items conceptually (and semantically) different in Japanese and English. It was a collateral test but aimed to measure and examine participants' interpretations quantitatively and to ensure the homogeneity of variances between the Japanese and Chinese groups. The idea behind the grouping of items was similar to that of the descriptive type test. The descriptive type test, the main test, consisting of 40 items, aimed to discover interpretative aspects qualitatively, focusing specifically on the relationship between the wordings and/or concepts of expressions and interpretation phenomena. Further details are given in Section III.1.

Quantitative analysis was conducted using SPSS 15E, while qualitative analysis was made by attentive reading of the given answers on the test paper, where participants were asked to write figurative meanings, and using (unstructured) interviews, where interviewees were asked individually about the schemas they employed, the strategies they used, such as their utilization of mother-tongue (NL) knowledge and drawing images in their mind when they encountered the expressions, and in the case of non-ENSs, linguistic problems. Participants were recruited in America, Australia, Britain, China and Japan in 2007-2012.

#### **Research questions (RQs):**

RQ1 What kinds of cognitive aspects are prominent in the interpretations demonstrated by three different native-language groups (English, Chinese and Japanese)?

RQ2 What are the causes and effects of successful and unsuccessful interpretations?

#### **III.1. The Metaphor Cognition Test (MCT): descriptive type test**



In the MCT descriptive type test, there are 40 test items of idioms, proverbs, and figurative expressions, of which a majority of the Japanese and the English expressions are (metonymic) metaphors that are frequently and/or emphatically used in daily life (e.g., on television, in newspapers or in daily conversations) (Appendix). The test was devised to obtain sufficient quantitative and qualitative data; however, in reality it was hard to fulfill the two antimononic requirements of quality and quantity, like ‘getting fish in troubled waters’. Although there is Deignan’s (2005, p. 118) indication of the content of a text for a figurative test, the author compromised by using test items of short and medium lengths considering the time and mental/physical stress for participants, with one or two words for the short items, phrases for the medium length items; and (short) sentences for the longest. The nature of these kinds of test items may result in limitations to this investigation: one limitation is the test items being for testing, i.e., tests for a test-sake, while another limitation is giving insufficient information regarding the target expression, and therefore, insufficient figurative stimuli.

These 40 test items are classified into two major groups: Groups A and B. Group A consists of linguistically and conceptually shared items among the three languages, i.e., the wordings and concepts of the expressions existing in these languages; therefore, a universality group, while Group B consists of linguistically and conceptually (though slightly or largely) different expressions, originating in Japanese and English; therefore, a non-universality or specificity group. For the verification of Chinese expressions to contrast with Japanese expressions, three Chinese academics assisted the non-Chinese NS author. The items that had perfect consensus were chosen to belong to Group A. For example, item 1 *Time is money* in Group A shares the concepts among the three languages, and the wording (in translation) is exactly the same; whereas item 23 to *cast a shrimp to catch a bream* in Group BJ is originated in Japanese. It shares the embedded concept that makes it possible to stimulate the human analogy of large size contrasting small, but the wording ‘bream’ is likely to be problematic for non-NSs. If a bream, ‘tai, 鯛’, is not understood to mean a valuable white fish as is served in Japanese cuisine, the expression may not be interpreted successfully. Group BE includes expressions originated in English. Item 19 *My sides split* and Item 20 to *spill the beans* may well be interpreted by English NSs, but may cause some interpretive problems by non-ENSs. Thus, some test items purposefully conceive ambivalent elements with the expectation of soliciting similar/different interpretations from English, Japanese and Chinese NSs. These features of the provision and classification of the test items may give a skew to the investigation, but those features may also give a clue for finding out interpretative tendencies or aspects of figurative interpretations in different languages, such as linguistic, schematic and cultural aspects in the interpretations of figurative expressions that originated in different languages.

The MCTs were presented in English; therefore, there was a presumable problem of translating the Japanese expressions into English. The translation of test items of Japanese origins took the method of literal or “raw” translation<sup>2)</sup>. Figurative expressions usually take a broad translation, i.e., using a literal paraphrase or an equivalent figurative expression in English; for example, item 31 *a horse out of a bottle gourd* appears in a Japanese-English dictionary (Masuda, ed., 1988, p. 501) as ‘Unexpected things often happen’. The broad translation makes the meaning clear, but it loses a delicate nuance or vivid imagery and makes the original expression non-figurative. This study takes an unprocessed (“raw” because it is in its natural state) measure of translation in order to preserve original nuances.

In the appendix, an item is laid out in the list, where the item number is preserved from the actual test and

an expression in English, the corresponding expression in Japanese, and its abbreviation and a Chinese expression (Group A) are listed.

Groups BJ and BE include the expressions that originated in Japanese and in English respectively. Some items are shared between the two languages, e.g., item 18 (Group BJ) You and I are *united with a red thread* is shared by Japanese and Chinese, but not by English, whereas there are items in the groups that are not shared (either slightly or greatly) among the three languages. However, some may possibly be understood linguistically, e.g., item 24 I need to *pick your brains*. A majority of the items were hypothesized to create a great (or small in some cases) difference between English and non-English interpretations. There are other items hypothesized to create a great difference between English and the other two languages, i.e., culture-bound items, for example, item 25 Tim must be *soft in the head* to do such a thing.

### III.2. Scoring of the answers and correctness ratio

Scoring of the answers took the following procedure: two raters (a British metaphor researcher and the author) set up an appropriate boundary of interpretation for each item and allocated one point for a correct answer or null for an incorrect answer. If there was any discrepancy between the two, they discussed the issue until they reached an agreement. After this procedure, the correctness ratio for each item was calculated (a hundred was the maximum; a zero was the minimum), and it was used for data analysis throughout this study. The use of the correctness ratio was intended to obtain an independent value of an individual item's interpretative performance by participants.

### III.3. Participants

The participants were 100 English native speakers (hereafter ENSs: 31 American ENSs (AmE hereafter), 32 Australian ENSs (AuE hereafter) and 37 British ENSs (BrE hereafter) aged 18 to 65; 72 Japanese native speakers (hereafter JNSs who majored in English in Japanese universities) aged 18 to 24; and 96 Chinese native speakers (hereafter CNSs who majored in English in a Chinese university) aged 18-22. None of the non-ENSs had had any experience of living in English-speaking countries for more than a year. The number of interviewees was 90 (56 ENSs and 34 JNSs).

## IV. COLLATERAL ASPECTS OF PARTICIPANTS

The participants in this study consisted of three different NL groups, and the ages of the three ENS groups had a wider range than those of their counterparts. The counterpart groups belonged to different Asian language groups, namely, Chinese and Japanese. Hence, we will first look at several collateral aspects in the following sections before we go into the main discussion.

### IV.1. Statistical examinations of the ENSs

First, we will examine the homogeneity of the three ENS groups and the feasibility and reliability of the test items, using the results of the MCT multiple-choice type test. The mean and std. deviation of the 100 ENSs were 95.2 and 6.1 respectively, in which those of 32 AmE, 31 AuE and 37 BrE comprised 95.8 and 5.3, 93.1 and 8.0, and 96.6 and 4.3 respectively; the homogeneity of variance for the ENS groups did not show a significant difference [ $F(2, 97) = 3.03, p = .053$ ] ( $\eta^2 = .05$ ); the feasibility of the test items was ensured by non-parametric tests; and the reliability (.981) was given by the Cronbach  $\alpha$  coefficient. Next is

the examination of the MCT descriptive type test. The ENS participants were from three regional areas; in other words, they spoke three different variations of English. It was possible to take advantage of this fact with regard to the regional similarities/differences and interpretative aspects of figurative expressions in an English context as a whole, which could be utilized as a benchmark and an index for non-ENSs. In contrast, the Chinese and Japanese participants could provide their primitive figurative interpretations. The presentation of the test in English gave a greater advantage to the ENSs than the non-ENSs. The correctness ratios of the multiple type of the ENSs vs the non-ENSs were 95 vs. 59.

With regard to the regional similarity/difference among the three ENS groups, the Pearson correlation between the three ‘Englishes’ is very strong as shown in Table 1. This result regarding the feasibility of the test items is remarkable, because if there was a chance for the JNSs, for example, to use the kinds of expressions featured in this study in communication with the ENSs, there would be no major interpretative difference among the ENSs.

Table 1 Correlations between Australian, British and American ENSs (no. of items: 40)

Measures	31 AuE	37 BrE	32 AmE
31 AuE	—		
37 BrE	.97**	—	
32 AmE	.96**	.97**	—

Note. \*\*  $p < .01$

#### IV.2. Statistical examinations of the non-ENSs

Here we will examine the equality of the two groups of non-ENSs, and consider whether or not the homogeneity of the variances is guaranteed, by using the MCT multiple-choice test of 15 items on the JNSs and CNSs. The mean and std. deviation of the 72 JNSs were 8.61 and 1.84 respectively and those of the 96 CNSs were 9.15 and 2.11 respectively. In the Descriptive Statistics, the means were slightly different between the two groups; however, the ANOVA showed there was no significant difference between the groups [ $F(1, 166) = 2.95, p = .088$ ] ( $\eta^2 = .02$ ). The results indicate that the equality of the two groups is substantially ensured.

A reliability examination is easier to run for a quantitative test; however, a reliability examination of the descriptive type test may not be as easy or suitable as would be the case with the multiple-choice type, because the descriptive type test includes test items that present complex problems, such as the inclusion of expressions of different origins and presumably delicate interpretative differences among languages; therefore, these complex interpretations should be looked at qualitatively.

#### IV.3. Differences in gender, age and major (English major or non-English major)

Classification according to gender and majors formed two separate groups. The results of the independent-samples-t-tests for gender showed no significant differences between the scores for male and females. However, the results for English and non-English majors showed differences, with the English majors’ scores being higher. Note should be taken of the indication by Gibbs & Colston (2012, p. 277) regarding tendency of female to avoid miscommunication when using figurative expressions.

Concerning age groups, there were two age groups for the non-ENSs: one was under 20 years old and the other was 21-25 years old. There was no significant difference between the scores for the two groups. In the case of ENSs, they were put into five groups as shown in Table 2. The table shows the averages of the correctness ratios. From these results, we can draw a line between the groups under 25 and those over 26 years old. The older generations' correctness ratios were higher.

Table 2 The averages of the ENSs age groups

age	-20	21-25	26-30	31-40	41-
correctness ratio average	74.8	76.7	78.6	78.5	78.8

## V. RESULTS AND DISCUSSION

First, we will take a holistic view of the results, such as the strategies used in interpretations that may relate to cognition, item features in the MCT as measurements, and contextual support or surrounding information in the items. This discussion will answer part of RQs 1 and 2.

### V.1. An overview of the results

#### Strategies used in interpretations

The interviews investigated what kinds of strategies for interpretation the participants utilized in their answers. In the interviews, firstly, general questions were asked, such as why the interviewees had written such answers, and secondly, the interviews focused on the kinds of strategies employed for interpretations: analogical reasoning or logical thinking, images or pictures drawn in their minds (i.e., activation of images or image schemas), and the use of knowledge generated from the participant's mother tongue. Among these strategies in general, because analogical reasoning functions as a fundamental interpretation and is related to the whole cognition, it is hard to identify single strategies. Hence, the following two outstanding categories are focused on in the investigation: resorting to conceptual schemas, that is, such schemas as are rooted in or generated from the knowledge of a mother tongue and culture (K stands for this category); and activation of image associations (i.e., visualization) (V stands for this category).

The overall phenomena shown in the MCT descriptive type indicate that, in general, the three nationalities seemed to utilize their general knowledge or schemas (cultural and/or knowledge generated from the mother tongues) in the interpretations of the items, especially that of the shared wordings and concepts. Table 3 summarizes the results of strategies (excluding three items, 14, 17 and 24, which showed large discrepancies in the correctness ratios between the ENSs and the non-ENSs). A further investigation into these three items is delineated later.

The strategies most often used were those of resorting to general knowledge or schemas and of image association. In the interviews, the ENSs and JNSs said that they initially searched for similar expressions in their mother tongue and associated the meanings with each other. If they could not resort to knowledge of their mother tongue, they went back to literal meanings or searched for connotations. When these strategies failed, they guessed figurative meanings using a clue derived from a part of the expression. If this failed, they gave up answering. They commented that another strong device for interpretation was image association. They visualized or pictured independently, or used as a clue either previously known images or instantly activated images (ad hoc imagination) generated from the mother-tongue knowledge. This proved to be a

successful strategy, which was especially useful in metonymic expressions or for those expressions having clear images. Some JNSs said that they translated the expressions into Japanese and attempted to deduce the meanings, by decomposing a whole expression into its constituent parts or by combining (re-composing) the words into an expression, in other words, a strategy using a part-whole relation to the expression.

Table 3 Strategies used in interpretations (no. of items: 37)

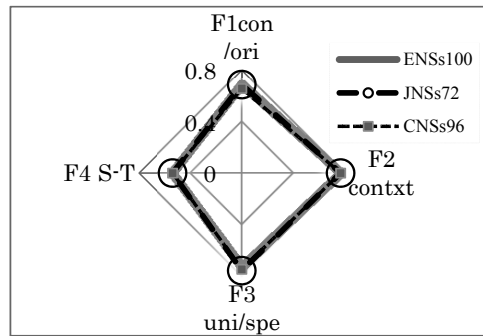
NS groups	strategies: K, K+V, V contributing to the correctness ratios		
	K (19 items)	K+V (10 items)	V (8 items)
100ENSs	79.2	54.3	85.4
72JNSs	41.9	60.6	49.6
96CNSs	41.8	37.8	34.1

Notes: K: knowledge; V: visualization; K+V: the combination

### Effects of item features

The features of the linguistic, cultural and schematic characteristics embedded in the MCT's 40 items may have contributed to the correctness of the interpretations. Four categories were formed as features: [Feature 1] shared wordings and concepts and different origins; [Feature 2] with/without context (sentence, phrase or word); [Feature 3] nature of universality or specificity (i.e., image clearness, culture-boundedness, idiomaticity); and [Feature 4] S-T (source-target) relation. In addition, in the case of CNSs, Chinese knowledge may have affected their interpretations so Chinese knowledge was also examined [Feature 5]. The values of these features were calculated by scoring items with each item being given one point maximum. Graph shows the four features in common (excluding [Feature 5]). The graph shows that the

Graph: Effects of item features



effects of F1, F2 and F3 were greater than those of F4. As the differences may look small in the graph, a Regression analysis was conducted. The 100 ENSs'  $R^2$ : .355, and Standardized Coefficients $\beta$  indicated positive in F1, F2, F3 and negative in F4. The  $R^2$ : .355 of the ENSs did not necessarily mean that it was low; it may have indicated that the participants did not rely on a single feature but may have taken into account all the features of the item. The 72 JNSs'  $R^2$ : .496, and Standardized Coefficients $\beta$  indicated positive in F4 but negative in F1, F2 and F3. The results were opposite to those of the ENSs. This may have indicated that they

had more of a tendency to rely on S-T or the part-whole relation. They may have attempted to interpret the expressions using some clues in them; for example, they tried to find out the meaning from parts of the words or a combination of words they knew. Finally, the 96 CNSs'  $R^2$ : .748, and Standardized Coefficients $\beta$  indicated positive in all of F1 to F4, and positive in Chinese knowledge, so they may have utilized all of these features.

### Contextual effects

Table 4 summarizes the results of the contextual examination, i.e., whether or not items are presented with surrounding supports. It can be deduced from the results of the ENSs that the contextual support is effective to some extent, but the nature of the expressions, for example, the wordings and concepts and origins of the expressions, contributed to the interpretations. The non-ENSs' interpretations were also better in the sentence-based items and the phrases than in the word-based items, except for the items that originated in their NLs. The lowest-scoring interpretations were the items that originated in the other languages. It can be seen from the results that if the expressions share the wordings/concepts between English and their NLs and if they are used in context, misinterpretation in communication may be reduced. However, it is necessary to investigate further what expressions share wordings and concepts among languages and it is possibly better to build database of such figurative expressions.

Table 4 Contextual effects

	context	no. of items	item no	100ENSs	72JNSs	96CNSs
Group A	sentence	11	1,3,7,8,9, 10, 26, 28, 29, 30, 33	91	68.8	77.1
	phrase	3	2,4,5	94.7	46.3	31.6
	word	1	15	91.6	16.7	48.9
	ave			91.8	60.8	61.7
Group B (BJ)	sentence	4	11,12, 18, 25	37.7	46.9	17.5
	phrase	7	21,22,23, 31, 32, 34, 35	29.8	59.9	17.5
	word	0				
	ave			32.7	55.2	17.5
Group B (BJ)	sentence	8	17,19,24, 27, 36, 38, 39, 40	92.6	31.3	33.2
	phrase	4	14,16, 20,37	90.3	4.9	12.2
	word(s)	2	6,13	87	13.2	25.6
	ave			91.2	21.2	26.1

### V.2. Figurative interpretations of the MCT items

The following sections discuss more of the qualitative and quantitative results of the interpretations revealed in the MCT items to answer RQ 1 and the causes and effects revealed in the interpretations to answer RQ 2.

### Overall phenomena in the interpretations of the MCT 40 items

Table 5 summarizes the results. As a whole, the ENSs showed high correctness ratios in the items sharing wordings/concepts (Group A, universality group) and in the items with English origins (Group BE), but

showed low ratios in items with non-English origins or meanings (Group BJ). One of the reasons for the high correctness ratios of the ENSs may have been due to the presentation of the test. The ENSs' mother tongue, English, might have linguistically, semantically and schematically favoured ENSs. Taking this condition into consideration, the average of nearly or over 60 percent for the correctness ratios of the non-ENSs should be evaluated as being sufficiently high. The JNSs possibly took advantage of their knowledge of Japanese in the interpretations of the items of Japanese origin but not in the items of English origin. The CNSs also seemed to apply their general, schematic knowledge to English and Japanese meanings. Their interpretations were successful in Group A, compared with those of BJ and BE.

Table 5 The averages of the 3 language groups

groups	Group A (universality)	Group BJ	Group BE
	15 items	11 items	14 items
100ENSs	91.8	32.7	91.2
72JNSs	60.8	55.2	21.2
96CNSs	61.7	17.5	26.1

(answers in Japanese meanings)

(answers in English meanings)

### Interpretation conformity/nonconformity or disparity of Group A

In order to identify the cognitive aspects, such as the linguistic, semantic, and schematic aspects (including the effects of mother-tongue knowledge), the logical thinking, and the cultural influence, the numerical results are viewed from high to low correctness ratios, with the focus on: interpretation conformity, nonconformity, and other phenomena.

Conformity issues in Group A are as follows: While a majority of the ENSs achieved a more than 90% correctness ratio in Group A, one item obtained a correctness ratio (hereafter abbreviated as cr in brackets) of only 53.6. That was item 7 At the age of 96, she *set out on a journey* to her husband. It is based on a conceptual metaphor and used as a euphemism. Furthermore, the age 96 in the expression was expected to assist interpretation, but the reality was different: the correctness ratio of the three NS groups were not as high as expected; the ENSs' correctness ratio (53.6cr) was not higher than that of the CNSs (77.1cr) and only slightly higher than that of the JNSs (48.6cr). The answers resulted in equity between literal and figurative interpretations. On the contrary, the reason for non-ENSs having taken the journey figuratively is that this expression is very popular and familiar in their languages, and the popularity and familiarity of the expression would have assisted them in making a successful interpretation.

Among the 15 items in Group A, the perfect conformity of the correctness ratio obtained not only by the ENSs but also by the non-ENSs was for item 1 *Time is money*. The polarity, easy vocabulary, and clear S-T relation of this item certainly contributed to the high correctness ratio even for the non-ENSs. A similarly successful interpretation was found in four items: 28 He is my *right arm*, 3 We are *at the crossroads*, 10 We could feel *electricity* between us, and 8 She was only saved from falling under the train by *a hair's breadth*. Each of these expressions was embedded in a sentence that provided a certain amount of information, a strong S-T relation and a clear image (confirmed in the interview). Another item 5 *to slip through one's fingers* attained a 58.3% correctness ratio in the CNSs and JNSs. Since the item was presented as a phrase, using one stimulus (slip through) paired with the other stimulus (fingers), participants could take advantage



of this pairing or collocation and of the clear image of the phrase (confirmed in the interview). This indicates that if the test items had some clue or information, interpretations would be easier. These are the cases that provided the fields<sup>3)</sup> fairly sufficiently. The case of *slip through* was one such, but there is another case that we should take note of, that is, the case of 4 *a bolt from the blue*. The ‘bolt’ in this item was paired with ‘the blue’, and the meaning was thought to be shared between Japanese and Chinese. However, the correctness ratios of the two NS groups showed a difference (JNSs: 56.9cr; CNSs: 16.7cr). No specific causes for the CNSs’ figurative interpretation can be clarified at this stage, but their notes mentioned in the test referred to a vocabulary problem: the word ‘bolt’ was unfamiliar to Chinese EFL learners.

High and medium conformities in the three language groups were revealed in two items: 30 You’ll need a *strong stomach* if you are going to be a surgeon (ENSs: 99.1cr; JNSs: 69.4cr; CNSs: 69.8cr) and 26 It is better to *tell a white lie* than to lose a friend (ENSs: 97.9cr; JNSs: 79.2cr; CNSs: 97.9cr). Item 30 is based on the container metaphor and embodiment schema. It seemed that these characteristics led to a successful understanding; although the word in the Chinese expression is liver instead of stomach, both organs are located in the middle of the body. In Asian thinking, the liver is a container of guts and the stomach is another container of physical/mental strength. Such cultural knowledge, the contextual support (the sentence-based item) and the utilization of analogy may have assisted the participants’ successful interpretation. However, it must be noted that in this expression, again, there was a linguistic and grammatical problem for some EFL learners: the latter part of the expression was taken to mean physically going to the doctor. This influenced whether participants gave a correct interpretation. The other expression, the word ‘white’ in item 26, shares the concept, image, and culture of the colour among the three languages. This characteristic led to a successful interpretation.

Medium correctness ratios (slightly more or less than 40%) were found in the interpretations of 15 *brainwave* (CNSs: 48.9cr; JNSs: 16.7cr) and 29 He has a *keen mind* (JNSs: 44.4cr; CNSs: 31.3cr). A majority of participants interpreted item 15 *brainwave* literally. Participants gave no specific cause, when asked in the interviews. A possible cause is that item 15 *brainwave* was presented as a word level item, i.e., stand-alone, in the test; therefore, it was literally taken to mean brainwaves as seen in a medical brain test. This is a warning regarding the research method. One of the other problems was the word ‘keen’ in item 29 He has a *keen mind*. It was found in the interview that this word was unfamiliar to EFL learners. Similar prominent phenomena among the non-ENSs that showed a big discrepancy from the ENSs were found in four items: 4 *a bolt from the blue*, 15 *a brainwave*, 2 *to bear fruit*, and 22 *double-tongued*. Some causes for the low correctness ratios found in the answers and confirmed by the interviews were vocabulary and grammatical problems (all 4 items) and a trend influence (e.g., item 4 being interpreted by the JNSs as the title of a popular song, ‘blue lightning’ 青い稲妻). Those participants who had a small vocabulary had a tendency to resort to intuition or tended to apply to interpretations the knowledge they already had (Azuma, 2005). One such example was ‘bear’ and ‘fruit’ in item 2 *bear fruit*. The item was decomposed into two parts, ‘bear’ and ‘fruit’, not collocated in a true meaning nor grammatically correct, which led to misunderstanding, such as ‘a bear’s food’ or ‘endure eating fruit.’ The medium correctness ratios in the ENSs but fairly low ratios in the non-ENSs were obtained in item: 33 He is a *weak worm* (ENSs: 63.4cr; JNSs: 52.8cr; CNSs: 40.6cr).

## Interpretation conformity/nonconformity of Group B



Here, the conformity issues are also viewed according to the correctness ratios of the items. The correctness ratios of the JNSs were high for those items with Japanese origins, those of the ENSs were high for the items with the English origins and those of the CNSs were high for some items with Japanese origins and for others with English origins, that is, midway between the other NS groups. Now we will focus, in line with high-low correctness ratios, on the following points: branching-off phenomena in interpretations; linguistic, analogical and schematic problems in interpretations; and linguistic, schematic and cultural problems.

Regarding high-low correctness ratios, one phenomenon common to the ENSs and the JNSs is that the frequently used conventional expressions in their NLs led to fairly high correctness ratios; the others showed individual aspects. Seven items are selected to represent phenomena typical of all three NSs for discussion, that is, those with medium or high correctness ratios: 18 *You and I are united with a red thread* (BJ), 36 *Curiosity killed the cat* (BE); and the items showing large discrepancies: 21 *a frog in the well* (BJ), 34 *a cry of a crane* (BJ), 20 *to spill the beans* (BE), 19 *My sides split* (BE), and 37 *to kick the bucket* (BE).

For item 18 *united with a red thread* (the Japanese origin), the JNSs' correctness ratio was 100. This item gives the JNSs' highest correctness ratio, and at the same time, the non-JNSs' branching-off phenomenon. The field and the situation sufficed for a relation schema and a bonding material to be used. The expression is so popular in Japanese communication as to be abbreviated to 'akai ito, 赤い糸' 'red thread', referring to the destined bonding of male and female. The Chinese expression is '紅糸' 'hon shen' which connotes the same bonding as in the Japanese expression; however, there are some debates in Chinese about the bonding material: thread, string or rope. The choice of the word for translation in the MCT may have influenced the CNSs' interpretations (43.8cr). The bonding meant by the Japanese expression is so delicate that the fine and thin texture of a thread is the best choice as a translation. The correctness ratio of the ENSs was not extremely low (37.4); their interpretation of this expression is quite interesting, because we can see analogical thinking and cultural influences functioning in the interpretations. The word 'united' indicates a relationship between the characters 'you and I', and some kind of relationship is expressed by 'a red thread'. These two words (red and thread) must have logically made the interpreters aware of the existence of a relationship; however, in the case of the ENSs, their logical and cultural inferences of 'a red thread' led them to a branching-off phenomenon: one branching-off resulted in a blood connection, i.e., a family relationship, which was inferred from the association of 'red' with 'blood', and the other branching-off resulted in a love relationship because it was inferred from the association of 'red' with 'passion.' Similar branching-off phenomena were found in ten items in Group B (items 11, 12, 16, 18, 20, 21, 31, 32, 34 and 40) and two other items (items 14 and 17).

Other high-low correctness ratios can be discussed for item 21 *a frog in the well* (BJ). This phrase is a part of the whole proverb 'A frog in the well does not know the ocean'. Had this item been presented in its entirety, the ENSs' correctness ratio would probably have been high. However, since a part was quoted from the whole, they interpreted it as meaning an environmentally harmful thing (a frog) was in the drinking water; hence, the interpretation referred to contamination. This is an example of how logical thinking worked. This popular expression shares the same wording and concept in Chinese and Japanese; therefore, were the non-ENSs' vocabulary sufficiently large, it would be understood without fail. Next is item 34 *a cry of a crane* (BJ), which refers to an authoritative voice (鶴の一声); this is often used by Japanese journalists, for example, 'Everything was determined by the president's cry of a crane'. The JNSs' correctness ratio (61.1)

was much higher than that of the other groups (the ENSs: 1.1; the CNSs: 0). The analogy the ENSs used in the interpretation was scientifically and culturally true up to a point: the interpreter commented in the interview on the nature and quality of the vocal sound a crane emits from its beak as a crane has a long beak, and so a sound emitted from such a long beak would be sonorous, penetrating, and reach the heavens. This is exactly what this expression is all about.

Typical linguistic, analogical and schematic problems of conventional English expressions are revealed in such items as 20 to *spill the beans* (BE) and item 19 *My sides split* (BE). The conventional expressions have rigid meanings, and it is hard for non-NSs to infer the meaning either from parts or from the whole, or other clues. The correctness ratios of the ENSs (97.8 for item 20 and 88.6 for item 19) make a clear contrast to those of the JNSs (4.1 for item 20 and 0 for item 19) and those of CNSs (13.5 for item 20 and 1.04 for item 19). Details of the answers are as follows: the CNSs' incorrect answers for item 20 were varied, and some indicated no knowledge of the words in the item, whereas the JNSs' incorrect answers for the same item converged on 'dropping beans' and 'spreading troubles', as inferred from the image of small things like beans scattering (31.9% of the JNSs), and 'mamemaki', 豆撒き, a spring event to get rid of evils (27.7% of the JNSs). Only 1% of the CNSs made an interpretation similar to that of the JNSs. The interpretation was quite logical until a branching-off occurred. In summary, the JNSs interpreted the expressions using logical thinking or by associating the expressions with Japanese culture.

Regarding literal and figurative connotations, the following phenomenon was found in item 19 *My sides split*: the element 'sides' (as a container) was interpreted figuratively by the ENSs, but literally by most of the JNSs; the action 'split' shared the interpretation. The CNSs' answers to item 19 varied: a number of the participants commented about having no knowledge of the meaning of the words used in the expression; a few provided an answer. Unique interpretations were found in the JNSs' answers, ranging from 'splitting of a group to which they belong' (18.1% of the JNSs) to 'clothes broken on both sides, because of, for example, getting fat' (6.9% of the JNSs). Only 4.2% of the CNSs provided an interpretation that was similar to that of the JNSs. One of the reasons for interpreting 'sides split', in the case of JNSs, was influenced by a Japanese katakana word 'saido, サイド' as in my 'side' meaning my 'group' or my 'team'; hence, the JNSs took 'sides split' to mean 'my group split'. This interpretation is a typical compensation example of applying the mother-tongue knowledge when the participants did not know the true meaning.

Another unique result was found in item 37 to *kick the bucket* (BE). The correctness ratio of the ENSs was high (81.1), while those of the non-ENSs were 1.3 (the JNSs) and 12.5 (the CNSs). The CNSs' correctness ratio for this item was higher than that of the JNSs, but this does not indicate that the CNSs' correctness ratios were always higher than the JNSs'. The level of the correctness ratios varies from item to item. Those who had learned figurative meanings in an English lesson or on their own could respond correctly. The JNSs' incorrect answers focused on the idea to 'kick a bucket (container)' (44.4% of the JNSs) and 'going wild', as in 'venting of one's anger' meaning to 'get angry' (19.4% of the JNSs).

The above results suggest that cognitive similarity among the NSs may be feasible in gaining an understanding of the expressions to a certain extent; however, there is a line beyond which different mother-tongue users cannot attain a correct interpretation only by linguistic or cognitive ability. Cultural knowledge and schematic flexibility may help improve understanding.

### Interpretation of culture-bound expressions

There are several items that specifically examined the tendencies of different NSs as confusable expressions: 17 to *pull someone's leg(s)*, 25 Tim must be *soft in the head* to do such a thing and 14 to *come to a head*. It was assumed for these culture-bound expressions that participants possibly tended to rely on the schematic knowledge generated from their mother tongues. As hypothesized, the tendencies shown in the interpretations generally reflected the meanings in participants' mother tongues. Table 6 shows that the ENSs interpreted the expressions in accordance with the English meanings but not the Japanese meanings. There must have been a reverse phenomenon in the other two NSs. However, some CNSs' interpretations tended toward the Japanese meanings; others, who knew the English meanings, interpreted the items according to the English meanings. Of course, some JNSs also interpreted them according to the English meanings; they said in the interviews that they thought they were expected to give the English meanings because the test was presented in English. The interpreters' considerations might have influenced the answers, but we can take this phenomenon to suggest interpretative ambivalence or an ambivalent element of choice. In short, these three nationalities utilized their familiar knowledge and employed previously known schemas in their interpretations.

Item 17 to *pull someone's leg(s)* (the singular 'leg' is used in the original English expression but a plural form was used in this study due to the two 'legs' of a human body) means to 'joke' in English, while the Japanese expression means to hinder or obstruct someone's action or performance, as in 'I pulled my team's legs' to mean 'I pulled down my team's progress'. The Chinese informants also indicated that this term was used in the same way in Chinese as in Japanese. Superficially, the interpretation looked different; therefore, misunderstanding may occur in daily use. However, as a figurative expression, we can find a notion of hindrance lying at the bottom, that is, 'joke' refers to taking an utterance as being of null value as in the ENSs' interpretation, while hindrance in the JNSs refers, also, to taking a null action.

Item 25 *soft in head* also has contrastive meanings: a negative meaning (stupid, empty brain) in English, and a positive meaning in Japanese (flexible thinking) and in Chinese (optimum).

Item 14 to *come to a head* means to come to a crucial point in English, whereas its similar Japanese expression, 'atamani kuru' is often used to mean 'get angry', where blood reaches the brain as if heated vapour were rushing up to the top of the body and exploding. The stimulus used in this expression shares between English and Japanese the sense of something coming to a critical point. However, the actual linguistic meanings differ. Some of the CNSs took the meaning in the Japanese way.

Table 6 Conformity/nonconformity of culture-bound items

BJ/BE	E meaning			J meaning		
	100ENSs	72JNSs	96CNSs	100ENSs	72JNSs	96CNSs
17pull leg(s)	96.8	8.8	12.5	0	75.2	44.8
25soft in head	94.8	7.1	22.9	0	80.5	12.5
14come to head	88.4	1.8	3.1	0	72.6	39.6
ave	93.3	5.9	12.8	0	76.1	32.3

In brief, there was a large difference between the interpretations of the ENSs and those of the non-ENSs: the CNSs' interpretations took an in-between position, and a good number of the CNSs tended to take the

Japanese meanings. Understanding of some Japanese figurative expressions seems to be feasible among the CNSs. However, in an actual discourse, conventional meanings take priority; therefore, users must follow the convention.

The figurative expressions that had similar wordings but different meanings were hypothesized to have interpretation nonconformity, and therefore, were problematic in cross-cultural communication. However, conceptually, the understanding/interpretation of culture-bound expressions, such as items 18 *united with a red thread* and 31 *bottle gourd* can be feasible to a certain degree with our intuition, logical thinking and/or linguistic/cognitive ability. It is better to take note again that if the conceptual knowledge about stimuli was weak or different from NSs' own, the interpretations would be varied.

## VI. CONCLUSION

Human cognition shows humans' vast and profound capacity for thought and action. This capacity is wisely and flexibly used in many ways when treating familiar and unfamiliar matters. It is also used in the interpretation of figurative expressions, as evidenced in the results of this study.

The prominent cognitive aspects found in the three language groups (RQ1) are summarized as follows. The familiar expressions used in Group A were fairly successfully interpreted by the different NSs; however, the results also suggest some problems, where misinterpretation occurred among the non-ENSs, who were especially affected by their linguistic knowledge or inclination, and Groups BJ and BE provided varied results.

Analogical capacity was manifested in the interpretation of both the familiar and unfamiliar expressions. One of the examples of an expression that was successfully interpreted was item 10 *We could feel the electricity between us*. For interpretation, an analogy of the nature of electricity was activated and mapped onto the relationship of the two. Another example is of the ENSs' low correctness ratio for the expressions with Japanese origins, but the analogy utilized in item 34 *a cry of a crane* was remarkable, as stated earlier. Although the correctness ratio of this item was very small (1), the analogy used in it exemplified human cognition.

Idiomatic, figurative and metaphorical expressions have a formulaic structure, and the true meaning is rooted in the specific language. Users must observe such rigid formulae. This is one of the hardships that EFL learners face. As a strategy for understanding idioms and figurative expressions, it was found in this study that the participants relied on their intuition and general knowledge, and on knowledge generated from the mother tongue in some cases. In other cases, some of which overlapped with the former, they made use of some clues residing in the expression, e.g., images arising from the expression or a part-whole relation of the expression. However, an ambivalent effect may also reside in the utilization of intuition or mother-tongue knowledge alone, as revealed in the ENSs' interpretations of the expressions that originated in a different language and in the non-ENSs' interpretations of some items (represented by item 14 *to come to a head*). Analogical reasoning is a very useful device for understanding figurative expressions. Nonetheless, this, too, has an ambivalent effect; for example, some items are understood by way of analogical reasoning to a certain degree, after which there is a possibility of a branching-off phenomenon, as revealed, for instance, in the interpretation of item 18 *you and I are united with a red thread*. In other words, analogical reasoning or logical thinking was successful up to a certain point, but after that, a branching-off occurred leading ultimately to an unsuccessful interpretation. In addition to this phenomenon, some of the misinterpretations

in Group B were due to linguistic problems, especially of the non-ENSs, such as items 19, 27, 38 and 39. An example (39 *getting cold feet*) shows the JNSs resorted to a phrase they knew: ‘catch cold’.

With regard to schematic and cultural effects, which were discussed earlier regarding the issue of item 18, we cannot get rid of cultural influence (and mother-tongue knowledge) when processing languages. Mother-tongue knowledge and culture are lodged in our body and mind and remain there. They are beneficial, but, at the same time, can have an ambivalent effect. A deliberate utilization of these benefits may expand our cognitive activities.

The causes and effects for the successful and unsuccessful interpretations (RQ2) are summarized as follows. From the results of Group A, if the wordings and concepts of the expressions are shared between languages, analogy and schematic knowledge may also be shared; therefore, there may be fewer problems in the interpretations. If the wordings of the expressions are shared but not the concepts, the interpretations may be affected either by analogy, schematic knowledge or cultural elements. This may cause a slight or a serious problem. One suggestion resulting from the study is regarding the need for caution about expressions having specific cultural traits and/or having similar wordings but different meanings. Another suggestion is to use figurative expressions with contexts.

This study attempted to investigate the aspects of figurative interpretations using the measurements that combined the concepts underlying language with the wordings exemplifying the concepts. We have found the interpretation conformity/nonconformity was possibly caused by innate individual cognitive abilities such as intuition and/or logical thinking, or analogy, and immune/external influence such as mother tongue or culture. Both had positive and negative effects on figurative interpretation.

## REFERENCES

- Aitchison, J. (1987). *Words in Mind*. Oxford: Basil Blackwell.
- Azuma, M. (2005). *Metaphorical Competence in an EFL Context*. Tokyo: Toshindo Publishing.
- Azuma, M. (2012). Chapter 9 English native speakers’ interpretations of culture-bound Japanese figurative expressions. In MacArthur, F., et al. (Eds). *Metaphor in Use*. John Benjamins Publishing Company.
- Boers, F. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics*, 21 (4), 553-571.
- Cameron, L. (2003). *Metaphor in Educational Discourse*. London: Continuum.
- Charteris-Black, J. (2002). Second language figurative proficiency: A comparative study of Malay and English. *Applied Linguistics*, 23 (1), 104-133.
- Deignan, A. (2005). *Metaphor and Corpus Linguistics*. John Benjamins Publishing Company.
- Deignan, A., Gabrys, D. & Solska, A. (1997). Teaching English metaphors using cross-linguistic awareness-raising activities. *ELT Journal*, 51 (4), 352-360.
- Gibbs, R.W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8 (2), 149-156.
- Gibbs, R.W. (1994). *The Poetics of Mind*. Cambridge: Cambridge University Press.
- Gibbs, R.W. (1999). ‘Researching metaphor’. In Cameron, L. & Low, G. (Eds). *Researching and Applying Metaphor*, 29-47.
- Gibbs, R.W. & Colston, H. L. (2012). *Interpreting Figurative Meaning*. New York: Cambridge University Press.
- Johnson, J. & Rosano, T. (1993). Relation of cognitive style to metaphor interpretation and second language

- proficiency. *Applied Psycholinguistics*, 14, 159-175.
- Kövecses, Z. (2005). *Metaphor in Culture*. Cambridge: Cambridge University Press.
- Kuroda, K. (2007). 「第 11 章比喩理解におけるフレームの知識の重要性」(楠見孝編『メタファー研究の最前線』217-238). 東京: ひつじ書房. (類似内容の論文 online, 2012 : 黒田航&野澤元「比喩理解におけるフレームの知識の重要性」)
- Kusumi, T. (Ed.). (2007). 『メタファー研究の最前線』東京: ひつじ書房.
- Lakoff, G. & Johnson, M. (1980, 2003 revised). *Metaphors We Live By*. Chicago: The University of Chicago Press.
- Lakoff, G. & Turner, M. (1989). *More than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago: The University of Chicago Press.
- Langacker, R.W. (1986). Introduction to cognitive grammar. *Cognitive Science*, 10 (1), 1-40.
- Langacker, R.W. (1991). *Concept, Image, and Symbol*. Berlin: Mouton de Gruyter.
- Littlemore, J. (2001). Metaphoric competence. *TESOL Quarterly*, 35 (3), 459-491.
- Littlemore, J. (2008). Chapter 10 'The relationship between associative thinking, analogical reasoning, image formation and metaphoric extension strategies. In Zanotto, MS, Cameron, L. & Cavalcanti, M. C. (Eds). *Confronting Metaphor in Use*, 199-222.
- Littlemore, J. & Low, G. (2006). *Figurative Thinking and Foreign Language Learning*. Basingstoke: Palgrave Macmillan.
- Low, G. (1988). On teaching metaphor. *Applied Linguistics*, 9 (2), 125-147.
- Masuda, K. (Ed.). (1988). *Kenkyusha's New Japanese-English Dictionary*, 4<sup>th</sup> edition. Tokyo: Kenkyusha.
- McCarthy, M. (2001). *Vocabulary*. Oxford: Oxford University Press.
- Nabeshima, K. (2011). 『日本語のメタファー』東京: くろしお出版.
- Oxford, R.L. (1990). *Language Learning Strategies*. Boston: Heinle & Heinle Publishers.
- Oxford, R.L. (1996). *Language Learning Strategies Around the World: Cross-Cultural Perspectives*. Hawaii: University of Hawaii.
- Taira, T & Kusumi, T. (2012). Relevant/irrelevant meanings of Topic and Vehicle in metaphor comprehension. *Metaphor and Symbol*, 27 (3), 243-257.
- Taniguchi, K. (2006). 『学びのエクササイズ 認知言語学』東京: ひつじ書房.
- Taniguchi, K. (2011). 「応用認知言語学と語彙学習 - 文法理論を英語教育に活用する (2)」『大阪教育大学紀要第1部門』 59 (2), 63-74.
- Wierzbicka, A. (2006). *English meaning and culture*. New York: Oxford University Press.
- Winner, E. (1988). *The Point of Words*. Harvard University Press.
- Yamanashi, M. (1988). 『比喩と理解』(認知科学選書 17) 東京: 東京大学出版会.

## NOTES

- 1) Part of this study (data collection of JNSs and ENSs and research at the earlier stage) was funded by JSPS KAKENHI ((C) 18520469). The data collection of CNSs was assisted by an academic in Hunan University, China. The author is grateful for their assistance.
- 2) Figurative expressions with Japanese origins in this study are translated literally (raw translation). There are arguments regarding the translation of figurative expressions, especially the pros and cons of literal translation (Dagut, 1976; Dobrzynska, 1995; Kellerman, 1986, 1995; Newmark, 1985, 1991; Nida, 1975;



Wright, 1999). The raw translation was used in this study to preserve the original figurativeness. As shown in the results, some raw translations were feasible (see Azuma, 2012).

3) The *field* is defined as the subject matter, or informational content, of the text (Deignan, 2005, p. 119).

## ACKNOWLEDGEMENT

The author would like to express her gratitude to the JKJ reviewers for their useful comments.

## APPENDIX

**MCT Groups:** test items, meanings/comments on specific items and abbreviations used in discussion

**Group A** 15 items (Note: C in brackets stands for Chinese.)

1. *Time is money.* 時は金なり (C: 時間之金錢)
2. *to bear fruit* 実を結ぶ 2*bear fruit* (C: 獲得成就)
3. *We are at the crossroads.* 我々は十字路に(居る) (C: 做出抉)
4. *a bolt from the blue* 青天の霹靂 (C: 晴天霹靂)
5. *to slip through one's fingers* 指をすり抜ける (C: 從指尖溜逃)
7. *At the age of 96, she set out on a journey to her husband.* (夫の許へ)旅立った(C: 上路)
8. *She was only saved from falling under the train by a hair's breadth.* 間一髪で (C: 千鈞一髮)
9. *The Mayors have been distinguished doctors for generations. It runs in the blood.* 血の中に流れている (C: 根深蒂固)
10. *We could feel the electricity between us.* 電気を感じる、ビビッと感じる (C: 有觸電感)
15. *a brainwave* 脳波、靈感 15*brainwave* (C: 脳電波、靈感)
28. *He's my right arm.* 私の右腕 (C: 得力助手)
29. *He has a keen mind.* 鋭敏な知性 (C: 思惟敏捷)
30. *You'll need a strong stomach if you are going to be a surgeon.* 丈夫な胃 (C: '肝'を使った表現はある)
33. *He is a weak worm.* 弱虫 (C: 病虫)

**Group B** (Note: The items of BJ & BE marked with \* indicate a large interpretation discrepancy between English and Japanese.)

**Group BJ** 11 items

11. *I cannot sleep with my feet turning toward him.* (彼のほうに)足を向けて寝られないほど恩がある (owe him a great deal)
12. *Prime Minister made iridescent remarks on the matter.* 玉虫色の (ambiguous)
18. *You and I are united with a red thread.* 運命の赤い糸で結ばれている (destined to be united)  
18*united with a red thread*
21. *a frog in the well* 井の中の蛙 (one who does not know the world)
22. *double-tongued* 二枚舌
23. *to cast a shrimp to catch a bream* 蝦で鯛を釣る
25. *Tim must be soft in the head to do such a thing.* \* E: 頭が空っぽ/ J: 頭が柔らかい (flexible thinking) / C: similar to J. 25*soft in head*
26. *It is better to tell a white lie than to lose a friend.* 無害の嘘 (C: similar to J. 善意的嘘)
31. *a horse out of a bottle gourd* 瓢箪から駒 (surprise) 31*bottlegourd*

Figurative interpretations demonstrated by three language groups:  
English, Chinese and Japanese

32. *to wet eyebrows with saliva* 眉唾 (doubtful)  
34. *a cry of a crane* 鶴の一声 (authoritative voice)  
35. *the carp on a cutting board* 俎板の上の鯉

**Group BE** 14 items

6. *a body blow* ボディーブロー  
13. *a loose tongue* ゆるい舌、しまりのない口  
14. *to come to head*\* E: できものが膿む、重大な局面になる / J: 頭に来る (angry, blood comes to head) 14come to head  
16. *to be off one's head* 理性を失う  
17. *to pull someone's leg(s)\** E: 冗談を言う、からかう / J: 足を引っ張る (obstruct) / C: similar to J 17pull leg(s)  
19. *My sides split.* 脇腹が割けるほど笑う、大笑いする  
20. *to spill the beans* 秘密を漏らす  
24. *I need to pick your brains.* 知恵を借りる  
27. *You have bought yourself a white elephant.* No one will stay in this house. 無用の長物  
36. *Curiosity killed the cat.* 好奇心旺盛は怪我の元  
37. *to kick the bucket* くたばる  
38. *Since Andrew started his own business, he has been making money hand over fist.* 素早く大量に  
39. *Simon is getting cold feet about advancing you the money.* 怖じ気づく 39getting cold feet  
40. *What does it matter what your uncle thinks of you? He only visits you once in a blue moon.* 稀にしか (来) ない