

**JLTA Newsletter No. 34**  
**日本語テスト学会**  
**The Japan Language Testing Association**

JLTA Newsletter No. 34 発行代表者：中村優治 2013 年 (平成 25 年) 3 月 15 日発行  
発行所：日本語テスト学会 (JLTA) 事務局  
〒389-0813 長野県千曲市若宮 758 TEL 026-275-1964 FAX 026-275-1970  
e-mail: [youichi@avis.ne.jp](mailto:youichi@avis.ne.jp) URL: <http://jlta.ac>



\*\*\*\*\*

**専門を学び究めることによって将来社会に還元する報恩奉仕の精神**

第 16 回全国研究大会実行委員長  
片桐一彦 (専修大学)

日本語テスト学会第 16 回全国研究大会は、2012 年 (平成 24 年) 10 月 27 日土曜日、専修大学生田キャンパス (神奈川県川崎市) にて開催された。ワークショップも前日に同じ場所で開催された。天候にも恵まれ気温も快適な穏やかな秋の二日間であった。2004 年 9 月に私の前任校の麗澤大学 (千葉県柏市) にて第 8 回全国研究大会を引き受けて以来、8 年ぶり 2 回目の会場校担当であった。

会場となった専修大学は、その前身である「専修学校」が 1880 年 (明治 13 年) に創立されたことから始まり、今年で創立 133 年を迎えた。「専修学校」は、日本で初めて経済科を、私学では最初に法律科を設けた高等教育機関であった。「専門 (経済学・法律学)」を学び究めることによって将来社会に還元する (報恩奉仕) という建学の精神から、「専修学校」そして「専修大学」の名前が由来している。その 130 周年を記念して建てられた 10 号館 (130 周年記念館) において、第 16 回全国研究大会を挙行政した。私も、自分の専門 (英語教育学・言語テスト理論) を学び深く研究することによって社会に還元したいと思っている。

日本語テスト学会の第 1 回全国研究大会は、1997 年の秋に東京経済大学で開催された。私はその時はまだ大学院修士課程の 2 年生であった。ちょうどスピーキング評価に関して修士論文研究をおこなっている最中であった。自分が研究したいと思っていた学問分野の専門学会が、まさに産声をあげて初めての全国研究大会を実施する歴史的瞬間であった。よって私は、言語テスト専門の本学会に初めて参加した時は、大変興奮していた。以来私は、全国研究大会には今のところ皆出席させていただき、また幸いにも『学会誌』に自分の論文を今まで複数掲載していただき、学会には大変お世話になっている。日本語テスト学会に、そして学会の役員や仲間の先生方にとっても感謝している。ということで、専修大学の報恩奉仕の精神で、2 回目の会場校を引き受けた。

大会を開催するにあたっては、学会事務局の中村洋一先生と法月健先生や大会運営委員会の先生方や専修大学の同僚をはじめ、多くの先生方にお世話になった。特に事務局次長の小泉利恵先生とは、1 年弱もの長い間、二人三脚で一緒に大会関係の仕事をこなしてきた。大変ではあったが、おかげで楽しくやりがいをもって仕事ができる。ここに深甚なる謝意を表したい。

**第 35 回日本語テスト学会  
研究例会報告**

**2012 年 6 月 2 日 (土)**

**於： 龍谷大学 (深草学舎)**

会場校の Lee 先生, お世話になりました。

(事務局, 広報委員会)

**e ラーニングを用いたテスト**

**樋口三郎 (龍谷大学)**

多様なテスト実施において、紙媒体ではできないところが e ラーニングでは可能になった。例えば、動画や音声を Quiz 内に搭載し、自由記述や選択肢で回答させることが可能となったことが上げられ、テスト実施の可能性を広げることができる。本報告は、龍谷大学の Moodle (e ラーニングと同義になっている) の活用状況、教育効果、そして樋口氏が授業で使用されているクlicker (クlickerとは、オーディエンスレスポンスシステム(ARS)、すなわち「講義者と学習者の双方向コミュニケーションを可能にするコミュニケーションツール」の1つ) の効用が紹介された。

e ラーニングによるテストは、紙媒体のテストよりも精度の高いテスト作成が可能となるが、e ラーニングによるテストには課題もある。例えば、「本人が受験しているのか?」「非参照で受験しているのか?」「非相談で受験しているのか?」などの紙媒体を使ったテストと違った不正が考えられる。不正対策として、「本人が受験しているのか? → ID のログイン」「非参照で受験しているのか? → WEB フィルタリング」「非相談で受験しているのか? → ファイヤウォール+WEB フィルタリング」などが報告された。また、情報機器が設置されていない通常教室の場合、e ラーニングを実施することが困難となる。しかし、その対策として「LMS である Moodle をモバイル端末から使う」「スマートフォンで、公式アプリケーションやテーマを使う」「フィーチャーフォンで Moodle for Mobiole モジュールを使う」「モバイル専用サービスと連携する」「クlicker か

らデータをインポートしてコンピュータで処理を行う」「携帯出席ツールからデータインポートしてコンピュータで処理を行う」という対策法が説明された。

普通教室で e ラーニングを使いたい場合、日本では高機能端末と化した携帯電話を使って行う場合が考えられる。その場合、Moodle と携帯電話を連携させる事や、携帯電話単体で様々な WEB サービスを使って行うことも考えられる。もしくは普通紙による OCR 処理など、紙媒体をデジタル処理して、データだけはパソコンで処理をすることも考えられる。近年、様々な WEB サービスがあり、それらは URL 埋め込みなど、連携が考えられてある場合も多いので Moodle 単体だけではなく、それら WEB サービスと組み合わせることも可能であることが紹介された。WEB サービスの中にはデータ蓄積型のサービスがある。YouTube は動画を蓄積しており、それらは別の WEB サービスで容易に使えるため、動画は Moodle にアップロードして格納し利用するのではなく YouTube から Moodle にリンクして使用するなどを考えると、リソースの節約や動画自体の使い勝手を考えると好ましい。e ラーニングで使う教材やテスト素材には、フリー素材をネット上で複数の人々がマッシュアップして使う事も考えられ、テスト素材を各先生方が組み合わせてテストを作り、それを新たな素材として他の先生方に提供することも可能となる。

李洙任(Lee Soo im)(龍谷大学)

**日本語テスト学会**

**Workshop**

**2012 年 10 月 26 日 (金)**

**於： 専修大学生田キャンパス**

**確認的因子分析**

**講師：印南洋 (芝浦工業大学)**

芝浦工業大学の印南洋先生を講師に迎え、「確認的因子分析 (confirmatory factor analysis)」の理論と実践についてのワークショップが開催された。昨年度は「探索的因子分析

(exploratory factor analysis) ]についてのワークショップが開催され、統計手順として、次に「確認的因子分析」を行うことが多いことから、昨年度に引き続き参加する人が目立った。

確認的因子分析の基礎的理論の説明の後、『教育・心理系研究のためのデータ分析入門』の書籍で扱っているデータを用い、その分析が出来る Amos というソフトを実際を使って、データの読み込み、モデルの作図のデモンストレーションが行われた。その後、算出された結果の読み方・解釈の仕方が丁寧に説明された。質疑応答では、探索的因子分析で仮定したモデルを検証する場合、そこで使用したデータとは別のデータを使って、確認的因子分析をすることが大切であるとのことであった。また、正規性 (univariate / multivariate normality) が満たされていない場合の対処法について、数多くある適合度指標のどれを参照し、論文に報告すべきかについての助言がなされた。最後に、高次モデル、MTMM モデル、多母集団モデルなどの応用モデルの事例も示された。基礎から応用まで幅広く説明がなされ、非常に内容の濃いワークショップであった。講師と 16 名の参加者の間で活発な議論がなされ、予定の 3 時間を 10 分ほど超過してもまだ足りないほどであった。

平井明代・横内裕一郎 (筑波大学)



## 基調講演 Keynote Speech

### Evaluating extensive reading

David BEGLAR

(Temple University Japan Campus)

The topic of Dr. David Beglar's keynote speech was relevant not only to researchers in language assessment but also to all the teachers who are involved in language teaching at any levels of education. Despite its importance, the assessment of extensive reading does not seem to have received due attention, mainly because of the difficulty of defining the construct, not to mention the difficulty of measuring it at the acceptable level of reliability and validity.

The ultimate goal of extensive reading programs, Dr. Beglar claimed, is developing students into competent readers, who enjoy reading and are capable of communicating what they read with others effectively. In order to achieve the goal, an attempt has to be made to help students engage in large amounts of pleasurable reading, with the expectation that it would eventually promote their cognitive, affective, and social as well as linguistic benefits. Evaluation in the extensive reading programme should be conducted on the assumption that assessment practices per se should support such fundamental educational goals. The classroom-based assessment for learning should readily encourage students to use all the modalities including speaking, writing and listening as well as reading. By doing so, the alternative approach to traditional assessment practices is expected to help

make assessment practices more authentic, in the way in which they would lead students to engage in the activity which would naturally occur outside of the classroom.

Dr. Beglar's presentation was informative, interesting and insightful. He interspersed his speech with many useful observations taken from his own personal experience in teaching, which contained both successful and unsuccessful attempts. It was consoling to know that he had a hard time managing the class the first time he taught the students who were not motivated to learn English. It was encouraging likewise to learn how he overcame the challenge. To implement the ideas that he shared with us in our own classroom, we have to inevitably adjust them in some way or other. But the basic principle has to hold in any context. For example, it may be difficult to have our students share what they have read with other students if they are not confident enough in speaking. In such a case it may be tempting for us to prepare a pre-task which is specifically designed to help them to speak before reading. However, perhaps this type of artificial arrangement, particularly if it goes too extreme, is at odds with the principle that Dr. Beglar illustrated,



because “extensive reading can serve as the basis for the development of all language skills” (Beglar, 2012, p. 1), which I believe is one of many lessons we should learn from the keynote speech.

Yoshinori WATANABE (*Sophia University*)

## 研究発表 Research Papers

第 1 室

### Mobile System to Monitor Extensive Reading

**Bill PELLOWE** (*Kinki University*)

**Trevor HOLSTER**

(*Fukuoka Women's University*)

This presentation began with an explanation of how extensive reading requires the reading of comprehensible and interesting text in large quantities. Detailed tests or other intensive reading type activities may reduce fluency and motivation and thus interfere with the purposes of extensive reading. The presenters demonstrated an innovative solution by having students do a quick survey that is stored in an online database. This was done using a free open-source module of a mobile audience response system called MOARS (<http://moars.com>). Students can answer the quick surveys online using a mobile phone, iPod, computer or any other device with an internet connection. This provides an unobtrusive extensive reading management system. Teachers can then monitor students' reading patterns by downloading book counts. Researchers can download reader and text data along with a Facets control file for many-faceted Rasch analysis.

The presenters concluded with a successful example using recently collected pilot data and an invitation for other schools to use the system.

J. LAKE (*Fukuoka Women's University*)

### **Can we treat the EPER Form A and Form E as alternate forms?**

**Kiyomi YOSHIZAWA** (*Kansai University*)

**Atsuko TAKASE** (*Kansai University*)

**Kyoko OTSUKI** (*Kansai University*)

Many extensive reading classes use the placement/progress test developed by the Edinburgh Project on Extensive Reading (EPER). Three forms of the placement/progress test are available: Form A, Form B, and Form E. These forms have been treated as alternate or parallel forms of the EPER placement/progress test. Although the conversion table is available, the researchers felt it necessary to confirm the difficulty levels of the forms when those forms are administered to the EFL learners for assigning appropriate reading materials and assessing their progress.

The presentation had two aims. The first aim was to conduct a content analysis of the EPER Form A and Form E. The second aim was to examine the difficulty levels of the two forms. To this end, two forms were linked using a common-person design. A total of 200 students at two four-year universities in Japan participated in the study. There were three groups of students: one group took Form A, another group took Form E, and the last group took both Form A and Form E. Rasch analysis was conducted to analyze the data files of the two forms.

The results suggest that Form E would be more difficult than Form A in terms of the length and readability, and the scopes differed from each other. In the Q and A session, the necessity of comparing the three forms and cloze test scoring methods were discussed.

Hideki IIMURA (*Tokiwa University*)

### **Learning by Assessing In Second Language Writing**

**Trevor HOLSTER**

(*Fukuoka Women's University*)

**Bill PELLOWE** (*Kinki University*)

**J. LAKE** (*Fukuoka Women's University*)

The presentation was to offer us valuable insight for the pedagogical usefulness of peer assessment of writing in EFL. The researchers used four-faceted-Rasch analysis to analyze the data collected from 26 university students in their academic writing courses. The instruction of the 15-week course was focusing on paragraph and essay organizations using a commercially available textbook for ESL writing. As part of the workload of the course, essays written by the students were randomly distributed and the students were instructed to assess them using the peer assessment rubric, which consists of 10 items: Thesis statement, Introduction, Conclusion, Organization, Unity, Support, Coherence, Cohesion, Relevance, and Formatting. The Mobile Audience Response System (MORAS), an open-source program, was used to rate each other's work on a four-point Likert scale. The system also allowed the students to access their own ratings instantaneously. The four-faceted

analysis indicated improvement between the first and second essays. However, some of the factors such as the item on the rubric and rater performance appeared to be a cause of misfitting with the statistical model.

The presentation prompted questions from the audience. Several questions focused on the research method as well as MORAS. One member suggested to interview students to obtain qualitative data. Although the current study did not provide enough information to value peer feedback as formative and diagnostic assessment, sharing rating process will surely give positive attitudes toward learning. The study will be pushed forward by the cooperation of the researchers with more samples in the near future.

Yuko SHIMIZU  
(Ritsumeikan University)

### **Gems of English: A tale of triangulated testing**

**John BLAKE** (*Japan Advanced Institute of Science and Technology*)

In this presentation, an end-of-course test for the staff training in a jewelry store in Thailand was introduced. This test was designed to meet multiple demands from different perspectives. The test consisted of two elements: proficiency tests and achievement tests. In consultation with the store management, a battery of tests was used with each one testing different aspects of the sales process.

After the presentation, many questions were asked to the presenter. One interesting question was “What skills were

required to succeed in the test besides the language skills?” John introduced one of the sales people who could tell which jewelry a certain customer wanted to buy just by taking a look at the person’s eyes, and so she was able to sell a lot of expensive jewelries in a limited amount of time. In such case, he insisted, two pages of English were sufficient for a sales person to give a good performance if he or she had a good personality and sales technique. I felt it was a typical example of a performance test, which regarded the language performance in a particular situation as a whole, rather than assessing each construct separately.

Ken YANO  
(*Hitachi Daini Senior High School*)

### **A Testing Environment: The Experiential Side of Testing**

**Myles GROGAN**  
(*Momoyama Gakuin University*)

This presentation detailed some of the qualitative experiences university teachers went through in developing a language placement test and some of the practical constraints that were faced. The presenter began with his own experience of learning how to develop and analyze tests. The presenter noted that the realities of making and giving tests are often quite different from the ideals that are presented in language testing textbooks. Using Excel, the presenter then showed colleagues how they could develop test items using classical test theory item statistics. This increased awareness of testing resulted in a type of professional development that created in

faculty members an interest and commitment for test development. The awareness of item and test statistics then led to improvements in the quality of items and the placement test through the development of a set item specifications and test blueprints. An audience member asked about the use of item writing guidelines such as are commonly found in books on test development and the presenter agreed that presenting a set of guidelines would be helpful to present to colleagues.

J. LAKE  
(*Fukuoka Women's University*)

## 第2室

### **Is a letter grading system appropriate for language classes?**

**Dong Wan CHO** (*Pohang University of Science and Technology, Korea*)

This study discusses the impact of two grading approaches on language learning and teaching: a dichotomous satisfactory/unsatisfactory (S/U) grading system and a multi-level letter grading system. The researcher collected views on the two systems from university students and instructors in Korea. Their preferences proved to be mixed and inconclusive. The S/U system appears to lessen the pressure on students to get better grades. On the other hand, the system seems limited in that students' motivation for learning is lowered because they are less likely to make an extra effort once they receive a *Satisfactory* grade.

This presentation and Q&A revealed that the question "Is a letter

grading system appropriate for language classes?" is too complex to be answered with a "yes" or "no." It was suggested that the choice of grading system should at least depend on (a) students' proficiency level and (b) the purpose of grading. Low-proficiency students may benefit from an S/U grade because it gives them a greater sense of achievement than a letter grade. On the other hand, the letter system may motivate high-proficiency students since it further distinguishes among high achievers in the course. In addition to students' proficiency, the purpose of grading should be taken into consideration. For example, the multi-level grading system is considered more suitable for providing specific diagnostic information (e.g., how well a student can do a certain task).

Importantly, this study has touched on an issue of the grading system that language teachers rarely give careful consideration. We should conduct further investigation in a variety of local contexts.

Takanori SATO (*University of Melbourne*)

### **Comparing Spoken Performances at Different Levels on the Same Tasks: What Linguistic Features do Japanese Learners of English Demonstrate at A2, B1 and B2 and Above?**

**井上千尋** (旭川医科大学)

ヨーロッパ共通言語参照枠 (CEFR) は汎言語的に作成されているため、具体性に欠ける面がある。Hawkins and Filipovic (2012)は、学習者書き言葉コーパスである Cambridge Learner Corpus (CLC)に基づき、CEFR 各レベルの具体的な言語的特徴 (基準特性、*critical features*) を22種類特定した。また、タスクの発話に対する影響

(task effects) は、語彙には見られるかもしれないが、文法には問題にならないレベルと論じている。そこで、本研究では同程度の難易度と思われる2種類のタスクを、3つの運用能力の学習者、合計64名にスピーキング課題として課し、Hawkins and Filipovic (2012) が記述した22の基準特性が当てはまるかを検証した。その結果、先行研究結果と一致したのは、2つの基準特性のみであった。また同等と思われる2つのタスク間で、5つの基準特性において違いが見られた。これらの差異はおそらく、Hawkins and Filipovic (2012)が書き言葉データを使用したのに対し、本研究は発話データを利用したこと、またサンプル数（特に高いレベル）が少なかったことによると思われる。今後は、使用語彙や正確さのレベル間の比較や、スピーキングの基準特性を検討する必要性を発表者は示した。

質疑応答では、task effectsと学習者の運用能力の交互作用があった可能性が指摘された。また、発話にかかった時間と基準特性に何らかの相関がなかったのか、という質問が挙がった。リーディングのtask difficultyの研究は既に存在するため、それを比較、参照することも提案された。

金子恵美子（会津大学）

### **Folk Judgments of Second Language Oral Communication Ability: A Preliminary Study**

**Takanori SATO** (*University of Melbourne*)

In an exploration of potential discrepancy between the rating criteria for oral performance in the commercial English tests and the aspects of oral performance that the general population might pay attention to in the real world, this study has investigated the perceptions of “folk (i.e. non-linguistic professionals)” on the video-recorded oral performances of 10 test-takers of the national College English Test-Spoken English Test of China. From the

analysis of the verbal reports and interviews of the four postgraduate students (non-linguistic majors), it was found that the folk paid relatively strong attention to fluency, pronunciation (accents), and content. Some of the comments also mentioned non-verbal behaviors, demeanor, and interactional behaviors. The folk appeared to have paid much less attention to the aspects commonly found in the rating criteria in commercial tests (e.g. grammar and vocabulary).

The results of this study suggest that the rating criteria in the commercial English tests may not be adequately addressing the aspects that are valued in real-life oral English communication. The questions from the floor during the Q&A session mainly addressed the potential issues in the definition and selection of the “folk” as well as the practical implications that this study might have on the field of language testing. The study results are preliminary because it is a work-in-progress based on the presenter’s PhD research, and the audience gave him a round of applause with hope to hear about the further results in the near future.

Chihiro INOUE

(*Asahikawa Medical University*)

### **What Makes Peer Comments More “Formative” in Speaking Activities?**

**Hidetoshi SAITO** (*Ibaraki University*)

The purpose of the study is to find effects of peer assessment on students’ learning. The research was conducted at a junior high school, 151 students participating, and 784



peer comments were collected. Those comments were analyzed using simple descriptive statistics, correspondence analysis, and correlation and regression. For the analysis, the comments were put into three evaluation categories; “praise”, “advice” and “mixture of both”. The comments were also coded using In Vivo coding, which is descriptive categorization. Time (early, mid, and late periods) and students’ achievement scores were also used as covariates. Descriptive results of evaluation coding showed that the majority of the comments belonged to the “praise” category in all periods. The correspondence analysis indicated that “mixture of praise and advice” seemed associated with the early period, and “advice” with the last period. In terms of In Vivo coding, students made more comments related to the assessment criteria, such as “voice volume”, “eye-contact”, “comprehensibility”, and “fluency”. Those comments, however, decreased as time passed. On the other hand, “comments on continuing speech” and “asking and answering questions” increased. The students’ achievement scores and In Vivo coding were analyzed, but there was almost no correlation ( $r = -0.037$ ). Based on the results, two suggestions were made to make peer assessments more formative--use of multiple comments by students, and giving multiple opportunities for peer assessment.

In the question-and-answer session, the reason for choosing the peer assessment criteria was asked. They were chosen according to what junior high school teachers considered to be important in

their English classes. There also was a comment that students who were not good at English seemed to hesitate to make comments to students who had higher English abilities. There were more questions and comments from the floor, which showed strong interest in this presentation.

Makoto FUKAZAWA (*Ibaraki University*)

### 第3室

#### **Examining the structural validity of the achievement test of English lessons at public elementary schools in Japan**

**Keita NAKAMURA**

(*Eiken Foundation of Japan*)

This presentation reported part of the validation study of the newly developed achievement test for young learners at Japanese public elementary schools. The test, consisting of 44 listening items, is designed to measure the basic knowledge of English vocabulary, English phrases, and Q & A expressions learned in the classroom. 1039 fifth grade students and 988 sixth grade students from Japanese public elementary schools participated in this study. Based on the result of the Rasch model analysis, it was found that the test could successfully discriminate students with different English proficiency. In addition, a series of factor analysis was conducted in order to examine the structural validity of the test. As the results of analyses, it was found that the higher order model could best explain the test structure and the score feedback (i.e., listening ability composed of the abilities of vocabulary, phrases, and Q & A expressions). The argument for the structural validity of

the achievement test was partially supported by these results. In the question and answer session, one of the participants asked the purpose of the study. The presenter answered that the main purpose of the study was to investigate how appropriate the test was for measuring the achievement level of the English activities at Japanese public elementary schools. Also, the presenter agreed that some follow-up studies to further investigate the test are also needed.

Shuichi TAKAKI (*Graduate School  
University of Tsukuba*)

### **The national achievement test of primary English in Korea: Two different perceptions**

KyuNam SHIM (*Cheongju National  
University of Education, Korea*)

This presentation was about how Korean stakeholders perceive the development and implementation of the national achievement test of primary English (NATPE). The NATPE is a 50-minute mostly-multiple-choice test purportedly covering four skills, taken by 6th graders in July every year. Survey data were collected from 691 students from 9 middle schools in 3 different cities, who took the NATPE in the previous year, 75 head teachers or deputy head teachers from 50 primary schools in the same area, and 10 NATPE item writers.

Examination of their responses revealed the following: (a) Generally all the stakeholders hold their own firm beliefs regarding NATPE; two-thirds of all the participants accepted that the NATPE was worthy of implementation while a

considerable number of respondents regarded the test as unnecessary. (b) The impact of the test was not so big as was expected. Teachers were much more concerned with the test than the students were.

Based on these results, Dr. Shim concluded that in order to make the NATPE more successful, positive washback effects should be enhanced and negative ones minimized. To make that happen, Dr. Shim emphasized, educational authorities and testing experts should make efforts to help teachers to understand why NATPE is necessary, should provide teachers and students with guides and materials, and should improve the report scheme.

One of the post-presentation questions from the floor addressed the discrepancy between the high quality teaching and not-so-communicative testing in Korea. Dr. Shim answered that the authorities should continue to make efforts to narrow the gap.

Tetsuhito SHIZUKA  
(*Daito Bunka University*)

### **Language testing competence in the Iranian National University Entrance Examination**

Alireza AHMADI (Shiraz University)

This paper was cancelled.

### **英語外部テストを利用した単位認定の波及効果：英語履修免除学生の英語学習状況と満足度**

島谷 浩 (熊本大学)

本発表は、英語外部テストとしてTOEICが単位認定に利用されている状況と問題点を概観し、さらに実際に外部テストによる単位認定を受けた学

生に実施したアンケート調査の結果を報告することにより、この制度に対する提言を行うものである。

T O E I Cによって単位が認定される大学は2008・2011年度共に40%を越えているが、スコアと認定単位数の関係は大学により大きな差異があり、運用の仕方は多様である。このような現状において、単位認定を受けた学生はこの認定制度についてどう考えているのか、発表者は2009・2012年にアンケート調査を行った。対象科目は1・2年の教養英語、アンケート調査を行ったのはその学生が4年生になった時点で、回答者数は2009・2012年それぞれ13名と30名である。

アンケートの質問は、単位認定申請時期、外部テストのスコア、認定された科目名と単位数等の基本事項の他、認定後の英語学習状況や空いた時間の活用法、上級レベルの英語科目受講希望の有無、認定制度や教養英語に対する意見を問うものである。

回答は「認定制度に大半が肯定的」「51%の学生は自主的な英語学習を継続」「空き時間は専門の勉強などに活用」「33%が上級レベルの科目を希望」とまとめられるが、現状の問題点として「外部テストで測定されていない能力（会話能力など）も認定してしまう」「より高度な授業が提供されていない」が挙げられた。学生の回答にも「高校までの学習で得た知識によって大学の講義を受けなくて済む制度は、大学の英語教育の意義を考えさせる」とあるが、外部テストによる単位認定は「英語教育の存在価値に大きな影響を及ぼすものであり、その実施には十分な検討が必要」との提言がなされた。

小山由紀江（名古屋工業大学）

### 潜在ランク理論に基づくコンピュータ適応型テストシステムの開発と日本語語彙サイズテストへの適用 －シミュレーションによる評価－

秋山 實（東北大学大学院生）

松下 達彦（東京大学）

潜在ランク理論（荘島、2007）は、アイテムバンクを構築するためのプレテストに必要な受験者データ

が少なくとも、モデルにフィットしたアイテムパラメータをデータ数に応じた精度で推定することができる（秋山、2012印刷中）という特長を持っている。秋山は潜在ランク理論に基づくコンピュータ適応型テストシステム（LRT-CAT）をオープンソースソフトウェアのMoodleのモジュールとして開発した。このモジュールは、過去の受験データを用いてアイテムバンクを構成した場合のシミュレーションを行なうことができ、アイテム数が少ないアイテムバンクであっても最適な実行条件をシミュレーションによって確認し、設定することができる。また、LRT-CATは、その処理の大半をクライアント側（受験者のパソコン等）で実行するので、全ての処理をサーバ側で処理する従来の方式に比べサーバへの負荷が同時に受験する受験者の数に大きく左右されないという特長を持ち、プレースメントテストを一斉に受験させて実施する場合、Moodleの小テスト機能を使う場合と比べてほぼ同じ程度のサーバ側の負荷で実施できる。

松下は、日本語を読むための語彙量テスト（アイテム数150）を開発し、プレテストによって検証した（松下、2012印刷中）。リニアテスト（全員が同じ問題に回答する従来のテスト）では半数程度で十分な精度で語彙サイズの推定が可能であるとの感触を得ているが、さらに受験アイテム数を少なくできれば、利便性が高まると考え、LRT-CATによる語彙サイズ推定を試みることにした。

秋山は、予め日本語語彙サイズテストの回答データをEXAMETRIKA5.3（荘島、2008-2011）を使用して分析し、その結果を含むEXCELファイルでLRT-CATに読み込ませ、LRT-CATのシミュレータ機能を使用してシミュレーションを行なった。

#### 第4室

### コンピュータによる日本語スピーキング自動採点システム—日本語教師による採点との相関による検証—

今井新悟（筑波大学）

石塚堅吉（筑波大学大学院生）

山田武志（筑波大学）

篠崎隆宏（千葉大学）

西村竜一（和歌山大学）

酒井たか子（筑波大学）

菊地賢一（東邦大学）

本発表は、日本語スピーキング能力をコンピュータによって測定するシステムの信頼性を検証するため、コンピュータによる評価と日本語教師による評価の比較に関する報告であった。テストの形式は、文読み上げ、選択肢読み上げ、空所補充、文生成、自由回答などがあり、これらの特徴量の有効性について、音響尤度、母語話者と受験者の発話タイミング距離、母語話者と受験者の基本周波数距離、キーワード数、語彙多様性、発話量、発話速度について能力推定値を計算し、さらに、日本語教師による採点平均値と比較している。その結果、コンピュータと日本語教師による採点結果は、文読み上げ（0.72）、選択肢読み上げ（0.84）、空所補充・文生成（0.72）、自由回答（0.88）という高い相関係数を示したと報告された。発表者は、日本語教師間のものにも匹敵するほど高いこれらの相関係数は、日本語スピーキング能力をコンピュータによって測定することが十分に可能であることを示したと言えると締めくくった。

発表のあとの質疑応答では、評価の信頼性をめぐって、「コンピュータか、人間か？」について非常に興味深く有意義な議論があった。日本語教師の評価が正しく、コンピュータによる評価をそれに近づけるべきか、コンピュータによる評価がより正確で、日本語教師の評価をそれに近づけるべきか、というものである。これは永遠の議論であるが、コンピュータの評価基準はより数量的で、日本語教師による評価の基準と内

容的に異なる部分があるので、比較することは容易ではないという流れで終了した。

藤田智子（東海大学）

### 統合タスクにおける文章と質問内容との構造的類似性がスピーキングパフォーマンスに与える影響

矢野賢（茨城県立日立第二高等学校）

本発表では、スピーキング面接テストで与えられる文章理解と質問内容との構造類似性が受検者のパフォーマンスに影響を与えるかどうかの検証とその結果が論じられた。与えられた文章やプロンプトに含まれる情報から、オブジェクトレベルの類似と高次構造レベルの類似に焦点を当て、(1)文章トピックとその内容を参考にする転移との関連性、(2)文章内容の転移と発話量、複雑さ、適切さとの関連性、(3)その転移と類似レベルとの関連性を、高校生53名を対象に、英検2級と準2級の面接問題を使用して調査した。分散分析や相関分析等を使用した検証の結果を踏まえ、以下のことが指摘された。(1)文章トピックにより転移レベルは異なるため、トピック親密度を更に調査する必要がある。(2)文章内容の転移は発話量、複雑さ、適切さを促進するものもあれば阻害するものもある。(3)転移は高次の類似レベルでは促進し、表面的な類似が存在しない場合は適切さを阻害する可能性がある。また、馴染みないトピックでは転移が起こらない可能性もあり、今後の研究が必要である。更に、低次構造レベルの類似ではあるがオブジェクト類似の場合は、回答理由説明の促進にはつながらないという指摘もなされた。

平成25年度より実施される高等学校新学習指導要領において、4技能を有機的に関連付ける統合スキル指導が奨励されている中、本発表は現場での統合評価設定を考察するという点で大変意義深いものであった。この研究が今後の現場でのスピーキング指導方法やテスト開発等への示唆を与えて行くことを期待したい。

宮崎 啓（慶應義塾高等学校）

## 大学生の文法用語認識度を調べるための基礎知識テスト

徳永 美紀 (福岡大学)

本発表では、英語の基礎力が定着していない日本人大学生のメタ言語（文法用語）認識度を検証した。簡単な語彙からなる英文の下線部や文全体を表す文法用語を語群から選ぶという形式のテストを行い、文法用語の項目困難度を検証するとともに、受験者能力推定値別の誤答の違いにも焦点を当てた発表であった。品詞に関しては、副詞と形容詞を混同している学生が多く、他の品詞も正確に認識出来ない学生が多いことが判明した。文全体の時制などにおいては、「if」や「will」、「have+過去分詞」などが認識の助けとなっていると同時に、「過去分詞=完了形」などという誤認も多いという結果であった。能力推定値の低い受験者は、主語や述語動詞の判別、人称代名詞の格の判別などにも困難がみられたと発表された。

さらに、文法用語認識度と英語習熟度の関係を示す為、対象学生の TOEIC Bridge®スコアが英語習熟度の目安として使用された。文法用語テストによる受験者能力推定値と TOEIC Bridge®の全てのセクションとの間に中程度の相関がみとめられ、リーディングセクションとの相関が最も高かった ( $r = .658$ )という結果であった。

フロアからは、授業での文法用語使用や文法指導そのものに関して是非を問う意見や、学生が中高のどの時点でなぜ英文法が理解できなくなったのかを検証する必要があるといった意見があった。発表者は、全ての文法用語が必要というわけではなく、最低限必要であると考えられる用語を学生が理解できているかどうかを教員側が知ることが大切であるという点を強調した。

木下 正義 (福岡大学)

## 物語文、叙述文エッセイの評価信頼度考察

宮崎啓 (慶應義塾高等学校)

本発表は、物語文や叙述文というジャンルにおける評価スケールと評価者間信頼性の検証の報告で

あった。英語ライティングの評価研究は、説明文や議論論文のようなアカデミックなジャンルを対象とすることが多いが、実際の教育現場では、高校や大学の初級・中級学習者は物語文や叙述文からライティング学習を始めることが多い。よって、論理展開が中心ではないよりパーソナルなエッセイの評価についての検証は非常に重要だと思った。

この研究では、8つの構成要素から成る評価スケールを使って、4名の背景要因が異なる評価者（日本人教員2人、ネイティブ教員2人で、それぞれ文学と英語教育を学んだ人が1人ずつ含まれ、外国人は英国人と米国人、日本人は更に教歴が短い人と長い人に分かれる）が30人分の物語文と叙述文を評価した。その結果評価スケールと評価者間の信頼性は高かったが、FACET を用いたラッシュ分析により、評価者と構成要素によって厳しさの度合いに差が見られることが分かった。採点后に評価者に求めたコメントを分析することにより、評価スケールの中の descriptors（記述子）の解釈が評価者によって異なったり、背景要因による見方の違いが差をもたらしていると推測された。

会場からは、事前にトレーニングをすることでかなりこのような差は解決できるのではないかという意見や、コメントを書いてもらうだけでなく詳細なインタビューをすれば、構成要素と評価者の背景要因との関係性をもっと見えてくるのではという指摘があった。

物語文や叙述文のジャンルで構成要素と評価者の背景要素にまで踏み込んだ研究はあまりないので、今後はもう少し評価者を増やして背景要素をより制御した研究ができればとても将来性と教育的意義があると思う。

松本佳穂子 (東海大学)

## 日本人学習者ライティングコーパスに基づいた CEFR レベル別基準特性の分析

長沼君主 (東京外国語大学)

工藤洋路 (東京外国語大学)

根岸雅史 (東京外国語大学)

CEFR では各言語でレベル別のプロフィールとして語彙や文法の基準特性リストが開発されつつあり、英語では Cambridge ESOL を中心として、English Profile の開発が現在進められている。しかし、これとは別に日本人英語学習者に特徴的な基準特性 (例えば定冠詞の脱落) が報告されたり、日本で作られた JACET8000 と比べると難易度の逆転現象があることから、日本人学習者に特徴的な基準特性の抽出が必要であると指摘した。本発表では、この問題提起に従い、日本人学習者に特有の英語ライティングにおける CEFR 各レベルの基準特性を明らかにするため、Eメール、物語文、意見文の3つのジャンルの作文をコーパス化し、共起ネットワーク分析、コレスポネンス分析した結果から以下の結果を得た。

- i) 全体では A1/A2 レベルの語彙をほぼ使用しているが、意見文で A2/B1 語彙の頻度が高かった。
  - ii) 作文レベルによる頻度差は少なかった。
  - iii) 使用語彙の種類は、Eメール、物語文、意見文と高レベル語彙に多様性が見られた。
- 今後の課題としては以下が挙げられた。

- i) 基準コーパスとの比較の必要性
- ii) B1+/B2 学習者データの増加の必要性
- iii) 学習者のパフォーマンスを最大限に引き出すタスク開発の必要性
- iv) 低頻度データの分析の必要性

今回の報告は日本人学習者に特有の基準特性を抽出するための基礎研究と位置づけられるであろう。今後、本研究が English Profile に見られるような具体的な基準特性のリスト化に発展することが期待される。

今井新悟 (筑波大学)

## 第5室

### 能力記述文による自己評価

木村哲夫 (新潟青陵大学)

近年、受験生は様々な入試形態を経て大学へ入学するため、教員が学生の英語力を把握することは難しい。また学生もどの程度英語力があるのかわからず、どのように学習していいのかわかっている学生は多い。英語力を客観的に自己評価でき、自分の学習計画や目標が立てられるように、と行われたのがこの研究の背景である。自己評価の対象に選んだのは STEP の Can-do-statement (CDSs) で、これは英検 1 級から 5 級合格者、約 2 万が英語を使用する際の自信の度合いに基づき作成されたものである。この指標を用いて学生自身が自分の英語を自己評価したものを潜在ランク理論で用いて分析したところ、自己評価と CDSs には非常に高い相関がみられた。つまり、CDSs は自己評価のツールとして機能することが明らかになった。しかし、これとは別に実施された英語テストによる結果を比較したものを各技能別に比べてみると 60% は CDSs と一致したが、リーディングに対して過小評価し、リスニングに対し過大評価する学生がいることが分かった。

この研究は学生にとって身近な英検の CDSs が自己評価のツールとして十分機能することを明らかにし、過大 (過小) 評価した学生へ客観的なフィードバックを与えることで自分の能力を知り学習計画を立てさせるきっかけになるであろう。なぜならば、このような具体的なフィードバックを与えることにより、学習者は自分の英語力を知ることにより学習への動機づけにもつながるからである。診断的に CDSs とテストを両方用いることにより学習者を支援していく試みを示した非常に貴重な研究である。このように 'learner friendly' なテスト研究がこれから増えていくことを切望する。

秋山朝康 (文教大学)

## ビジネス分野の can-do statements による自己評価の分析—言語能力と業務経験の関係から—

伊東田恵 (豊田工業大学)  
川口恵子 (芝浦工業大学)  
太田理津子 (慶応義塾大学)  
佐々木裕 (豊田工業大学)

本発表では、TOEIC 得点を指標としたビジネス現場での can-do リスト(Ito, Kawaguchi & Ohta, 2005)に基づき、ビジネス環境の変化に対応して項目を見直した上で、低得点者層へのデータの偏りを補正すべく新たに取直したデータをもとに、自己評価結果の比較と経験要因の自己評価への影響を検証した。また、テキストマイニングを用い、自由記述の分析も行った。

自己評価アンケートの改訂にあたっては、ビジネスと関連したタスクに絞り込み、出張や訪問などの旅行場面に関する一般的タスクは削除された。代わりに時代のニーズに合わせ、テレビ会議や異文化理解に関する項目を追加した結果、can-do statements は 65 項目から 45 項目となった。タスクは 6 つのビジネス場面 (来客対応、日常業務関連、電話、発注・支払・クレーム、商談・プレゼンテーション、専門業務の会議・講演)、5 つのスキル領域に分類された。

経験者と非経験者で自己評価の差の大きかったタスクと小さかったタスクを各 15 項目ずつ、得点上位者と下位者に分けて分析した結果、高得点者では、商談・プレゼンテーションや会議・講演などの発信スキルで経験による差が大きく、低得点者では日常業務、電話、発注・支払い・クレームなどのより一般的タスクで 4 技能に渡って差が大きかった。質疑では、得点による業種の偏りの影響なども指摘されたが、業務シミュレーションなどを通して、自己評価が変化するか、言語不安が減少するかなど、今後の研究が期待される。

長沼君主 (東京外国語大学)

## CEFR 読解尺度に基づいた日本語読解テスト開発の試み

谷誠司 (常葉学園大学)

本発表はキャンセルされました。

## 英語教科書におけるリーディングの内容理解確認問題は何を測っているのか

高木修一 (筑波大学大学院生・日本学術振興会 特別研究員)

高木氏の発表は、中学校英語教科書本文の内容理解確認問題 (英語による設問) に焦点を当てて、それが生徒にどのような読む能力を求めているかを分類しようとしたものである。分析の枠組みは(a)パラフレーズ、(b)推論、(c)テーマ、(d)照応、(e)語彙、(f)テキスト構造に分類した清水(2006)に準拠した。分析対象は、採択率の高さに従って選定された 4 種類の新学習指導要領準拠中学校用検定教科書であった。

すべての質問がパラフレーズに分類されたために、(1)設問と解答の情報の重なり (設問の主語・動詞と解答の主語・動詞の重複)、(2)問題形式 (問題が TF 型かどうか) の 2 観点から追加分析を行ったところ、約 70%の問題で設問と解答の主語・動詞が重複し、約 30%の問題が TF 型の設問であるという結果を得た。主語・動詞の重複に関しては、全体的には設問と解答で情報が重複する傾向にあるが、個別にみると NC、SS で情報の重複がない問題が 3 年生用で多くなる傾向があることが分かった。問題形式に関しては、教科書間で傾向が異なることが報告された。

フロアーからは、教科書の問題は何かを測定するものではなく、内容理解を助けるために設けられたものであり、中学校段階では言語材料が限られているので字義的な意味が理解できれば十分ではないかという意見や、授業中に教師が内容理解問題をどう活用しているか、それを生徒がどう捉えているかを調べてはどうかといった示唆があった。

島田勝正 (桃山学院大学)

## VELCテストによるTOEICスコアの予測: リスニングとリーディングについて示唆されるもの

静哲人 (大東文化大学・VELC研究会)

日本人大学生向けの熟達度テストとして開発されたVELCテストの概要を示した上で、利用者からよく質問されるVELCテストのスコアとTOEICのスコアの関係について、TOEICのリスニングスコア(T\_R)とリーディングスコア(T\_L)を別々の目的変数として、VELCの6パートのスコアを予測変数として行われた重回帰分析の結果と、両テストの8つのスコアについて行われた共分散構造分析(SEM)の結果について報告された。

VELCテストは、リスニングセクション60問、リーディングセクション60問、合計120問からなる所要時間約70分のテストである。両セクションは各20問ずつの3つのパートから構成される。詳細についてはVELCテスト(<http://www.velctest.org/>)のサイトを訪れることで、サンプル問題や結果通知のサンプルを見ることができる。VELCテストはラッシュモデルによって開発・標準化されたテストであり、等化された複数フォームがあり、どれを受けても同一尺度上での結果比較が可能である。

重回帰分析で予測変数の絞り込みが行われ、T\_LはVELCのリスニングセクションの3つのパートのスコアで57%が予測可能であり、それにリーディングセクションのパートのスコアを加えても予測はほとんど改善されないのに対して、T\_RはVELCのリーディングセクションのパートだけでは60%であるが、リスニングセクションのパートもすべて使って予測すると64%予測可能であることが示された。この2つの予測の合計によりVELCテストの結果からTOEICのトータルスコアは68%予測可能であることが報告された。また、SEMによる複数のモデルが試されたが、最も解釈しやすく適合度がよいのは3因子モデル(リスニング因子、リーディング因子、語彙因子)であった。今回の重回帰分析とSEMから得られた知見は、VELCテストが十分な妥当性を有するテストであることの例証と考えられる。

木村哲夫 (新潟青陵大学)

## Symposium

### Evaluating extensive reading at school

Coordinator : Yo IN'NAMI

(Shibaura Institute of Technology)

Panelists:

Atsuko TAKASE (Kansai University)

Rob WARING

(Notre Dame Seishin University)

Discussant: David BEGLAR

(Temple University Japan Campus)

Professor In'nami opened the symposium with a brief description of extensive reading (ER) and presented shared issues, for example, what to measure and how to measure it. Then, he introduced the speakers for the symposium.

The first speaker, Professor Takase explained what ER entails, in terms of reading in quantity, purposes of evaluating ER programs, and the goals of ER. She discussed various tests utilized in ER studies such as ACE, BACE, EIKEN, EPER, GETC, SLEP, and TOEIC. She also showed the audience a half a dozen ER empirical studies.

The second speaker, Professor Waring discussed ER research in action in cases where research emerges from a teacher's concern or problems. In addition, he shared some typical research questions regarding ER: "Do students read faster if they have easier books?", "Why do they go down difficulty levels as the semester goes on?", "Do positive reports from



previous ER classes lead to positive attitudes to ER?”, and “What are their favorite books and genre they read?”.

Finally, from the two talks, Professor Beglar asked about reading fluency and difficulty of assessing reading aloud. Specifically he asked what kind of guidelines to use for ER, for example, a book a week, to stimulate discussion. The answers to the question from the two professors included that ER needs to have a good environment and should be applied early when students are young, but also instructors have to be careful of setting a goal, because once learners have reached the goal, he or she might stop reading.

During the Q and A session, participants commented about the problem of translation and concluded that students should start reading before their exposure to translation methods if instructors want to avoid the problem. Participants also questioned why ER is performed silently, what instructors think of intensive reading, and the need for it and so forth. Professor Waring suggested that if learners read every word aloud, it limits reading speed. The need for intensive reading and motivation to read were also part of the discussion.

Takayuki Nakanishi (*Tokiwa University*)



## 日本言語テスト学会 第4回最優秀論文賞 結果と経過

### 2012（平成24）年度日本言語テスト学会最優秀 論文の審査の過程および結果報告

JLTA 最優秀論文も本年度で第3回目を迎えた。候補論文はどれも力作揃いである。5名の委員が1カ月を懸けて全て丁寧に読み、独創性（originality）、学術研究への寄与（contribution to the academic field）、教育への寄与（contribution to education）の3つの基準で評価する。本年度は9篇の候補論文のうち、片桐一彦氏（専修大学）の Longitudinal Measurements and Developmental Patterns of Receptive Vocabulary Size : A Study of Japanese University EFL Students が受賞した。大学生の語彙力の経年変化を調査し、1年次には変化しないが、2年次になって増加すると言う傾向があること等有益な発見を報告したもので、委員全員一致の上理事会で承認を得て本論文を受賞論文とした。総会終了後授与式が行われ、記念の盾と賞金が贈られた。それに引き続き片桐氏が受賞の言葉を述べられた。今後益々のご活躍を期待するとともに、次年度も多くの方々が御寄稿下さるようお願いする次第である。

報告者

2012年度第4回 JLTA 最優秀論文賞委員会  
委員長 渡部良典（上智大学）



<編集後記>

全国研究大会報告は今号で最後となりました。ニュースレターにふさわしい記事を広報委員会がこれから探していかなければなりません、ぜひこのようなものをというご提案を歓迎するとともに、様々な原稿(博士論文の概要、書評、論考なども)も歓迎いたしますのでまずご連絡ください。

日本言語テスト学会事務局  
〒389-0813 長野県千曲市若宮 758  
TEL 026-275-1964 FAX 026-275-1970  
e-mail: youichi@avis.ne.jp  
URL: <http://www.avis.ne.jp/~youichi/JLTA>



編集： 広報委員会

委員長 齋藤英敏 (茨城大学), 副委員長 長沼君主 (東京外国語大学)

委員 秋山實 (東北大学大学院/株式会社eラーニングサービス),

片桐一彦 (専修大学), 佐藤臨太郎 (奈良教育大学), 宮崎 啓 (慶應義塾高等学校)