

JLTA Newsletter No. 29

日本語テスト学会

The Japan Language Testing Association

JLTA Newsletter No. 29 発行代表者: 浪田克之介 2009 年 (平成 21 年) 11 月 21 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 URL: <http://jlta.ac>



巻頭言: JLTA からの恩恵

法月健 (静岡産業大学)

JLTA Newsletter 巻頭言の執筆依頼のメールをいただき、一瞬目を疑い、驚き、戸惑ったが、謹んでお引き受けすることとした。

2009 年の全国研究大会は北海学園大学で開催されたが、1991 年の北海道大学に於ける JACET 全国大会で実質的な学会デビューをした私にとって、18 年間の月日を経て、思い出の札幌の地で、自分を育ててくれた JLTA の大会に参加し、JLTA で知り合った先生方と共同の研究発表ができたことは、真に感慨深いものであった。

1996 年の第 1 回 JLTA 全国研究大会以降、全国研究大会研究発表の司会業務、第 9、10 回全国研究大会、第 18、30 回の研究例会の運営等に従事する機会をいただき、JLTA から多くのことを学び、励まされ、様々な貴重な体験を得てきた。この 13 年間、あるいは 18 年間に、研究者や教育者として誇れるような業績はほとんど残せなかったが、恩師、大友賢二先生から言語テスト研究のご指導をいただいていたなければ、JLTA に入会していなければ、今頃は目標意識に欠ける単調な日々を送っていたかもしれない。それ故、JLTA や JLTA 関係の方々からのご依頼は少々荷が重いものであっても、できる限り、お引き受けするようにしている。

今、私の大きな関心の一つは、言語テストの研究や言語テストに関連する学習の支援を通じて、ゼミや教職課程の学生の自己啓発力を育成することにある。幸い、英検や TOEIC の勉強を通じて、ゼミ生たちは、時に挫折感を感じながらも、自らの可能性を信じ、仲間との連帯意識を強めながら、自身を高めていく気概に満ちている。単位取得すらおぼつかない学生は、私の研究に被験者として参加したことを契機に、英語学習や研究に打ち込む模範生となり、下級学年の頃から、ゼミ生に交じって英検や TOEIC を勉強していた学生は、卒業研究に言語テスト研究を選び、来年度から大学院に進み、さらに研究を追究することになった。「言語テスト」を介して、学生たちと心の架け橋を築く喜びも、究極的には JLTA から受けた恩恵なのかもしれない。

**第13回 日本言語テスト学会
全国研究大会 研究発表**

於：北海学園大学

平成21年9月7日（月）

VERSANT™ 試験活用におけるコミュニケーションデザイン構築に向けて
上村武司（株式会社ピアソン・エデュケーション）・鈴木正紀（Pearson Knowledge Technologies）・井狩幸男（大阪市立大学大学院文学研究科）

まず、VERSANT™について説明がされ、その後、大阪市立大学での実績が報告された。

VERSANT™は、受験者が英語をリアルタイムにどの程度理解し話すことができるかを、電話（或いはPC）によって受験するテストで、音声認識を用いたコンピュータシステムが即時に判定し、スコアを出す口頭英語能力テストシステムである。また、CEFR、TOEFL、TOEICとのスコアを対比することも可能であり、広く企業や教育機関で利用されている。

海外の企業、アメリカの大学、日本の企業、日本の大学における受験者のテストスコアの累積度数曲線の例が紹介された。アメリカの大学では留学生がTAになるために受験し、日本の企業の例は毎日のように英語を使う環境にある企業であったので、比較すると日本の大学生が最も低い能力を示す例であった。

大阪市立大学の英語教育開発センターでは、英語コミュニケーション能力、特に発信能力（ライティングとスピーキング）習得をめざす授業のプレースメントテストや標準テストとして

VERSANT™を導入している。これで学生を習熟度別クラスに分け、成績を決定する資料にすることは授業の内容とテストの整合性があり非常に理にかなっていると思われた。またテスト結果を分析すると、4月から7月にかけての第

一学期のスコアの伸びが、それ以降より大きく、学生たちがコースの初期にコミュニケーションな英語能力を伸ばすことに注力したことが伺えた。

VERSANT™の評価について Fluency はどのように評価されるのか？という質問に対して、評価項目のなかで音読の配点は小さいとの返答があった。

藤田智子（東海大学）

オンラインテストの回答時間で習得度を測る試みー英語の語彙・文法に関するテストの回答時間の場合ー

秋山實（東北大学大学院／株式会社eラーニングサービス）

本発表は、2つの大学の学生によるオンライン化されたプレースメントテストの回答データのうち特に回答時間に注目し、回答時間から各項目の習得度を測ろうとする試みである。検討されたのは、高校卒業レベルを対象としたプレースメントテストの文法、語彙のセクションである。大学Aの学生とAよりも英語の熟達度の高いと思われるBの学生の結果を比較したところ、回答時間の平均および標準偏差ともにBはAの約半分という結果が得られ、習得度が上がると回答時間が短く、分散も小さくなると考察された。さらに、回答操作に要する時間を考慮するため、オンライン化に携わった者3名が1人20回受験した際の回答時間のデータも利用されることで習得度を測る基準となる回答時間の設定の可能性が示唆された。そして、これらを踏まえて、基準時間内の回答かつ正答が習得とみなされ、回答時間のモデル化について提案された。今後の課題として、再選択の時間の扱い、モデル化、短すぎる／長すぎる回答時間の扱いなどが挙げられた。最後に回答時間の応用例として、即答かつ正答をN回できると習得というように判断可能な暗記カードが紹介された。フロアからは、中高の指導に役立てるための文法項目ごとのデータ提供についての要望や議論の前提

となると考えられるプレースメントテストの結果と回答時間の相関についての質問などが出された。

滝沢雄一（福島大学）

「Moodleで簡単eテスト」を実現するテストスイツの開発

中原敬広（株式会社eラーニングサービス）

本発表では、Moodleで教師の負担を軽減しながらCATを行うための様々な支援ツールの開発についての紹介、および実際にそれらを利用してCATを作成する流れについて説明された。テストングツールとしては、発表者が開発に関わった、テスト分析プログラムのTDAP3.01、アダプティブテストを実現するCATモジュール、多肢選択式のアイテムを作成するMultipleChoiceMaker、語彙レベル判定機能を有するJ-EditおよびE-editアイテムエディタ、Cloze形式のアイテムを作成するClozeエディタ、語彙などの学習用「暗記カード」などについて、それぞれの主な機能や旧バージョンからの改善点などの説明があった。その後、これらのツールを用いて実際にCATを作成した際の流れについて説明がなされた。作成は「問題の作成（MultipleChoiceMaker）」「予備テスト」「分析（TDAP）」「テストレット作成（TDAP）」「CAT作成（CATモジュール）」「シミュレーション」「終了時間の確認」の手順で行われ、作成上の具体的な課題等についても言及があった。発表後フロアからは、主にテストレットに含めるアイテム数についてや能力値の設定などCAT作成上の作業について質問やコメントが出された。

滝沢雄一（福島大学）

よく読める学習はよく話せるのか：

TOEIC®テストおよびTOEIC®SWテストを用いた日本人英語学習者の4技能能力値の関係

石川慎一郎（神戸大学）・石川有香（名古屋工業大学）・水本篤（流通科学大

学）・三橋峰夫（国際ビジネスコミュニケーション協会）

本発表では、108名の日本人大学生のデータに基づき、TOEIC®における4技能間の関係、および学習者属性の影響が詳細に検討された。被験者はTOEIC®テスト、TOEIC SW®テストの2種類を6ヶ月間以内に連続受験し、さらに語彙テストおよび性格、動機、学習経験、英語使用経験の各種アンケートに回答した。分析は、被験者全体および4つの学力群に分けて行われた。

その結果は、(1) 変動係数による弁別度はスピーキングが高く、中位群で特に高い、主成分分析によればスピーキングが総合得点への貢献度が高い、(2) 4技能間の相関は、どの組み合わせでも上位群と下位群では高いが中位群、準上位群ではバラツキがでる、(3) 回帰モデリングを用いた総合得点によるスピーキング・ライティングスコアの説明力 (R^2) はそれぞれ0.6, 0.45で、総合得点が高い者はおおむねよく話せるがよく書けるとは限らない、(4) 受容語彙サイズとスピーキングにやや高い相関が観察されたことは興味深い、下位群では語彙力が、リーディング力と密接に関係するが、上位群では語彙力はむしろスピーキングとの相関が高い、(5) 因子分析とクラスター分析の結果、性格要因、動機要因および学習経験要因は、テスト得点との関係が深いとは言えない、の5点に要約できる。

これらの分析から4技能は4つのテストで直接的に測定すべきであるという総括があった。フロアからは学力群分けによる被験者数の減少に伴う切断効果の問題点の指摘や、学力群の数に関する質問があった。

島田勝正（桃山学院大学）

錯乱肢とテキストの分析による新旧 TOEIC® の比較検証

法月健（静岡産業大学）・伊藤彰浩（西南学院大学）・島谷浩（熊本大学）・木下正義（西南学院大学非常勤）

2006 年に改訂された新 TOEIC® の問題形式に見られる特徴は、テキストの長文化及び一問題あたりの問題数の増加、アクセントの多様化、誤文訂正問題の廃止と空所補充問題の追加の 3 点であるという。この改訂に関連し、136 人の日本の大学に在籍する学生に旧・新の順で 2 つの形式の TOEIC® テストを実施し、古典的テスト理論、項目応答理論に基づいて新旧のテストデータを比較分析した。

その結果は、(1) 新テストの平均点が旧テストよりも向上した、読解問題の平均はほぼ同等なので、新旧テストの得点差は聴解問題の容易化に起因する、新旧テスト間の相関は読解、聴解とも類似し、信頼性もほぼ同様である、(2) 受験者能力値に対して項目難易度の分布は新テストの方が広がっている、(3) 新テストにおいて聴解に項目困難度、項目弁別力が低い項目が増え、不適合項目が多い。(4) 旧テストと新テスト間で一部の受験者に大きな得点の変動があり、それは聴解にも読解にも見られる、と報告された。

次に、読解に関してうまく機能しなかった項目の事例を挙げ、新テストでは使用単語が難しい、状況が理解できない、ダブルパッセージの問題で推論を必要とするなどの原因を特定した。また、聴解に関しては、テキスト分析の結果、使用語数や品詞の使用頻度が新旧で異なる、新テストに主語が同一の選択肢（人称代名詞）の多用が見られ、直接的な解釈を求める設問が多い等の報告があった。

フロアからは、複数のバージョンを用いて比較すべきではないか、錯乱肢とテキストにより特化した分析が望ましいのではないか等の指摘があった。

島田勝正（桃山学院大学）

Establishing guidelines on how to select databases for meta-analysis in applied linguistics

Yo In'nami
(Toyohashi University of Technology)
Rie Koizumi (Tokiwa University)

This paper was an attempt to establish practical guidelines on how to select databases for meta-analysis as a statistical approach to integrating the results of separate empirical findings in the field of applied linguistics. Three research questions were raised and examined, regarding what databases are used, how they are used in combination, and which databases provide a comprehensive coverage of 24 representative journals in the field. It was claimed that ERIC, LLBA, ProQuest Dissertations and Theses, and PsycINFO are most commonly used and the combined use of these four databases is to be recommended. Meanwhile, ERIC, Linguistics Abstracts, LLBA, MLA, and Scopus were found to provide a comprehensive coverage in terms of the coverage rates of different journals and different periods of publication. It was concluded that optimal number of various databases should be used ideally, depending on one's research purposes and needs, in order to preclude the meta-analysts' classification errors, at least two of the seven databases indicated above should be crosschecked, among which ERIC and LLBA are the most essential databases in the field, and the use of other types of databases, such as the search engine available on journal web pages, CD-ROM versions of journals, and those databases only accessible from the university libraries with special permits, might be taken into account for specific search. Active questions and comments from the floor reflected a strong interest in the use of databases not only as tools of meta-analysis but for the general purposes of literature review as well.

Ken Norizuki (Shizuoka Sangyo University)

Comparing ratings of a face-to-face test and a telephone-mediated speaking test.

Yujia Zhou

(Tokyou University of Foreign Studies)

This study examined the validity of TSST test (ALC Inc.), a telephone-mediated speaking test, by comparing it to its face-to-face predecessor SST (standard speaking test), which is a version of ACTFL OPI. Using data collected from 81 participants who took both tests (Kaneko, 2004), the author compared perfect agreement ratios and correlations between ratings and found that the TSST estimates were comparable to SST estimates. Non-significant t-tests results confirmed this consistency between the two tests. Two remaining issues are the use of ratings (NOT raters) and arbitrarily numeric SST levels conversion. This conversion followed Dandonoli and Henning's (1990) study, but there were no clear basis for this change. Although the study supports the validity of TSST, further investigation with a new set of data is called for.

Hidetoshi Saito (*Ibaraki University*)

Does awareness of listening course objectives change student's self-can-do assessments?

Fujita Tomoko (Tokai University)

This study examined the relationship between the understanding of course objectives and changes in responses to can-do self-check lists of a new listening course in a university language program. Students from three different course levels (n = 230) underwent a series of four special lessons that focused on listening of the course objectives, while a control group underwent regular lessons. Both groups completed a self-check list containing listening can-do statements. The author reported notable increase in some can-do statements (e.g., conversation) in the experimental groups. This suggests that the understanding of the course objectives may lead to increase in certain can-do items in the course. However, this effect varies across different course levels, the relationship between understanding of

course objectives and responses on the can-do checklist seem negligible for lower level groups but significant for higher level groups. Some questions remain concerning the reasons for level and item variability and the possibility of "compliance" to can-do statements. The author plans to utilize longer treatment and achievement tests to reveal additional utility of can-do self-check lists for understanding learning gain.

Hidetoshi Saito (*Ibaraki University*)

Reconsidering the effectiveness of TBL in comparison with PPP in the Japanese EFL classroom

Rintaro Sato

(Nara University of Education)

Professor Sato commenced his presentation by comparing task-based language learning (TBL) with the presentation-practice-production (PPP) model. Although TBL has been regarded in SLA as a very effective device for honing learners' communication skills, Sato cautioned against the indiscriminate use of TBL. He mentioned that the effectiveness of a task is limited if the interaction in the target language is minimal. He cited another reason—the mismatch between TBL and examinations. Even though a student can substantially improve his/her communication skills through TBL, s/he would be dismayed on discovering that such a skill is not tested in in-house or university entrance examinations. Further, he stressed that an overemphasis on task completion in TBL would lead to the possibility of a student engaging in a task without heeding the grammatical form. This conclusion was based on his small-scale study where 15 of the 21 participants did not produce the target form during the task performance.

The audience posed several questions. In response, Professor Sato stated the following: Tasks are effective in improving learners' motivation levels, thinking abilities, creativity, imagination, and communication skills provided there is sufficient input and output; practice in PPP

does not always lead to learning or acquisition; according to the purpose of the class, tasks can be used effectively in the production stage of the PPP. Overall, the presentation was very well received, and the question and answer session was insightful.

Yo In'nami (*Toyohashi University of Technology*)

Constructing and validating a questionnaire: Using Rasch PCA and Confirmatory Factor Analysis

Keita Kikuchi (Tokai University)

In this presentation, Prof. Keita Kikuchi presented an idea on the process of questionnaire development. In his research, a 40-item questionnaire on demotivators was developed to elicit data from more than 1,200 high school students. As a result of the both Rasch PCA factor analysis and Confirmatory Factor Analysis, six originally hypothesized constructs (teachers, characteristics of classes, experiences of failure, class environment, class materials, and learners' interest) were re-combined into four factors (teacher behavior, experience of difficulties, class environment, and lack of interest). He concluded that a considerable amount of time should be spent in designing a questionnaire before it is administered. In a question-and-answer session, several issues regarding the research method and data analysis were discussed. Finally, Professor Kikuchi showed that his next study will examine whether questionnaire development would be confirmed using actual test data. It seems that designing a questionnaire needs to be scientifically investigated, so I hope the progress of this study will be reported in the near future.

Hiroshi Shimatani (*Kumamoto University*)

Developing an in-house placement test for measuring students' English proficiency

Yuji Nakamura (Keio University)

**Haruhiko Mitsunaga
(Tokyo Institute of Technology)**

Nakamura and Mitsunaga reported on the result of the in-house placement test that was developed in order to meet the needs of the Faculty of Letters of Keio University. The test consists of four sub-sections made up of grammar, vocabulary cloze and reading comprehension, containing a total of 50 multiple-choice items. The test has been administered to approximately 800 students at a time twice a year since the spring of 2006. The validity and reliability were established using the Rasch-based statistic program, and the changes in students' reading ability were traced over three years. Based on these results, a suggestion was made also for developing and implementing an item bank by means of equation. Amongst many useful findings and suggestions for the future research, perhaps one of the most insightful is that the two-parameter logistic model can be used as a powerful tool for exploring the item characteristics of tests measuring language ability. The presentation was followed by a Q&A session, where attendees asked pertinent, timely and engaging questions.

Yoshinori Watanabe (*Sophia University*)

Developing diagnostic test for teaching English in English

Kyung-Ae Jin

(Korea Institute for Curriculum & Evaluation)

In this presentation Dr Kyung-Ae Jin reported on the reliability and validity of the test that was developed to measure the ability, skill and knowledge of Korean teachers teaching English as a foreign language. The test consists of two modules. ETKT (English Teaching Knowledge Test) is a module for testing teachers' knowledge that is required to carry out an informed instruction, including the knowledge about teaching methodology, linguistics, curriculum development, and so forth. ETPT (English Teaching Practice Test) is the test for measuring teachers' performance in their classrooms. ETKT was administered to a total of 247 in-service teachers teaching English at primary school (N=127) and secondary

schools (N = 120), whereas ETPT was piloted on a total of 6 both primary and secondary schools. In addition to stabilizing the reliability and validity of the test modules, the standard setting was conducted based on the results by using a modified Angoff method. The presentation greatly made the audience aware of the importance of developing a contextually appropriate test measuring English teachers' ability, which in turn will be applicable to other contexts where English is taught as a foreign language.

Yoshinori Watanabe (*Sophia University*)



Developing and Implementing the National Assessment of Educational Achievement (NAEA) in Korea

Min-young Song

(*Korea Institute for Curriculum and Evaluation*)

Prof. Song gave a presentation on "Developing and Implementing the

National Assessment of Educational Achievement (NAEA) in Korea. According to Prof. Song, since 1998, NAEA in Korea has been administered to small portion of elementary and secondary students to diagnose their educational achievements and provide fundamental reference data for improving the National Curriculum, and the result has not been released to the public. However, Prof. Song says that as accountability of educators, schools, and school districts has been raised as an important social issue, it has been decided that from 2009 on, all the 6th, 9th, and 10th graders are supposed to take the test and information about school performance on the test should be provided for the public. Prof. Song discussed many significant changes in the procedures of developing and administering NAEA, and some of the major changes, especially focusing on the English subject. There was a fruitful discussion with questions and comments from the floor after the presentation.

Yuji Nakamura (*Keio University*)

Developing a Framework and Prototype of the Korean National English Proficiency Test for Secondary School Students

Oryang Kwon,

(*Seoul National University*)

Prof. Kwon gave a presentation on "Developing a Framework and Prototype of the Korean National English Test for Secondary School Students." According to Prof. Kwon, the project was motivated by two factors: (1) the necessity for accurately assessing the students' English abilities and (2) the necessity for qualitative improvement in students' English proficiency in order to meet the challenges of the rapidly changing world where international competitions are becoming tougher than ever before.

Prof. Kwon explained the purposes of the project as follows:

- 1) Establishment of criteria for secondary school students' English proficiency level in the system of the national curriculum
- 2) Development of a framework of measurement for secondary school English abilities according to the nationally accredited English language assessment
- 3) Development of a prototype of test that reflects the overall system of the nationally accredited English language assessment
- 4) Establishment of a system and direction to enhance the positive influence of assessment on education

Prof. Kwon described the nature of the test, the levels of the test, the skill areas to be assessed. He further showed the test development procedures, and the materials and types of linguistic input (i. e., spoken or written) for listening and speaking tests, as well as the criteria for rating speaking and writing samples. Prof. Kwon further talked about actual test items were written by professors, school teachers, and English native-speakers.

After the presentation, there was a lively discussion based on the questions from the floor.

Yuji Nakamura (Keio University)



「明示的知識・暗示的知識を測定するテストバッテリーの可能性と問題点」

島田勝正 (桃山学院大学)

本発表は、明示的／暗示的知識 (Explicit/Implicit knowledge)を識別するテストバッテリーとして、Ellis(2007)

モデルの妥当性について言及し、更に、より妥当なモデルを目指し 1) 処理過程の違い(Production/Judgment) 2)時間的制約(Timed/Untimed) 3)項目の正誤 (Grammatical/Ungrammatical) のそれぞれの特性がどのように関与しているかを Ellis(2007)の改良モデルを用いて日本人大学生 83 名に対して検証したものである。島田(2008)は、適格文(GR)と非適格文(UG)の 80 項目からなる、時間制約の有無による 2 種類に分けた文法性判断テスト、同一の 80 項目からなる誤文訂正テストを開発した。さらに、GR20 項目と UG20 項目から構成される口頭模倣テストを開発した。今回は、40 項目の口頭英訳テストを追加し、5 つのテストによる 8 つの変数間の相関関係を分析した。探索的因子分析の結果、時間制約の有無による 2 因子モデルよりも口頭産出、文法的適格性判断、文法的不適格性判断の 3 因子モデルの方が適合度が高いことが示された。

竹村雅史 (北星学園大学短期大学部)

「言語テストにおける段階評価の実際：入試とプレイスメントテストのデータ処理」

木村哲夫 (新潟青陵大学)

本発表は、順序尺度を想定したテスト理論として発表されたニューラルテスト理論(NTT)を紹介したものである。この理論 (Shojima, 2008) は、テストのデータを非連続の順序尺度として処理し、評価において段階評価を行うものである。測定方法論的、教育社会学的、教育現場の評価体制、品質管理・アカウンタビリティの 4 つの側面から段階評価の利点を述べ、テストは選抜やクラス分けに用いるだけでなく、どの程度どのような能力を持っているか、段階評価を導入することにより、各能力段階 (潜在ランク) の特徴を、Can-Do Chart との関連で示すことができ、言語テストの品質管理やアカウンタビリティの確保という観点からも、有効かつ重要な問題であると述べ

た。今回は、NTT による英語プレイスメントテスト作成の試みと評価（木村，2008）をもとに実施された大学における英語プレイスメントテストの結果を報告し、言語テストにおける段階評価の実際を示すとともに、大学入試英語問題のデータ処理に段階評価を導入した場合に、従来の連続尺度上での評価と比較して、どのような利点が生まれるかも論じた。

竹村雅史（北星学園大学短期大学部）

日本語能力試験（聴解）の国内外の正答率の差の大きい問題から見えるもの—日本滞在経験の有無を中心に—

楊 元（筑波大学大学院）

(a) 研究目的

試験問題を検討する際に、識別力の低い問題については多くの分析がなされているが、国内外受験者の正答率の差の大きい問題についての研究はまだ少ない。本研究では、国内外受験者の正答率の差の大きい問題に焦点を当て、中国人日本語学習者に解答を求めた後、プロトコル分析を行った。日本滞在経験の有無が正答率の差に影響を与える可能性を検討した。

(b) 研究方法

①『日本語能力試験分析評価に関する報告書』2001 年度－2005 年度の 1 級聴解問題中国語グループの正答率の差の大きい問題 20 項目を分析対象とした。

②日本国内（以下は CJ）及び日本国外（以下は CC）の中国語母語話者の学習者合計 25 名。両グループともに調査時の 4 ヶ月後に能力試験 1 級を受験するレベルである。

③発話思考法を用い、協力者の聴解問題の解答過程についてプロトコル分析を行った。

(c) 結果とその考察

プロトコル分析の結果から、日本滞在経験の有無が問題の解答に影響を与える可能性について検討した。正答率の差を大きくしたと思われる要因として、以

下のようなことが明らかになった。

① 日本の生活経験がなければ、学習する機会の少ない語彙が回答のキーワードになっているもの。例えば、「ステッキ」「半額」、「サービス」など。

② 日本の生活経験がなければ、その場面を想像しにくいもの。電車の時刻表のように、中国そしてほかの多くの国の人にとって、時刻表を見ながら予定を決める行動が普段の生活にあり得ないと考えられる。また、ある問題では、「もっと早いのに乗ればいいじゃない。ラッシュアワーの前の」を聞き逃すと、残りの情報は「身の動きもできない」「カバンがドアに挟まってしまい」であり、実際に経験していない国外受験者にとって、満員電車という状況を特定しにくいと思われる。

日本語能力評価を考える：質的視点より

森谷 浩士（熊本大学）

小林 美代子（熊本大学）

日本語学習者の日本語能力の測定に広く使われている試験に、日本語能力試験と日本留学試験がある。これら 2 つの試験は異なる目的をもって開発されていることから、測定する日本語能力も異なるのではないかと推測される。本研究は、この 2 つの試験が測定する日本語能力とはいかなるものかを、Bachman (1990) の言語能力及びテストメソッドの枠組みに従い、各設問項目を詳細に吟味し、それぞれの試験の特徴を特定しようとするものである。これら 2 つの試験は来年、大きく改訂されることになっており、日本語能力試験は日本語の習熟度の測定、日本留学試験は学業に必要な日本語能力の測定という、それぞれの目的に特化した試験へと変わっていくことが予想される。現在、予備研究として、現行版の試験を各設問の長さ、構成概念、質問のタイプ、談話的特徴などの観点から分析している。改訂版も同様の手法で詳細に吟味することで、両試験の変容の特徴づけが可能となると期待される。本

発表では、研究全体の概要と予備研究の中間結果を報告し、今後の研究課題への示唆を考察する。

Messick(1989)は、テストの妥当性の議論の中で結果的妥当性という概念を提唱し、テスト開発者は、妥当性を検証する際にテスト結果の解釈や社会への影響なども考慮に入れる必要があるとしている。今や、日本の外国人登録者数は200万人を超え、さらに、昨年、日本政府は留学生30万人計画を打ち出した。今後、外国人の日本語能力を判定する試験は、これまで以上に重要な役割を果たすものと思われる。日本語能力試験ならびに日本留学試験を利用する人のためにも、項目統計等の伝統的な量的分析に加え、本研究が提唱するような設問の質的分析を実施することで、両試験の説明責任を高め、結果的妥当性の確保が可能になることと思われる。

ワークショップ：

TDAP を使用するテストデータ分析・

中村洋一（清泉女学院短期大学）

中村先生がご用意くださった TDAP の CD-ROM とテストデータを使用し、古典的テスト理論に基づいたテストデータ分析を学ぶ実践的なワークショップとなった。まず、平均点、中央値、標準偏差、歪度、尖度、標準得点など、古典的テスト理論におけるテスト分析についてわかりやすく説明された。そして、TDAP 用にデータを整形する際のコツをご伝授いただき、実際に TDAP を参加者で走らせることで、その速さ、手軽さを実感した。標準得点についてのご説明では、わかりやすい例を用い、テストの得点分布によって素点1点の重みが異なってしまうことを示し、偏差値のような標準得点を使うことの意義を、説得力を持って述べられた。ワークショップ後半は、項目分析の話に進み、項目困難度、項目弁別力指数、実質選択肢数など、再び中村先生のご指導のもと参加者自らが TDAP を用い、算出したデータの解釈

を試みた。データの解釈には現場感覚が不可欠であり、テストの目的によってその解釈も変わる。受験者の特性、項目の特性など様々な観点から項目を見直し、検証するという項目の「質」の分析が必要であると述べられた。

今回のワークショップは古典的テスト理論によるものであったが、その限界にも触れられ、それを補うための項目応答理論による分析、そして、良いテスト項目をリサイクルできるようなアイテムバンキングの構築も推奨された。要所所で参加者からの質問があり、中村先生のご経験談も交えながらのパワフルなトークで、2時間のワークショップは瞬く間に過ぎてしまった。

金子恵美子（会津大学）

講演：

日本の英語学力評価の歴史

田中正道（広島大学名誉教授）

我々が専門として研究している言語評価に関する国内の歴史に関する講演であった。

まず、「そもそも何をすれば英語学力評価史研究なのか？」といった問いに対する回答から始まり、次のような研究分野や研究方法が示された。(1)「英語学力」概念の変遷を調べる、(2)評価方法（ツール）の発達・充実のプロセスをたどる、(3)評価の研究・実践の進展に貢献した人物を取り上げる、(4)各種の資格試験を調べる、(5)言語環境を考慮した評価国際比較、(6)生徒（受験生）と入学試験を調べる、(7)入試（選抜）制度の変遷を調べる、(8)大学における評価の授業史、学会（研究例会）史を調べる。

次に、日本の英語学力評価研究の歩みについて概説が行われた。配布資料には、明治時代や昭和初期の英語のテストの写しが添付されており、明治時代や昭和初期の英語のテストを見たのは、恥ずかしい話だが記録者にとっては初めて



であった。

本学会においても、質的研究、記述的研究、調査型研究、理論的研究、量的研究、計量心理学的研究など言語評価に関する様々な研究がなされているが、これら「共時的」（こういう術語を用いるかどうかはわからないが）な研究にばかり目が奪われていた記録者にとっては、言語評価に関する「通時的」研究は、とても勉強になった。まさに、温故知新であった。

片桐一彦（専修大学）

シンポジウム：

言語能力と評価に関する研究：現在・過去・未来

コーディネーター兼パネリスト：

小林美代子（熊本大学）

パネリスト：

佐々木みゆき（名古屋学院大学）

伊東祐郎（東京外国語大学）

「言語能力と評価に関する研究：現在・過去・未来」をテーマとして行われたシンポジウムは大会テーマ：History of Language Testing in Japanに相応しい誠に充実した内容であった。小林美代子氏は「ヨーロッパの言語評価研究の動向：日本への示唆」と題された発表において、特に英国ケンブリッジ大学において開催された「ヨーロッパ言語評価者会議の国際大会（ALTE 2008）」の報告を行った。現在のヨーロッパでは言語評価が社会や教育に与える影響、評価の使用目的や結果の解釈の持つ意味、評価の倫理や

説明責任といった事柄が大きな関心事となっており、大会発表タイトルの中にはimpact、washback、effect などといった語が多くみられるとのことである。研究発表、基調講演に触れながらヨーロッパの言語評価研究の動向が紹介され、日本の言語評価研究及び実践への示唆がなされた。ALTEの風景なども映像として紹介され国際大会の臨場感すら味わうことができるなど大変有意義な発表であった。

佐々木みゆき氏は「第2言語ライティング力研究」の過去、現在、未来」において、過去20年間に亘るL2 Writing Ability を研究対象としてWhat, How, Whyの観点から概観しながら、あわせてWhat→How→Whyと変遷してきたご自分の研究の歴史を具体例とし、能力とは何か、能力を測るとはどういうことなのかなど重要な問題について考察を深めた結果を紹介された。特に作文力の低い学習者に、事前に全体の枠組みを計画するなど作文力が高い学習者の方略を指導すると作文の質があがるのか、もしあがるならそれはその学習者の作文力と言えるのかなどの興味深い問題も提起された。さらにWhyの研究について、同じ作文の課題に取り組んでいても、学習者がただ課題をこなしていると思っっているか、より良い作品を生み出したいと思っっているのかで、作文の質は変わってくるのではないかという非常に興味深い指摘もなされた。その質の違いは実



力の違いと言えるのかなどという考えも述べられ、発表内容は哲学的に深くまた教育実践に有益であった。

伊東祐郎氏は「日本語教育における言語能力記述文（Can-Do Statements）とテストング」において、東京外国語大学における日本の大学学部進学を目指す留学生を対象とした1年間の日本語コースで取り組んでいる言語能力記述文（Can-Do Statements）の構築の試みを具体的な例とともに報告された。主な目的は大学進学に必要な日本語運用力の最終目標を明確にし、習得すべき運用力とそれを下支えする言語スキルを捉え直すことにあり、カリキュラムと評価のあり方に整合性をもたせることもねらいとしている。日本語コースにおける言語能力記述文の構築の経緯が紹介され、今後の言語能力の評価と日本語テスト作成のあり方について考察された。テストが教師の一方的な情報入手手段であった時代から、受験者や第3者に対しても様々な影響力を及ぼす存在となってきたなどという重要な指摘もあり、理論と実践のバランスのとれた非常に興味深い報告であった。

渡部良典（上智大学）

第30回研究例会 報告

7月4日（土）、静岡産業大学にて、第30回研究例会が開催された。最初の講演で、講師の今井隆夫先生（愛知みずほ大学）は、「用法基盤モデル（カテゴリー化の能力）から英語学習と評価法を考える---理論的枠組みと事例研究---（1）be + V-ing 形の場合」の演題の下、プロトタイプ事例を units として学習させ、そこから立ち上がるスキーマを提示した段階で、周辺事例の意味を問うテスト問題の作成指針を提唱された。講演の後、宮崎佳典（静岡大学）・長谷川由美（大真大学）・影山功（静岡大学大学院）「個

人向けにチューンナップした readability 式による日本語学習者用リーディングプログラムの開発」、法月 健（静岡産業大学）・伊藤彰浩（西南学院大学）・島谷浩（熊本大学）・木下正義（西南学院大学非常勤）「新旧 TOEIC®テストにおける項目及びテキストの比較分析」、鄭宗田（静岡産業大学学生）「日本と中国の大学入学試験における英語試験の比較」の3本の発表が続いた。参加者は出席した発表者も含めて10数名であったが、打ち解けた雰囲気の中にも熱気に満ちた半日間であった。

法月 健（静岡産業大学）



第31回研究例会 報告

日本言語テスト学会第31回研究例会は、平成21年9月26日（土）に東アジア英語教育研究会第91回研究会と共同で開催された（責任者：伊藤彰浩）。場所は福岡市の西南学院大学。参加者は31名。英語教育、ことばの学習や習得、テストの研究などに興味関心のある一般参加者、広島県、山口県、福岡県、熊本県、大分県、長崎県の大学教員、高等学校教員の参加者と、九州の大学の大学院生、学部生の参加も10名を数えた。JLTA会員の出席者としては、木下正義先生（JLTA 副会長）、島谷浩先生（熊本大学）、大津敦史先生（福岡大学）、今年

の全国大会での講演者である田中正道先生（広島大学名誉教授）であった。

研究会は、木下正義先生（JLTA 副会長）の小講演「韓国における英語教育：過去・現在・未来」で幕を開けた。8月にソウルで開催された KELTA（韓国言語テスト学会）に参加し収集した資料を基に、韓国における英語教育の変遷について触れながら、将来の日本の英語教育を考える際に必要となる視点について論じた。小講演の時間が若干足りず、十分な質疑応答がなされなく残念であったが、木下先生が準備した English Village のパスポートや資料などに全ての聴衆が強い関心を示した。「韓国は近い国ですが、結構知らないことが多いんですよ」との木下先生の言葉に、日韓の人的交流をさらに促進しながら JLTA の活動を活性化する必要性を感じた。

研究発表 I は、ペニントン雅子先生（西南学院大学）による「四技能を融合した英語指導法：ディベートを使って情報収集・分析・発信力を高める」であった。ペニントン先生は、英語ディベートは日本人英語学習者にとって二つの意味で大きな挑戦であると論じた。その理由として、日本の教育では重視されない論理的・説得的な議論展開力は必要であることから、ディベートの授業は、英語学力の高い学習者を対象に行われることが多いからだとした。しかし初級、中級の学習者であっても、言語活動と思考活動の二本柱を段階的に教授しながら、内容的に質の高いディベートを行うことは可能であると提案。言語活動では、四技能をバランス良く使う機会を与え、アウトプットの方法を身につけさせる。そして、情報収集の際に英語を「読む」、原稿作成で英語を「書く」、発表で英語を「話す」、相手の発表を「聞く」という言語活動を自然に融合した授業の可能性を模索した。思考活動では、情報の収集、分析、発信方法を身につけ、説得的に論理展開する力を養うとの観点から、実際に行った英語授業で収集したデ

ータや学生のコメントを踏まえて、今後の研究の方向性について論じた。この発表に対するコメントや質問は、具体的な授業の展開方法や資料収集法に関するものが多く、参加者から「大変参考になった」とのコメントが数多く寄せられた。

研究発表 II は、郷原光華さん（西南学院大学文学部外国語学科英語専攻 4 年）による「日本人中学生の英語学力と学習意欲：教育実習校で得たデータを分析し解釈する！」であった。この発表では、自らが教育実習先の中学校で収集した英語学力、英語に対する動機付け、英語に対する興味・関心に関するデータを統計的に分析した結果を報告した。卒業論文の中間発表であり、聴衆の暖かい雰囲気の中で、データ分析に関する質問や、今後、教職を目指す学生として言語テストを利用した得た知見をどのように活かして生きたいかなど、若い世代に対する厳しくも優しいコメントや意見が寄せられた。発表終了後、「緊張しましたが、このような舞台で発表できて光栄でしたし、先生方のコメントに感謝します。卒論の内容に生かせるように頑張ります」と郷原さんは感謝の言葉を述べた。

研究発表 III は、神崎沙織さん（西南学院大学文学部外国語学科英語専攻 4 年）による「英語の接尾辞 -able の意味拡張現象に関する実証的研究の試み」であった。この発表は David Lee (2005) が認知言語学の入門書で論じた仮説を検討しようとしたものである。接尾辞-able を有する英語単語が、「～できる」から「～しやすい」または「～するべき」に拡張する現象がどのような方向性を持ち、またどのような英単語が「～しやすい」または「～するべき」に意味拡張するかを検討するために、現代英語で使用されている 700 あまりの英語単語の特徴を分析した。聴衆からの質問としては、コーパスを利用した頻度係数を指標とした方がよりよいのではないかと、他の辞書を利用して意味を確認したほうがいいのではないかと、などであった。その一

方で、比較的理論的に検討すべき研究課題であるため、コーパスの使用よりもむしろ発表で使用されたクロス集計を用いたままで良いとの意見もあった。発表終了後、神崎さんは「いろいろな先生の意見が聞けてよかった。もっと統計分析法について勉強しなければ」と今後の研究への抱負を語った。

研究発表終了後、少し長めの休憩を取り、最後に伊藤彰浩（西南学院大学）と手島亜季子（西南学院大学文学部外国語学科英語専攻4年）によるワークショップ（「wh 疑問文って難しいの？みんなと一緒に考えようよ！」）が行われた。このワークショップでは第2言語習得研究で近年話題になっている主格疑問文と目的格疑問文の非対称性に関する問題を取り上げた。可能な限り難しい用語や説明を省き、聴衆の一人ひとりに意見を求めながら進行された。手島さんは、伊藤氏が数日前に収集し、分析したデータの報告をした。そして、この研究方法を利用し、日本語の習得研究を行う方法についても報告した。このワークショップで伊藤氏は「学生に、『勉強してね』という言葉は、自分自身に対する Yell かも知れない。今の時代、研究をやめる理由を説明するひとは多い。でも、あえて言わせてもらうならば、研究を続ける理由を模索すべきだ」と述べた。

本研究会は成功裏に終了した。研究会の目標として掲げた「風通しの良さ」「自由な意見交換」は達成できたと確認する。

伊藤彰浩（西南学院大学）



事務局よりお知らせ

- ◆ 転勤、転居等、JLTA の名簿記載事項に変更が必要な場合は、速やかに、事務局までご連絡下さい。また、銀行引き落としによる会費納入を利用している会員で、吸収・合併などにより、銀行名、支店名、口座番号等が変更になった場合は、必ず事務局まで、その旨をお知らせ下さい。
- ◆ JLTAの活動に対するご意見やご要望、Newsletter等への掲載希望記事などがありましたら、事務局までご連絡ください
- ◆ The JLTA office would be grateful if you could update us on your recent achievements relevant to the field of language testing and evaluation. Any information on your presentations, publications, awards, and so forth would be greatly appreciated. The relevance of the information will be evaluated by the office and given in the newsletter in due course.
- ◆ 会員の皆様の当該分野での近況をお伝えください。テスト・評価関係本を出版した、論文を発表した、賞を受けた、博士論文を提出した、など。随時報告していきたいと考えております。

編集後記

 ●研究が行われていない分野なんてあるのか、と思う。この分野はまだほとんど研究がなされていないところなどないのではないかと。何かしら研究がされており先行研究があるはずだ。もしなかったとしたら、探し方が悪いのではないかと。最近読んだ論文(Haladyna & Downing, 2006)で、構成概念と無関係の分散(construct irrelevant variance)に関する研究をテーマごとにわけて、あるテーマの研究が「たくさんある」、「不要である」、「必要である」、というようにわけてある表があった。たいへん感心したのだが、21項目あげられているこの分野の研究テーマのうち半分以上のテーマが「必要である」に入れられていた。うれしいことである。まだまだやられていないトピックはあるのだ。単に勉強不足ゆえ、研究されていない分野をしらないのである。(HS)

日本言語テスト学会事務局
〒389-0813 長野県埴科郡戸倉町芝原 758
TEL 026-275-1964 FAX 026-275-1970
e-mail: youichi@avis.ne.jp
URL: <http://www.avis.ne.jp/~youichi/JLTA>

