

**JLTA Newsletter No. 26**  
**日本語テスト学会**  
**The Japan Language Testing Association**

JLTA Newsletter No. 26      発行代表者: Randy Thrasher      2007年(平成19年)1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://www.avis.ne.jp/~youichi/JLTA.html>



\*\*\*\*\*

**第10回 日本語テスト学会 全国研究大会**

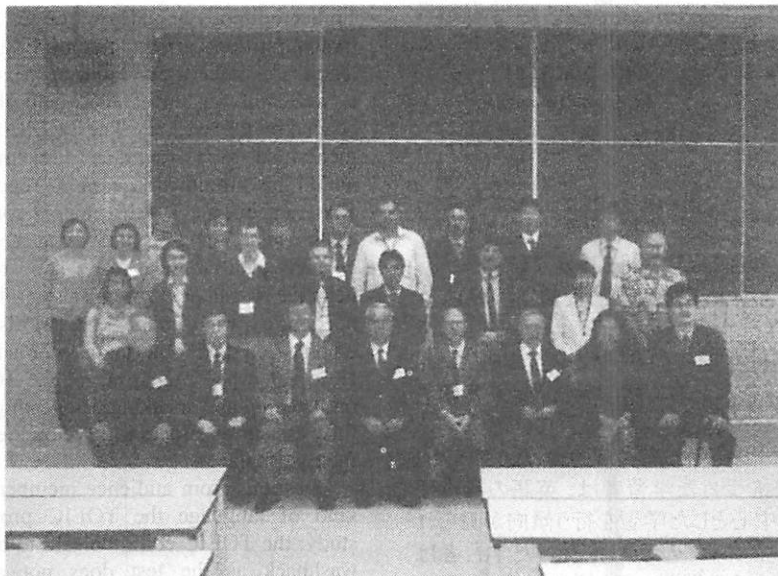
記念すべき第10回全国研究大会が、平成18年10月28日(土)29日(日)の2日間にわたって京都の龍谷大学深草キャンパスにて開催されました。10周年記念大会でもあった今回の大会では、海外から著名な研究者を3名もお迎えし講演・シンポジウムの他、初代会長の大友先生の講話と内容満載の大会となりました。

各々の発表については、以下に報告を記載致します。「報告」の執筆依頼をご快諾くださった先生方、どうもありがとうございました。

なお、今回の第10回全国研究大会の開催にあたりましては、龍谷大学様より、開催場所の無償提供の他、補助まで賜りました。学会を代表致しまして、学会事務局からも厚くお礼申し上げます。

大会実行委員長兼会場校の Soo im Lee 先生、たいへんお世話になりました。2年連続大会運営委員長を務められた法月健先生、どうもお疲れ様でした。

日本語テスト学会 事務局



第10回 日本言語テスト学会  
全国研究大会 研究発表

**A Study of Face Validity in a High School English Grammar Test: Inter-personal Authenticity and Affective Responses Made by Test-takers**

山下次郎 (東京都立国分寺高等学校)

妥当性の研究では、受験者の目に当該テストがどのように映っているかという Face Validity (表面妥当性) は、教育現場における受験者、つまり学習者への学習への動機付けにも影響を与える重要な問題である。本研究では、高校1年生(324名)を対象に様々なタイプの文法問題からなる定期試験を実施し、受験直後の質問紙調査を通じて、言語使用実感と難易度等の量的・質的分析を報告された。そして、コミュニケーション重視の授業展開を能力測定にも反映させることにより、学習者へのプラスの波及効果が期待できると結論付けられた。

報告者: 清水裕子 (立命館大学)

**清泉アカデミック Can-Do 尺度の開発とその評価**

長沼君主 (清泉女子大学)  
宮嶋万里子 (清泉女子大学)

Can-Do statements には、日常生活に必要な言語能力の尺度を中心としたものや、学習言語能力に関する活動を中心としたものがある。しかし、教育現場の様々な状況を考慮した Can-Do 尺度というものも必要であり、清泉女子大学では、4 技能・各5項目からなる独自のアカデミック Can-Do 尺度を開発されてきている。本研究では、大学1・2年生を対象に英語学習ニーズ調査として実施されたデータとプレイズメント・テストの結果をもとに、能力発達段階ごとの変化の分析が示された。本尺度では数値尺度ではなく質的な段階付けがされており、能力証明だけでなく学習・指導の指針としての機能も期待でき、今後の開発の展開に興味もたれる。

報告者: 清水裕子 (立命館大学)

**スピーキングテストの種類が持つ波及効果研究**

正頭英和 (関西大学大学院博士課程前期課程)

本研究は、ダイアログとモノログの2つの形式のスピーキングテストが中学生の英語学習に与える影響を調査したものである。調査の結果、これら2つのテスト形式の波及効果には顕著な違いがみられた。ダイアログタイプのテストを受けた学習者は、英語の音声面を意識し、音読を中心とした学習を行う傾向が見られた。一方、モノログタイプのテストを受けた学習者は、英訳や語彙を中心とする学習を行う傾向があった。

**パソコンによるスピーキング・テストと対面式スピーキング・テストにおけるパフォーマンスの比較**

周育佳 (東京外国語大学大学院博士課程)

本研究は、スピーキング能力を測る方法として近年急速に普及しつつある Computer-based speaking test (CBTS) と対面式のスピーキング・テストのスコア及びパフォーマンス指標の違いの解明を目的としている。今回は中間報告として、実験デザインが報告された。調査対象は高校生。CBTS として GTEC for STUDENTS のスピーキング部分を採用し、対面式テストでは、GTEC のタスクを修正し、難易度と内容面で同等のタスクを実施する。

報告者: 飯村英樹 (常磐大学)

**Are we talking about the same test? How do teachers and learners view the impact of the TOEIC on language learning?**

Mark Chapman (Hokkaido University)

The presentation addressed the links between testing, motivation, and language learning and focused on how the TOEIC impacts learners of English in Japanese corporations. The presenter suggested that the links between language testing and how it affected motivation for learning languages were under explored in the research literature. This is despite the existence of strong evidence that motivation is a powerful variable in the second language acquisition process. Literature from the field of general education also consistently claims that motivation for learning is harmed by high-stakes standardized testing. There is therefore the possibility that language tests may harm the developmental process of foreign language acquisition.

The presentation reported a series of interviews (n=28) with learners and teachers in Japan with regard to their views of how the TOEIC affected language learners and learning. The teachers and learners shared similar opinions about how the test provided extrinsic motivation for learners to study English. Several of the learners claimed they would not study English at all if it were not for the TOEIC and this view was also echoed by the teachers. The TOEIC was also frequently mentioned by learners as a factor in whether they would be promoted. Learners reported feeling some anxiety about the TOEIC, mainly concerning the length of the test. The presenter suggested that the TOEIC created positive washback for reading skills but negative washback for speaking. Learners described needing a variety of linguistic skills for their jobs, but emphasized the importance of reading and speaking. However, when preparing for the TOEIC they focused on listening and speaking, suggesting that the test is not fully assisting learners to develop the skills they need in their workplaces.

Questions from audience members focused on what kind of language the TOEIC prompted learners to study; the TOEIC could be said to be causing negative washback as the test does not actually contain a speaking section; and another question asked for more

detail as to what was actually causing the anxiety about the test.

Because more and more Japanese learners are eagerly interested in improving their TOEIC scores, this is valuable paper investigating how the TOEIC impacts learners of English in Japan. I would like to see the presenter's new work after the new generation TOEIC will be started.

報告者: 藤田智子 (東海大学・東京工業大学大学院)

### **Predicting Reading Comprehension Item Difficulty with Respect to Content Characteristics**

**Kiyomi Yoshizawa (Kansai University)**

**Soo im Lee (Ryukoku University)**

The purpose of this presentation was to examine the relationship between reading comprehension item difficulties and content (test task) characteristics. Three-hundred eighty one Japanese EFL learners took the reading part of the official TOEIC tests.

First, qualitative analyses were performed in order to investigate the characteristics of the test items found to be particularly difficult. Four characteristics were identified: 1) items that include domain specific vocabulary; 2) items that require test-takers to check the information spread across text or other types of prompts; 3) items whose stems are not specific enough for correct responses; 4) items that use paraphrased expressions in correct choices. Examples were presented for each of the four characteristics.

Second, quantitative analysis was conducted to find a small set of explanatory variables which would predict item difficulties. Correlation coefficients were calculated among the explanatory variables. The presenters reported that test/passage length and the familiarity of vocabulary in text and key sentences can be successful predictors of item difficulties.

After many findings were reported by the presenters, audience members actively responded. One of those responses was provided from Dr. Spolsky. He asked, "How shall we define test takers reading comprehension? How shall we measure reading comprehension?"

These active responses were evidence of how the audience members were interested in this study, understanding the complexity of reading comprehension ability and the difficulty of measuring it. The first half of the study focused on the qualitative approach, but I personally wanted to see the item difficulties and item information indices of the sample items. Overall, however, the presentation provided a great deal of important findings useful to item writers and test designers.

報告者: 藤田智子 (東海大学・東京工業大学大学院)

### **Meta-Validation of Institutional TOEIC Research in Japan**

**Steven J. Ross (Kwansei Gakuin University)**

**Hiromu Yamada (TOEIC Steering Committee)**

Critical analysis is necessary when conducting or reviewing validation research. The TOEIC test has in

recent years been not only a commonly used measure in such research in Japan, but has not been without its detractors. There is some ideologically based criticism of the TOEIC stemming from various issues of test-score interpretation and impact of its use in various institutions. However, Yamada and Ross point out that data-driven evidence used to support criticism of the TOEIC may be misleading if certain assumptions about the distribution and internal consistency of data samples are not tested beforehand. Among the assumptions behind sound use of correlation for validation studies is that both the predictor and criterion variables should have normal distributions, that sample variance and population variance should be shown to be equivalent enough for generalization, and that estimates of reliability for both sets of data be obtained.

Dr. Ross warned that correlation between the TOEIC and other measures in institutional studies might have been deflated because of the restricted range of non-random and relatively homogenous samples within an institution. In order to arrive at a more accurate picture, such correlation needs to be corrected for attenuation and for truncation based on standard deviation and reliability estimates.

A second method for obtaining validation evidence has involved the use of gain scores. For sound interpretation of the results, such studies need to consider possible pitfalls unless a way to monitor them is built in at very earliest stages of design. Aside from the need for use of parallel forms across administrations, the total hours, quality, and continuity of instruction need to be coded into the database so that pooling learners who have undergone different instructional time frames does not muddy the results. Case in point: a study reported low gains when there was a failure to account for a subset of students whose pre-and post-test administrations bracketed a summer recess. Such interactions should have been modeled in the analysis. Dr. Ross pointed also pointed out that regression to the mean should be modeled and that learning time should co-vary with gain, but that 50 hours of quality contact hours may be the minimum required for measurable gain.

I would argue that the TOEIC is essentially a domain-referenced test and as such is not as sensitive to instruction as, say, achievement or other criterion-referenced measures, but with the kind of care proposed above by the presenters, its usefulness for educational decision making was clear. As a third approach to validation of TOEIC use by institutions, Dr. Ross introduced a triangulation model for correlation among proficiency gains as measured by tests such as the TOEIC, use of syllabus-based Can-Do statements used for student self-assessment (pre- and post-instruction), and for post-instruction assessment of students by teachers. The degree of co-variance between post-instructional gains in proficiency, student Can-Do self-assessment, and the teachers' Can-Do assessments indicated higher degrees of consistency between reading & listening proficiency and student confidence than with the teacher's assessments. Furthermore, the use of Can-Do statements can open up areas for diagnostic applications.

Reported by Jeffrey Hubbell (Hosei University)

リッチコンテンツを用いた大規模一斉オンラインテストの開発と英語プレースメント・テスト「PLATON」の適用結果

秋山 實 (合資会社eラーニングサービス)  
井川悦朗 (NPO英語運用能力評価協会)  
白戸治久 (NPO英語運用能力評価協会)

CBTでネットワークを使用する場合の諸問題の解決策として、リッチコンテンツを用いた教材をPHPで作成し、コンテンツは flash movie 形式のCDで、解答行動はサーバとの応答で試験を実施する方法が提案された。これにより、1万人規模の同時受験が可能となり、解答の履歴もデータとして分析することができ、テストの信頼性も増す。この形式で作成された ACE オンライン版が現在モニター用として公開されている。

報告者: 小山由紀江 (名古屋工業大学)

タイトル: リスニングテストにおいて情報の検索および統合が項目困難度を与える影響

土方裕子 (筑波大学大学院博士課程、日本学術振興会)

土方氏はリスニングテストにおける解答に必要な情報 (NI) の位置と特性に注目し、それらが項目困難度や項目弁別度に及ぼす影響を調べた。大学生を対象に実施した英検3級の問題について分析したところ、NI の位置に関しては、困難度は NI が1節のみの場合に (特に初めの1節で顕著に) 高く、弁別度には差がないこと、NI の特性に関しては、困難度は発話速度や代名詞の割合、弁別度は質問文との距離が関連するという結果であった。フロアからは、推論を必要とする設問になると結果が変わるのではないかなど意見があった。

報告者: 島田勝正 (桃山学院大学)



技術研修生のための初級レベル日本語口頭能力試験のプリテストにおける評価表の妥当性に関する検討

庄司恵雄 (お茶の水女子大学)  
野口裕之 (名古屋大学)  
春原憲一郎 (海外技術者研修協会)

海外技術者研修協会が開発している日本語口頭能力試験のプリテストにおける評価表作成にかかわる調査研究の発表である。実際のテストの内容と実施手順が紹介されたあと、プリテストにおいて相対的に評価者間信頼性係数の低い項目についての分析と考察があった。特に評価表の妥当性の検証に焦点をあてた検討事項が報告され、課題の質に言及するなかで、分散分析モデルを適用して、評価表の再構成の必要性が述べられた。

報告者: 伊東祐郎 (東京外国語大学)

海外短期英語研修の口頭英語面における効果と測定

横川博一 (神戸大学)  
鈴木正紀 (Ordinate Corporation)

教育の目標があつて、それに沿つた実際の指導と学習があつて、その成果を評価する。これは当然のことであるが、まず、現実に起こりがちな乖離、すなわち海外短期英語研修の成果をいわゆるペーパーテストによって測定しがちであるという問題点を指摘した。そして、代替案として実際に留学中に培われたであろう能力を考察し、口頭で即応する能力を測定することをねらつて、「口頭英語能力自動試験 SET-10」を用いた結果を示した。口頭による英語運用能力というもの、捉えかたと測りかたについて示唆に富んだ発表であつた。

報告者: 前田啓朗 (広島大学)

テスト改訂における妥当性検証—新 TOEIC フィールドスタディー結果

三橋峰夫 (財団法人国際ビジネスコミュニケーション協会)

2006年からの TOEIC 公開テストの改訂に伴う、“More Authentic”というコンセプトや、新規に導入された開発手法である“Evidence-centered design、”そして項目分析の実態などが説明された。そして、新旧の形式における、様々な英語熟達度・特性を持つ受験者のスコアを比較したデータや受験者の印象評価のデータが示され、両形式において同程度の英語力を持つ受験者が同程度のスコアを取得すると判断できるものであつた。

報告者: 前田啓朗 (広島大学)

リスニング・リコールの採点について—アイデア・ユニットを分析単位にした場合

飯村英樹 (常盤大学)

本発表は、リスニング・リコールの採点において使用される分析単位であるアイデア・ユニットのうち、3種類のアイディア・ユニット— 1) Spencer version、2) Carell version、3) Sakai version—を取り上げ、それぞれのアイデア・ユニットの妥当性について検証している。発表者は、クラスター分析などの分析によれば、

上位群では Carrell version が、中・上位群では Sakai version が妥当性が高かったと報告した。

報告者: 塩川春彦(北海学園大学)

#### 並べ替え問題の部分採点法

安間一雄(玉川大学)

本発表は、並べ替え問題における合理的な部分採点法を求めて、様々な部分採点法を検証している。正解からの項目の異動回数分のペナルティを与える方法が合理的であるということが分かったが、計算が困難であるということが、まず指摘された。そこで発表者は、正解の連鎖と同じベクトルを有する項目連鎖を最大にする項目の組み合わせを見つける「項目連鎖最大化法」を考案した。これを用いれば目算で一意に計算できる、と発表者は結論づけた。

報告者: 塩川春彦(北海学園大学)

#### 学習者の語彙サイズはテストによってかわるのか—京都大学での事例研究結果からの考察—

笹尾洋介(京都大学大学院)

本研究は4種類の語彙テストが推定する語彙サイズを調べたものである。(1) 日本語訳を見て英単語を書く「発表再生」形式、(2) 英単語を見て日本語訳を書く「受容再生」形式、(3) 日本語訳を参照して4つの選択肢から該当する英単語を選ぶ「発表認知」形式、(4) 英単語の日本語訳として正しいものを4つの選択肢から選ぶ「受容認知」の各形式では、推定語彙サイズはそれぞれ3,020語、4,821語、6,158語、6,537語であり、かなりの差が見られた。

報告者: 土方裕子(筑波大学大学院博士課程、日本学術振興会)

#### ライティングにおける産出語彙の利用可能性: より良い評価へ貢献できるように

占部昌蔵(兵庫教育大学大学院)

自由英作文は「添削に時間がかかる」「評価が難しい」「自分に自信がないのでALTにまかせっきり」などの理由から、中高現場で避けられている。そこで本研究は、自由英作文を評価した結果を項目応答理論を用いて分析し、評価者の下した評価と産出語彙の客観的語彙指標の差異を調べたものである。フロアーからは、ETSの自動作成では何を基準に評価しているかを調べてみてはどうか、などの提案が出された。

報告者: 土方裕子(筑波大学大学院博士課程、日本学術振興会)

#### IRT, testing and online technology: the future of autonomous vocabulary learning?

Brent Culligan (Aoyama Women's Junior College)

Charles Browne (Aoyama Gakuin University)

Brent Culligan and Charles Browne delivered a well-received talk on the subject of vocabulary knowledge and testing. The talk addressed the development of a diagnostic test, which can predict the vocabulary items students already know and which ones they need to be taught. They suggested that currently available tests of lexical knowledge failed to accurately predict which items learners did not know. This was seen as important because if learners have not acquired the most high-frequency words of English, they will have problems reading fluently. If currently available tests only give a rough prediction of which words learners know there is the risk that these learners will progress through their language acquisition process without being exposed to some of the most vital lexis. The test the presenters have developed aims to overcome these problems.

The new online test was developed after a large-scale project in Japan, which applied Item Response Theory to lexical decision tasks. This allowed the authors to obtain difficulty estimates for over 5000 of the most frequently used words in the English language. This analysis enabled the authors to suggest that vocabulary frequency was not a good predictor of vocabulary difficulty. Students with large vocabulary size (3,000+ words) still had gaps in their knowledge of high-frequency words. In addition, the presenters claimed that vocabulary frequency does not allow the accurate calculation of a learner's vocabulary size. Finally, they said that their study enabled them to design individualized learning programs at the learner's level based on the most frequent words the students did not know. This clearly suggests the possibility of learners more efficiently acquiring the language they need to read and communicate in English through a well-developed diagnostic test.

The talk was well received and the presenters were asked several questions about their study and the new test. In addition to the details of their main study, the presenters commented on shortcomings of the teaching of English vocabulary within the Japanese education system. They suggested that although the first 1000 words of English are effectively covered after that Japanese learners are exposed to many low-frequency vocabulary items instead of going on to learn more important high-frequency lexis. This contributes to poor reading comprehension due to large gaps in the most important three to five thousand words of English.

報告者: Mark Chapman (Hokkaido University)

New ways of testing and teaching vocabulary online  
John Paul Loucky (Seinan Women's University)

John Paul Loucky gave a very enthusiastic talk on ways of testing and teaching vocabulary online. He attempted to compare a number of existing programs which aim to maximize the effectiveness of teaching and testing vocabulary. The presenter displayed and illustrated the online resources for the participants to demonstrate what they are capable of doing.

There were five main programs covered in the presentation. First the presenter introduced V-Check, which tests vocabulary size assessed by accuracy of recognition. Second was the DAVIE vocabulary Knowledge Scale. This mainly assesses accuracy of production including translation ability, L2 defining ability, collocational recall, semantic accuracy of meaning in context and syntactical accuracy of grammatical structure. The presenter then introduced semantic field keyword quizzes and the fourth program covered was Word Champ. This monitors online vocabulary learning using student-made flashcards, class- and teacher-made flashcards, computer generated tests of the flashcards and the use of CMS to monitor learning in detail. The final program covered in the presentation was Lexxi-Study's Online Vocabulary and Language Learning System. This program generates word lists and tests from any text along with combining common collocations, corpora and concordancers. It also tracks students and teachers writing and reading texts.

Participants were given a quick look at all the above programs and the presenter gave his views on the merits of each program. Due to the large number of programs covered and the enthusiasm of the presenter, there was unfortunately no chance for discussion or questions from the audience.

Reported by Mark Chapman (*Hokkaido University*)

#### 高校生英語リスニング力に関する Can-do statements の妥当性検証

吉池陽子 (*ARCLE 研究員*)

根岸雅史 (*東京外国語大学大学院*)

杵澤糸 (*株)ベネッセコーポレーション*)

発表では ARCLE による Can-Do 実地調査の1つの「英語でインターネットラジオを聞く」活動に対する英語能力調査の報告が行われた。自己申告型の Can-Do 調査が必ずしも現実の遂行能力を反映していないことから、実際にタスクを与えることにより、GTEC for STUDENTS のグレードとの対応が調べられたが、素材の難易度の解釈により結果が左右されることが分かった。

報告者:長沼君主(*清泉女子大学*)

#### 英検 Can-do 調査報告

柳瀬和明 (*(財)日本英語検定協会*)

松平知樹 (*(財)日本英語検定協会*)

発表では英研 Can-do 調査の能力記述文作成および調査方法の説明が行われた。調査は英検合格者を対象に3回に分けて実施された大規模な調査であり、英

語学習・使用状況調査も併せて行われた。結果、各級は依存型学習者(3級~5級)、自立型学習者(2級~準2級)、自立型使用者(1級~準1級)の3つの段階に分かれ、Can-Do による各技能の能力発達段階が示された。

報告者:長沼君主(*清泉女子大学*)

#### 日韓大学生の英語「方略能力」に関する一考察

達川奎三(*広島大学*)

本発表は、コミュニケーション能力の下位構成素である「方略能力」を測定する英語学習者向けの診断テスト(多肢選択式、30問)を用いて、日本人大学生と韓国人大学生における英語方略能力の特徴を分析した。その結果、日韓共通で以下の傾向が示された。(1)英語での「話題や談話を途中で回避したり、取り下げたりする」場面や、「対話者に対して理解の可否を自ら確かめる」ことに慣れていない。(2)「援助を求める」、「確認のための反応をする」、様々な「コミュニケーション上の障害を乗り越える」知識は備えているが、「自ら言い換える」ことによって、より積極的に、コミュニケーションの挫折を前もって回避する方略が身につけていない。

報告者:小泉利恵 (*常磐大学*)

#### 次世代英語教育におけるコミュニケーションとその評価ーコミュニケーションに対する理論的考察とそれを応用した評価モデルの提言ー

山中司 (*北陸大学非常勤、慶応義塾大学大学院博士課程*)

鈴木佑治 (*慶應義塾大学*)

本発表は、日本の英語教育において、あまり議論されていないプログラム評価の重要性と、体系的なプログラム評価に向けた「モデル構築」と「評価論の整備」の必要性を提言した。モデル構築に向けて考慮すべき要素として、(1) 教員だけでなく、学生・英語教育の専門家による評価を加える、(2) 学生の多様なニーズをくみ取り、教育プログラムをデザインし、成果の評価を行うことが提唱された。

報告者:小泉利恵 (*常磐大学*)

#### The Effectiveness of a 15-minute Writing Sample for Assessing Writing Ability

Melvin R. Andrade (*Sophia Junior College*)

Prof. Andrade reported on a work in progress addressing the utility of a 15 minute writing task designed as a placement test. His study involved six groups (N=136) stratified into high, mid, and low streams by TOEFL or TOEIC scores. The quick writing task was designed as a pen-pal letter with seeded topics such as a self-introduction, a vacation report, and a hometown description. No dictionaries were allowed. Ratings were of the tenses used by the writers. The presenter also used word counts and sentence length as



indicators of fluency. Prof. Andrade claims that the 15-minute quick writing task roughly corresponds to the groupings produced by the objective proficiency tests, and infers that quick writing can be used in lieu of longer tests if the goal is placement within an institution.

Steven Ross (*Kwansei Gakuin University*)

**Critical language testing: The General English Proficiency Test**

**Chih-Min Shih** (*Chin-Yi Institute of Technology, Taiwan*)

Prof. Chih-Min Shih presented research on qualitative interviews focused on learner attitudes to the General English Proficiency Test used in Taiwan. The goal was on examining the critical consciousness of among students and teachers on the use of the GEPT. Prof. Chih-Min Shih noted that some Taiwanese universities have adopted the GEPT as an exit test, thus increasing its perceived importance among Taiwanese students and teachers. As would be expected from interview data, there were many different and contrastive viewpoints about the validity and reliability of GEPT. An interesting trend was the observation that the expatriate teachers interviewed were in the main critical of the test, while the majority of the Taiwanese teachers saw the test as valid. Chih-Min Shih reported that the majority of the students interviewed saw the test as reliable and valid. Further, all of the students' parents and 60% of the students thought that the test ought to be used as graduation requirement. The results of the interviews provoked some interesting discussion among the audience. Many teachers (particularly Expatriate teachers) adopt the 'critical testing' perspective, seeing standardized tests like GEPT as hegemonic, while students and their parents, in Taiwan, at least, seem to be less critical of high stakes testing and seem to accept it as fact of life.

Steven Ross (*Kwansei Gakuin University*)

The Symposium began with an overview of what language assessment frameworks are and how they have been employed in Australia, Europe, and in the USA. The function of frameworks has in general been to provide criteria for making cross-language proficiency level comparisons. The issue confronting the symposium presenters was the fundamental need for an Asian framework for languages in general as opposed to a framework for EFL. It became apparent that the focus of the three presentations by participants was on the latter issue.



Prof. He Lianzhen provided a rich overview of EFL testing issue in China, with special focus on the College English Test (CET). The CET is a high volume, high-stakes test in China used not only for admissions, placement, and proficiency purposes, but increasingly as a graduation requirement. This latter function has increased the washback of the CET on Chinese society – in particular creating incentives for cramming and teaching to the test. It has also spawned a prestige race among universities to attract the most proficient students.



Prof. Oryang Kwon described the current state of high stakes testing in Korea. Prof. Kwon noted in particular the impact of the College Scholastic Ability Test used as the main college entrance examination criterion. Of particular interest was the description of the listening subtest on CSAT and the washback it has had on language teaching in Korea. A worrying trend in Korea is the *laissez-faire* approach to EFL testing at all levels of the educational system. Prof. Kwon explained how test providers large and small compete to market proficiency tests, and how these tests are given to candidates for whom the tests were not devised. Prof. Kwon's call for some kind of language testing policy in Korea resonated well with the audience, and brought the theme of the symposium into focus.

シンポジウム:  
Language Testing in an Asian Framework

コーディネーター

Steven Ross (*Kwansei Gakuin University, Japan*)

パネリスト

Lianzhen He (*Zhejiang University, China*)

Oryang Kwon (*Seoul National University, Korea*)

Yuji Nakamura (*Keio University, Japan*)





Prof. Nakamura Yuichi presented the last symposium paper. His theme was a critical exploration of the use of language standards based on North American and British standards of usage. Prof. Nakamura reasoned that if there a consensus in Asia to develop an Asian framework for pedagogy, there is an even greater need to form the basis for a language testing framework. While the need for an Asian framework for EFL testing is currently recognized, the actual content of such a framework is certainly more controversial. Prof. Nakamura pointed out that although there is by no means a consensus about the portability of the most widely known framework, the Common European Framework of Reference, there may be at least a set of common cultural touchstones linking China, Korea, and Japan.

#### IRT を利用したテストの改訂

藤田智子 (東海大学・東京工業大学大学院)  
前川眞一 (東京工業大学大学院)

2002 年に英語プログラムのプレースメントテストを IRT モデルを使い分析し、その結果をもとに新たな項目を作問し第一回目の改訂を実施したが、テストの精度という点で課題が残された。今後、成績判定のための学期末英語統一試験とプレースメントテストが共に英語能力を測定するテストであるとみなし、前者の過去問間で共通項目を係留項目として選定し、共通の受験者に受けてもらい、共通項目デザインによるテストの等化を行い、項目プールを構築し、プレースメントテストの改訂を試みる予定である。等化によるテスト改訂の方法、利点が説明された。発表後、学期末英語統一試験が成績に占める割合、プレースメントテストの基準点の設定方法など活発な質疑応答が行われた。

報告者: 吉澤清美 (関西大学)

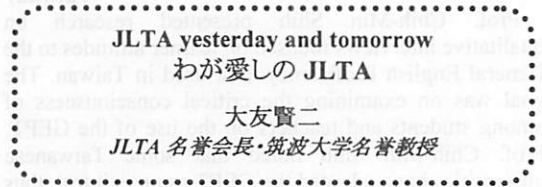
#### テスト自動管理・作成プログラム (Item Banking System) の開発

平井明代 (筑波大学)

Item Banking System は、テスト問題をデータベースに保存・管理し、そこから自動的に問題を組み合わせることでテストフォームを作成することを可能にしたテスト自動管理・作成プログラムです。テストフォーム自動作成の流れには、(1) データベースに保存したい問題項目

の準備、(2) 項目とその情報のデータベース化、(3) 項目データベースを活用したテストフォームの自動作成、(4) 使用項目の記録と自動更新、(5) 使用項目の統計機能が含まれる。筑波大学英語検定試験で実践されているテストフォーム作成の流れ、テスト実用性の向上に関する成果が報告された。発表後、保存したい問題項目の識別の方法、プログラムを導入する際の留意点など多くの質問が寄せられた。

報告者: 吉澤清美 (関西大学)



10 周年記念大会企画として、本学会のこの 10 年間の歴史について本学会名誉会長の 大友賢二 先生より講話を頂戴した。10 年前の本学会の誕生に至るまでの経緯から始まり、第 1 回研究例会と第 1 回全国研究大会について、そして本学会設立の最大のきっかけとなった 1999 年の The 21st Language Testing Research Colloquium の日本開催、ETS/JLTA Language Testing Workshop、言語教育国際シンポジウム、JLTA 言語テストワークショップ等について、それぞれ振り返られた。



最後に、これから未来にかけて本学会に対してのメールとして、特に次の 3 点についてお話された。(1) JLTA の設立は困難であった。しかし、その存続と発展は、それ以上に困難であろうこと。(2) JLTA という「樹」には、葉、幹、根がある。それぞれ連携し、協力することが大切であること。(3) JLTA は、言語テストの「研究」と言語教育の「現場」との架け橋になってほしいこと。

講話の後、引き続き、10 周年記念祝賀パーティが開催された。

報告者: 片桐一彦 (専修大学)

英語関係節の習得研究における可変的研究結果と文法性判断テスト



伊藤彰浩(愛知学院大学)

「名詞句の接近度階層」(主格>直接目的語>間接目的語>前置詞の目的語>属格>比較級の目的語)の妥当性を、文法性判断テストを用いて第二言語学習者から収集されたデータで検証すると、属格が主格に匹敵する高い正確度で解答されてしまう現象が報告されている。従来の第二言語習得研究は、この理由を測定対象の属格の特殊性について論じることで説明しているが、伊藤氏は、利用されてきた文法性判断テストの内容の問題点を指摘した上で、新たな文法性判断テストを構築し、それを日本人高校生 80 名に実施した結果から、「名詞句の接近度階層」の妥当性を再検証した。仮説を支持した結果が得られ、結果として、従来の文法性判断テストの質的問題が、属格の正確度上昇の要因であったことを立証した。今後の第二言語習得研究に、言語テストの専門的知識が貢献する余地が多いことを示唆した有益な研究報告であった。

報告者: 島谷浩(熊本大学)

ELPA 診断テスト(文法)の妥当性検証: 項目反応理論による分析を中心に

小泉理恵(常磐大学)  
酒井英樹(信州大学)

中学生・高校生を対象に、名詞句境界の認識と修飾関係の認識を調査した研究である。例えば These <1> small <2> books <3> English-Japanese <4> dictionaries の空欄に "are" を入れさせる問題では、正しい <3> を選択すれば名詞句 "these small books" および "English-Japanese dictionaries" およびそれらの中の修飾関係が正しく認識されているということになり、<2> や <4> に入れると修飾構造の概念がわかっていないという診断が下されることになる。しかしながら <1> に入れた場合(こういうケースは多いそうだ)、“These are small books” までは文法的である。これはワーキングメモリの制約による限定処理とは考えられないか? また、解答に至る判断を内省記述させたところ、70%以上が意図した結果と一致したそうであるが、この方法は本来このような基本的構文分析ができない学習者から情報を引き出したいにもかかわらず、能力レベルが低いと内省分析そのものができないという矛盾を抱えている。副題に「項目応答理論を中心に」とあるが、特にこの分析方法を取る必然性は感じられなかった。会場からは X パー理論との関係に関する質問および妥当性検証に関する質問があった。

報告者: 安間一雄(五川大学)

Creating pedagogic tasks based on analyses of English university entrance examinations

Keita Kikuchi (Waseda University)

Brian Wistner (Tokyo Junshin Women's College)

発表者からは、文部科学省の旗振りにもかかわらず高校現場ではタスクに基づく教授法がほとんど実践されていないこと、大学入試問題は 10 年前と比較しても英文和訳の多さとリスニングの少なさは相変わらずであることなどの指摘の後、大学入試のテキストを用いたタスクベースの活動の例が紹介があった。フロアからは日本の大学入試のテキストは学生のレベルを無視したほどむづかしいがタスク自体は下線部訳などローカルなものであること、個々の大学ではおそらく事後の項目分析がなされていないこと、逆にセンター試験では項目分析を繰り返しているにもかかわらずなぜか発音問題がなくなるということ、中学教師はタスクベースのアプローチを歓迎するのに高校教師はかたくなに文法訳読に固執していること、などの問題点が提起された。

報告者: 静哲人(関西大学)

### Asian language testing and the threat of scales

Plenary Speech by  
Dr. Bernard Spolsky  
(Bar Ilan University, Israel)

In our field of language testing study and practice, Dr. Spolsky is recognized by most of us as a master of measured words, but perhaps more accurately or characteristically, a master of thought measured by experience, wisdom, and wit. Employing the visual aid of Power Point technology, he introduced at the beginning of his speech with biological and ecological images from science; then of musical scales – which the ancient Greek culture studied as applied mathematics. To this reporter, there was a call to be prepared to use more than one side of our brains.

In his abstract for this conference, Dr. Spolsky presented participants with two deliberations:

1. Will the attempt to build a single scale for language proficiency for all countries lead to a single scale for language proficiency for all Asian countries?
2. Will the infection of a single scale based on current Western models force conformity rather than permit intelligent localized diversity?

The first question brings to mind issues of accuracy and fairness, but also utility. Dr. Spolsky presented a history of formal assessment beginning with its roots in Imperial Chinese examinations for civil service, through the Jesuits who brought the idea to the West, its growth in Europe in the 19th century, through developments in 20th century America, and beyond to the Council of Europe's common framework of

references for language learning, teaching and assessment.



Up through the 20th century, the utility of tests has involved, in one form or other, variations on the theme of control and the problems with validating the use of scales adopted to inform that control. Dr. Spolsky allusion to 'unavoidable uncertainty' characterizes the dilemma of unidimensional scaling of language proficiency, while so many of us perceive language as multidimensional. He offers as an alternative the use of profiles informed by multiple scales, which may include such devices as 'Can-Do' statements, but warns against the use of any single measure. Given the various uses and purposes for language assessment in Asia, judicious ad hoc weighing of several scales may be appropriate. By this, we may understand the gist of Dr. Spolsky's second 'question' as not a question at all, but a lead-in to his call for us to meet our social and ethical responsibility by careful attention not only to whether or not our means of assessment is accurate, but also attend to justifiable use of assessment as construct relevant.

Reported by Jeffrey Hubbell (Hosei University)

## 第24回研究例会 報告

### 学習者のためのテスト妥当性追求 そして Can-do の活用

第24回研究例会が、平成18年12月3日(日)に東海大学湘南校舎にて開催されました。「学習者のためのテスト妥当性追求そしてCan-doの活用」というテーマのもと、4件の発表が行なわれました。

各々の発表については、以下に報告を記載致します。当日直前の依頼であったにもかかわらず「報告」

の執筆依頼をご快諾くださった先生方、どうもありがとうございました。

なお、今回のJLTA第24回研究例会の開催にあたりましては、東海大学様より、開催場所の無償提供の他、補助まで賜りました。学会を代表致しまして、学会事務局からも厚くお礼申し上げます。

会場校の藤田智子先生、松本佳穂子先生、どうもお世話になりました。

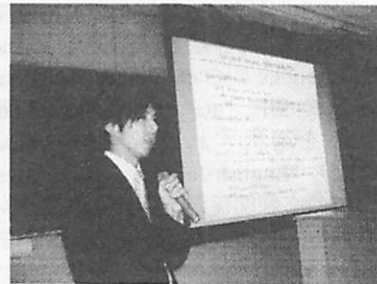
(日本語テスト学会 事務局)

内部評価としての Can-Do 尺度と英語フレームワークの開発～香住丘高校および清泉女子大学での開発研究をもとに～

長沼君主(清泉女子大学)

Can-Do 尺度開発の経緯として、自己能力評価のための CDS (Can-Do Statements) とテストスコア解釈のための CDS の実例を示して概説した後、実際の指導を考慮した、シラバスや目標に基づく CDS の必要性和開発過程が述べられた。香住丘高等学校、清泉女子大学、埼玉県英語運用力推進事業での CDS 開発事例を交えた、理論的背景、指導者のニーズ、学習者へのフィードバックや目標設定という多面的な示唆を得ることができる発表であった。

報告者: 前田啓朗(広島大学)

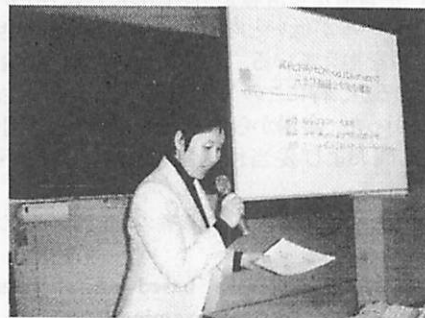


高校生向け Can-do statements の妥当性検証と今後の展望

吉池陽子 (ARCLE)

根岸雅史 (東京外国語大学)

沓澤糸 (ARCLE / (株)ベネッセコーポレーション)



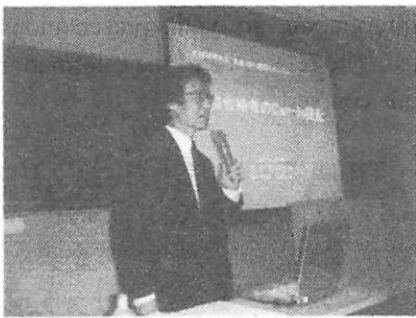
本発表は、GTEC for STUDENTS の Can-do statements (以降 Can-do statements) の妥当性を検

証した。その結果、経験に基づく自己申告の Can-do statements の中には、実際のパフォーマンスを比較的反映しているものと、そうでないものがあることが示された。改善策として、Can-do statements の場面や状況などを具体的に示す、実際のパフォーマンスに基づき Can-do statements を作成する等の方向性が提案された。

報告者:小泉利恵 (常磐大学)

### 英検合格者の Can-do 調査

柳瀬和明 (財)日本英語検定協会



2万人を超える英検合格者に対し、約3年の歳月をかけてアンケート調査を行い、『英検 Can-do リスト』が作成された。この一連の研究過程についての報告がなされた。

アンケートには、80~100項目にわたる4技能別の能力記述文があり、それぞれを5段階のリカート・スケール(自信の度合い)で自己評価してもらうように構成されている。これらのアンケートの回答結果を集計し分析することによって、英検の各級の合格者ごとにその級の合格者は英語運用にあたってどんなことができると合格者自身が思っているのかが記述された。

なお、『英検 Can-do リスト』は、(財)日本英語検定協会より頒布されている。

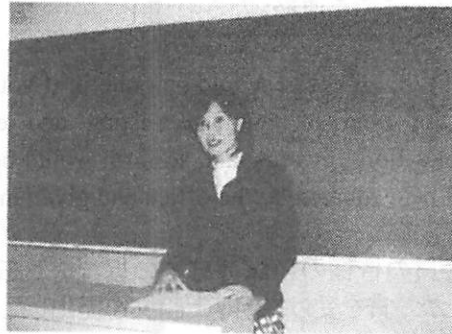
報告者:片桐一彦(専修大学)

### テスト問題の質の管理と実用的テスト作成をめざして: 教員が使用しやすい項目銀行システムの開発

平井明代(筑波大学)

平井氏は筑波大学の項目バンクシステムの開発方法、実用性、今後の可能性について発表された。過去のプレースメントテスト及び熟達度テストの項目を項目

応答理論による同一尺度によって連結し、新問題が加わる度に更新作業を繰り返して行くというシステムであるが、十分な蓄積が出来た今は汎用性のある「テスト自動管理、作成プログラム」として機能するものとなっている。今まで問題の管理、テストの作成に要していた重複した教員の労力が軽減されただけでなく、統計処理のされた過去の項目を個々の測定目的に合わせて編集することでより妥当性、信頼性の高いテストができるわけである。



発表では、問題の取り込み、整形、ファイリング、統計との連結、新テストの作成法が順を追って紹介され、コンピュータベース、紙ベースのテスト、そして授業単位の小テストにも適用できるという様々な可能性が示された。経費の制約などの理由で自前のプレースメントテスト、到達度判定テストを長年作っている学校には大変参考になる上に、外部テストを使うよりもその学校のカリキュラムと学生により適合したテストの作成にもつながる。このアクセススペースの市販向け項目バンクプログラムをモニターとして試行できる制度の紹介もあった。

報告者:松本佳穂子(東海大学)



## 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

●国際学会で口頭発表をするアジアの国々のテスト研究者が最近富に多くなってきたという印象を受けます。今回の JLTA の大会はまことに時宜を得たものでした。●「ギョエテとは俺のことかとゲーテいい」という川柳があります。Goethe をギョエテと発音していることが私たちにはよくあるように思います。●言語学を勉強し始めたころ伝統文法で有名な Kruisinga の発音がクロイズインハであることを知るまでにしばらくかかりました。言語教育関係でも時々正しい発音の仕方がわからない名前を見かけます。Kumaravadivelu, Wajnryb, Pakir, Morgenthaler などなど、難しい氏名には必要に応じて発音表記を付記する必要があるかもしれません。Flow 理論で有名な Csikszentmihalyi は最近カタカナでも見るのでだいぶ知られてきたように思いますが、英語の原著でも裏表紙に発音記号があるのでチクセントミハイとわかります。アジア系の名前、特に日本人の名前などはローマ字書きでは発音が大変難しいものも多いでしょう。●本来の発音から少し逸脱しているくらいは許容する必要があるでしょうが、やはり原語の発音に忠実に発音をしようとするのはよい習慣かもしれません。本大会のご講演(実に明快で示唆に富むものでした)の中で大友賢二先生が Popham の発音について言及なさるのを伺っているながら抱いた偶感です。(YJ)

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>.

