
JLTA Newsletter No. 27
日本語テスト学会
The Japan Language Testing Association

JLTA Newsletter No. 27 発行代表者: Randy Thrasher 2008年(平成20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>



北海道から沖縄まで

中村洋一 (常磐大学)

事務局に保管されている入会申込書のファイルを見ると、発起人会があった1996年12月14日の日付で入会の手続きを処理した会員数は20名でした。現在の会員数は約190名ですから、10年ほど、約10倍になった計算です。1998年に初めて得た賛助会員も現在は10社になり、貴重なご助力をいただいております。

私事で恐縮ですが、JLTAの発足とほぼ同時に生まれた、わが家の三女が、今年の春、中学校に入学する年齢になりました。12年は長かったような短かったような、しかし、確実に時は流れています。JLTAの事務局も、人間で言えば、中学生ぐらいになったかなあ、という印象です。振り返ってみると、事務局はいつも自転車操業で、会員の皆様にご迷惑ばかりかけてきてしまったと、反省しています。娘は4月の入学に向けて、初めて体験する英語の学習にも興味を持ち始めているようです。手の掛かる子供時代から、少しは成長して、来るべき大人の時代に向けて基礎固めを始める中学1年生のように、心を新たにしていきたいと思います。

2008年度のJLTAは、Randy Thrasher 会長から、浪田克之介新会長へと引き継がれ、12年目の活動に取り組むことになりました。Randy Thrasher 会長には、大変お世話になりました。特に、節目となった、創立10年目には、韓国や中国との密接な連携の扉を開いていただきました。ありがとうございました。

大友賢二初代会長の東京から、Randy Thrasher 現会長の沖縄へ、そして、浪田克之介新会長の北海道へと、まさしく、「北は北海道から、南は沖縄まで」となりました。そして JLTA の会員も、全国各地からご入会いただいているところです。

12年の歴史の上に立ちつつ、浪田新会長のもと、まだまだ発展途上にある、JLTA をさらに大きく、意義あるものへと育てていくために、「北は北海道から、南は沖縄まで」の会員が、より活発に、より広く、より深く、より高く、言語テスト研究の活動に取り組んで行かれることを、願っています。

頼りない事務局ですが、会員の皆様のご協力をいただきながら、活動を進めていきたいと思っております。よろしく願いいたします。

第11回 日本言語テスト学会
全国研究大会 研究発表

TOEIC®受験者の英語能力記述—TOEIC® Scale Anchoring Study
三橋峰夫 ((財)国際ビジネスコミュニケーション協会)

2006年にリニューアルされたTOEIC(新TOEIC)は、現実の言語使用をより良く反映した出題と、Evidence-Centered Design(ECD)やScale Anchoring Study(SAS)の手法による、スコア・ユーザーへのより有用な診断的情報の提供を、改善の主眼としている。ECDでは、スコア・ユーザーが欲する情報を裏付ける証拠の収集に配慮してテストが設計される。新TOEICの公式認定証は、測定領域ごとの正答率(Abilities Measured)やレベル別評価(Score Descriptors)を提示するが、特にレベル別評価内の長所と短所の記述は、ECDの理念に基づき、統計学者と言語学者が共同で行ったSASの成果と言える。具体的にはまず、統計学者が受験者を4段階の得点域に位置づけ、各レベルの受験者集団が各項目に正答する割合を算出、各レベルの受験者を明確に弁別する項目(anchor items)を特定し、それらの項目をクラスター化する。次に英語学者がクラスター内の項目を分析し、各レベルの弁別特性に関する詳細な注釈を作成して、レベル別の一覧表にまとめる。SASにより、より効果的にユーザーへ情報をフィードバックすることが可能になった。発表では実際のTOEICの問題を例に、レベル別評価の記述がいかにか実践的な解釈に適用できるかについても説明がなされた。

報告者:法月健(静岡産業大学)

Can-do statements 調査と指標テスト:ESPの観点から

小山由紀江(名古屋工業大学)

一般的な科学技術英語の習得を目標とする英語教育カリキュラムを背景に、名古屋工業大学で開発されたcan-do statements(CDS)の実践とその分析結果について報告が行われた。6月に行われたCDSの本調査の結果が、4月にクラス編成の目的で実施されたTOEIC、8月に実施された統一試験(期末試験)の結果と比較された。パイロット調査時のCDSは75問で構成されていたが、質問項目が多く、質問によっては現実の状況との隔離も示唆されたため、本調査では各スキル5問ずつ計20項目に精選して、実施された。分析の結果、TOEICの得点域を3段階で区切ってCDSの評定平均値と比較した場合、TOEICの得点域が高くなるにつれてCDS評定値も高くなる傾向が確認されたが、中位の学生が低めに評価し、readingに比べて他のスキルの評価がやや低かった。CDSはTOEICとの相関が0.22と、統一テストとTOEICの0.53の相関

に比べてかなり低く、統一試験との相関もTOEICと統一試験の相関よりも低い値を示した。総じて、自分の経験に結びつく活動内容に関しては評価が高いが、経験に結びつかないものへの評価は低くなる結果となった。CDSは学生に主体的に自らの言語活動について考え、評価する機会を与えたが、今後、他の評価指標やカリキュラムとの整合性を強化して、より実践的なCDSを構築していく必要性が述べられた。

報告者:法月健(静岡産業大学)

文法性判断テストによる明示的 / 暗示的知識
島田勝正(桃山学院大学)

島田氏は、文法性判断テスト(Grammaticality Judgment Test; GJT)において、問題文提示時間の制限の有無により、明示的知識と暗示的知識を測定することができるかを調査した。開発されたテストは、80個の問題項目から構成され、20個の文法範疇を取り扱っている。1つの文法範疇ごとに2項目の適格文(grammatical sentence; GR)と2項目の非適格文(ungrammatical sentence; UG)が配置され、さらに同一の80項目に対してUGの誤りを指摘させるテスト(Error Location Test; ELT)が付加されている。GJTは、コンピュータで出題される場合、問題提示時間を制御できる。大学生81名を被験者とした調査では、提示時間制限のあるテスト(Timed GJT)、時間制限のないテスト(Untimed GJT)、そしてELT間において平均点の差は小さく、相関も高かった。この結果は、時間制限の有無が明示的知識と暗示的知識の識別に必ずしも貢献するものではないことを意味している。一方、GRとUGの相関は、Timed GJT、Untimed GJT、ELTのいずれのテストにおいても低い結果が報告された。そして、GRとUGを正しく判断する能力が異なっていると考えられると報告された。従来の考え方では、明示的知識と暗示的知識は、その作用するスピードの違いによって主に識別されていたが、本研究結果は、作用する対象が異なっていることを示唆しており、今後の研究の進展が興味深い。

報告者:島谷浩(熊本大学)

多肢選択形式による語彙サイズ測定テストの誤差の推定方法について

野上康子((株)教育測定研究所)

多肢選択式の語彙サイズテストが、最近注目を集めている。学習者にとって自分の語彙サイズは、英語を学習する上で非常に貴重な指標となるのは間違いないであろう。野上氏が提案した語彙サイズテストの標準誤差を推定する手法の概略は次の通りである。1) U 個の語彙のうち本当に知っている語彙の数(語彙サイズ)が w であるとして、出題された n 項目に含まれる「知っている語」の数の分布を算出する。2) これを重みとして、受験者が本当に正解を知っている項目の数 R を与えた場合に観察される正答数 x の分布を合成する。3) U 個の語彙のうち本当に知っている語彙の数

(語彙サイズ) w に関して、選択肢数 k 、項目数 n の場合の正答数 x の標本分布を導き出す。詳細は割愛するが、野上氏がコンピュータにひたすら計算させて得られた語彙サイズの標準誤差および信頼区間が示された。驚いたことに、測定の精度が増すのは問題項目数が500を超えるぐらいからだという。語彙サイズテストの問題項目数が200以下の場合、結果は相当の誤差を含んでいると認識しておく必要があることがわかった。語彙サイズを推定する際の基本的情報として、更なる研究の深化を期待したい。

報告者: 島谷浩 (熊本大学)

スピーキングにおける生徒相互評価の妥当性: 日本人高校生の場合

深澤真 (茨城県立竹園高等学校)

スピーキングの評価方法の一つである生徒相互評価 (peer assessment) の妥当性について、高校での詳細な研究結果が報告された。参加者は日本人高校生 62 名と英語教師 4 名で、本番の評価の前にビデオを用いて練習を行った。クラスごとに生徒が 1 人ずつスピーチを行い、残りの生徒と教員がそのスピーチの評価を行った。妥当性は、併存的妥当性、表面妥当性、信頼性の観点から吟味された。主な結果は以下 3 点であった。第 1 に、生徒相互評価は、全体評価において、教員による評価と同等の十分な妥当性が認められた。第 2 に、生徒の熟達度の上・下位群には十分な妥当性が見られ、中位群には妥当性が部分的に認められた。第 3 に、生徒相互評価を行う人数については、評価人数が増えるに従い (5、10、20 人)、妥当性が高まる傾向が見られた。さらに、生徒相互評価をスピーキング評価で取り入れることにより、より実用的で、生徒にとってより安心感があり、よりダイナミックなスピーキング活動・評価が可能になることが示唆された。

質疑応答では、相互評価で用いた評価尺度の形式 (各レベルに記述のある 6 段階方式)、相互評価の分布の形 (正規性が見られた) 等が話題になり、相互評価を行うことにより、生徒自身が自己を客観的に振り返ることができ教育的にも有意義である点や、相互評価の研究は世界的に見ても少ないため、今後より研究を深め、英語で研究報告していくとよい点などが指摘された。

報告者: 小泉利恵 (常磐大学)

速読と精読におけるテキストタイプとタスクタイプの要因の分析

長沼君主 (清泉女子大学)・和田朋子 (工学院大学)

発表では、速読と精読に影響を及ぼす要因として、テキストタイプとタスクタイプに焦点をあてた研究報告がなされた。テキストタイプは、テキスト内の情報の関係の強さ (マクロ構造における情報の緊密さ) から Association (AS), Description (DS), Causation (CS), Problem-solution (PS) の 4 種類 (前から後に行くにつれて、loosely-structured から tightly-structured にな

る; Kobayashi, 1995) に分類した。タスク (項目) タイプは概要型、部分型、比較型、参照型の 4 つを用いた。参加者は大学生 70 名で、速読として、異なるテキストタイプの 4 種類のテキストをすばやく読み、そのスピードが測定された。テキストを見ずに、理解度を測る 4 つのタスクタイプの問題を解いた後、精読として、同じ 4 種類のテキストを時間制限なしに読み、テキストを見ながら 4 つのタスクタイプの問題を再度解いた。

その結果、主に以下 3 点が報告された。第 1 に、速読時には、構造の緩やかなテキスト (AS, DS) で読むスピードが早かった。理解度については DS や CS で高かった。そのため、速読では、読むスピードと理解度の 2 つの軸によりテキストタイプが分類できる点を示された。第 2 に、タスクタイプとの関連では、速読と精読時ともに、緊密度の高いテキスト (CS, PS) では参照型タスクで、低いテキスト (AS, DS) では概要型タスクにおいて理解度が高かった。第 3 に、速読では、概要型、参照型、比較型、部分型の順で理解度が低下したのに対し、精読ではタスクタイプによる差が見られなかった。今後は、指導や評価面でのテキスト・タスクのよりよい組み合わせや、速読と精読の能力バランスのあり方を探求し、速読・精読に関する Can-do 項目の洗練と評価タスクの開発などを行うとのことだった。質疑応答では、トピックの影響の相殺をどうするか等が議論された。

報告者: 小泉利恵 (常磐大学)

日本語手紙文を評価する際に留意すべき点とは - 評価過程の分析をもとに -

宮島良子 (名古屋大学大学院博士後期課程)

日本語作文評価の中で手紙文を評価する場合、実際にどのように評価しているのかを明らかにする研究の発表であった。

評定者は、日本語教師歴 2~8 年の日本人 (JNT) 7 名と、日本語能力試験 1 級を有し日本語教師を目指し日本の大学院で日本語・日本文化専攻の中国人 (JSL) 7 名 (教師及び作文評価の経験なし) であった。評定した素材文は、中国の大学生 (日本語学習期間: 1 年 7 ヶ月) が書いた手紙で、相手 (読み手) は年上の文通相手を想定し、手紙の内容は旅行へのアドバイスとした。発話思考法及び半構造化インタビューを用いて、発話プロトコル資料を収集し、作文評価の過程を分析・考察した。

JNT も JSL も、高い信頼性係数が確認され、違いはなかった。しかし、評定者間において、課題を達成しているとはどういうことであるのかという認識のずれ、呼称の選択、表現の選択といった読み手への待遇表現に関する考え方のずれ、書き手の意図をどのくらい汲めるかの違いなどがあり、評定結果に影響していることが示唆された。

報告: 片桐一彦 (専修大学)

能力記述による日本語教育カリキュラムの改善 —
Can-do-statements の実施を通して—

村上京子 (名古屋大学)

名古屋大学の日本語教育において、7レベルに分かれた会話・独話・聴解・読解・作文の各クラスに関して、CEFを参考に能力記述によるレベル設定が行なわれた。その能力記述作業を行なった一連の過程と、記述されたものの妥当性の検証に関する発表がなされた。

まず、「行動目標」をレベル別4技能別に作成した。その際、(1)外部の人や学生にもできるだけわかりやすい、簡潔なものであること、(2)行動目標の形(can-do)で書かれること、(3)何を教えているかではなく、何が身につくかという proficiency が記述されること、(4)単なるレベルの記述にとどまらず、教材や活動の具体的な内容との関連性を常に見直し、目標自体も進化していく可能性を持つものであること、という方針を採用した。その後、各レベルごとの行動目標に基づき、「can-do-statements」を作成した。やりとり、話す、聞く、読む、書く、それぞれにつき各20問作成した。

67クラスで合計756枚のcan-do-statementsシートが回収された。can-do-statementsの信頼性係数、各テスト項目の困難度係数と識別力係数、修了試験との相関係数がそれぞれ算出・分析・考察され、全体的にはよく機能していることといくつかの今後の課題が示された。

報告:片桐一彦 (専修大学)

共通教育「英語」の成績評価における課題とその解決に向けた試み—愛媛大学における事例

折本 素 (愛媛大学) 廣森友人 (愛媛大学)
田中英理 (愛媛大学) 山西博之 (愛媛大学)
山本武志 (ベネッセコーポレーション)

本発表は、愛媛大学・英語教育センター(平成13年度設置)における、少人数クラス編成による学生活動中心型の英語授業の実施、学生の実態を考慮した独自の英語教科書の作成などによる、英語による実践的なコミュニケーション能力の育成を目指した教育実践に基づいて、評価の観点からの研究と分析であった。まず、愛媛大学・英語教育センターの取組は、英語教育・学習に対して一定の成果を上げている一方で、各教員による評価の(極端な)ばらつきや成績評価の妥当性の確保といった新たな課題を生みつつあるという問題提起がなされた。次いで、このような課題の解決のために、平成19年度からは、各授業の成績評価の30%分を英語能力試験(GTEC for Students)に充てる、学生の実態に応じた独自の英語運用能力判断基準(Can-Doリスト)を作成するといった新たな試みに着手したことが報告された。さらに、前者の取組により、教員間の成績評価のばらつき相殺や、毎学期の成績評価(GTEC for Studentsによる30%を除いたもの)とGTECの結果の比較などが可能になり、後者の取組により、愛媛大学の学生の実情に応じたより適切な指

導と評価が可能になったと報告された。まとめとして、英語能力試験を成績評価の一部に取り入れることの意義やその効果などについて肯定的に議論された。

報告:塩川春彦 (北海学園大学)

言語コミュニケーション力の三次元的理解

柳瀬陽介 (広島大学)

本発表は、言語コミュニケーション力の全体像を理解しようとする哲学的探究の試みとして提起された。このような探求は、言語コミュニケーション力の一部分を精確に測定しようとする科学研究と共に重要である、と発表者は主張した。本発表では、言語コミュニケーション力を、(1)言語知識、(2)心の理論・関連性理論、(3)身体性の三次元のベクトルの合力として、全体像を捉える見方が提示された。この全体像は、これまでの言語コミュニケーション力理論の発展史を踏まえながら、“capacity”、“strategic competence”といった用語で現されていた言語知識を使いこなす力を、認知科学の「心の理論」と「関連性理論」に沿って解釈する試みであり、また1990年のBachmanのモデルでは存在していた“psychophysiological mechanisms”の重要性を再主張するものでもある、と主張された。まとめとして、本発表では、この三次元的理解により、言語をコミュニケーションで使う際の相手との相互作用性についての理論的理解が進み、また実践的にも、各種英語教育で狙われている言語コミュニケーション力を統合的に理解することが期待できる、と今後の研究の方向性と可能性が示された。

報告:塩川春彦 (北海学園大学)

A meta-analysis of task type effects on listening and reading test performance

Yo In'nami
(Kanda University of International Studies)

Prof. In'nami talked about “a meta-analysis of task type effects on listening and reading test performance.” In his study, by contrasting a narrative approach to synthesizing findings, Prof. In'nami conducted a meta-analysis on the effects of three types of tasks on listening and reading test performance. A total of twenty-nine studies retrieved through a thorough search of the literature was the basis for estimates of mean effect sizes for listening and reading task type effects. Results using the fixed and random effects models of meta-analysis show that task type effects vary and are not sufficiently similar across studies except for in L2 listening. The difficulty order of multiple choice (MC) being the easiest, followed by open ended (OE) and then by summary gap filling (SG) tasks is observed across skills, languages, and research designs (independent-group, and repeated-measure with or without counterbalancing). The degree of task type effect difference is medium to large for MC vs. OE tasks and small to medium for OE vs. SG tasks, although the degree of MC vs. SG task type difference is wide-ranging and not very clear. Prof. In'nami

discussed implications and directions for future research as well.

Reported by Yuji Nakamura (*Keio University*)

A Dual-Testlet Approach to Student Placement Toshihiko Shiotsu (*Kurume University*)

In his presentation, Prof. Shiotsu discussed an in-house development and administration of a placement test which was to place incoming students into three level divisions of the newly introduced Core English course. The test was designed to have a close interface between the test items and the prospective contents of instruction. Thus, the test consisted of two "testlets" of distinct difficulty levels corresponding to the higher two of the three level divisions. Student performance on each testlet constituted the primary source of placement decision; i.e., sufficient performance on the Easier Testlet was judged as evidence of readiness for at least the Middle Level, in which case performance on the More Difficult Testlet was also consulted for decision on the student's readiness for the High Level. A Rasch-based calibration of item difficulty was conducted following a pilot testing with 1028 students, and the test revision took account of both quantitative test data and qualitative feedback data from individuals, who proctored and took the pilot test. The completed test was conducted to the target population of 1460 for the actual student placement and it functioned in a reasonable way. Prof. Shiotsu also illustrated some of the known issues and limitations of the approach.

Reported by Yuji Nakamura (*Keio University*)

Validation of the EBB Scales for the Story Retelling Speaking Test

Akiyo Hirai (*University of Tsukuba*)
Rie Koizumi (*Tokiwa University*)

The EBB (Empirically derived, Binary-choice, Boundary-definition) scoring scales are claimed to be easy to use and highly reliable though required less time for rater training. However, it has been questioned whether the EBB scales can be applied to any tasks and whether scores rated by the EBB scales can be closely related to the scores rated by other existing scales. This presentation focused on these validity issues regarding the EBB scales. Hirai and Koizumi developed a highly practical semi-direct speaking test called the Story Retelling Speaking Test (SRST), and in the context of the SRST, they compared the reliability and characteristics of the EBB scales with those of the modified FSI analytic scales. Fifty-two low-intermediate EFL learners were asked to read and retell each of the four stories in the SRST sessions. Later, their SRST performances were evaluated by six raters using the EBB and the analytic scales, both of which had four rating criteria. D-studies in the generalizability theory were carried out using mGEMOVA to examine how many stories are needed to achieve high reliability. Multi-facet Rasch analyses (partial credit scale model) were also made using Facets

to identify the characteristics of the EBB and analytic scales. The results showed that in order to achieve a generalizability coefficient of .70 or more, two stories would be necessary except for the grammar criterion in both rating scales. The results also showed that the EBB scales, though less practical, were more appropriate than the analytic scales in terms of the stability of rater severity and the difficulty of levels in each criterion.

Reported by Katsumasa Shimada
(*Momoyama Gakuin University*)

Exploring Korean Students' Perceptions of Test Fairness

Jungtae Kim (*Yonsei University*)
Seo Young Yun (*Hankuk Univ. of Foreign Studies*)

Professor Jungtae Kim of Yonsei University and his student, Seo Young Yun of Hankuk University of Foreign Studies reported on their research using the questionnaire data to examine the perceptions of test fairness concerning the TOEFL iBT among Korean students. As you might expect, they found that the perception of fairness differed according to the reasons the students had for taking the TOEFL iBT. But, in my opinion, the most important finding was that perceptions of cultural fairness are multi-faceted. The findings of Professors Kim and Yun should make us wary of speaking of test fairness without considering the context of the test and the test takers.

Reported by Randy Thrasher (*Prof. Emeritus, ICU*)

Ratings of Korean students' English language oral proficiency

Hyun-Ju Kim (*Dankook University*)

Professor Hyun-Ju Kim of Dankook University, began by pointing out the variety of English used in the world and her study examined the attitudes of Korean, Malaysian, and UK English teachers toward the teaching of these varieties which have been called World Englishes. She discovered that not only did different groups of teachers have different attitudes toward World Englishes, their attitudes impacted not just their choice of materials but their teaching as well. Those with a negative attitude toward World Englishes were more likely to consider the teaching correct pronunciation and intonation more important than the teaching of other language skills. The issue of varieties of English has not been deeply considered by testing scholars in Japan and Professor Kim's study should give us much to think about.

I found both presentations from the Korea English Language Testing Association (KELTA) very stimulating and my only regret was that so few members of JLTA were able to attend these sessions.

Reported by Randy Thrasher (*Prof. Emeritus, ICU*)

シンポジウム：
カリキュラムの変革における言語テストの役割

コーディネーター 小泉利恵 (常磐大学)
パネリスト 李洙任 (龍谷大学)、
堀内香予子 (金沢工業大学)
片桐一彦 (専修大学)・
興水以久子 (山梨県立甲府第一高校)



シンポジウムでは、小泉がはじめに、Brown (1995, 2005) に基づき、カリキュラム・デザインの6要素(ニーズ分析、目的と目標の設定、言語テスト、教材開発、指導、プログラム評価)と言語テストがカリキュラム改善に貢献できる部分、そのために言語テスト使用者が考慮すべき点を概括した。具体的には、言語テストは、カリキュラム・デザイン要素の必要な側面を決定・修正する際に有効に使え、テスト結果に基づき、(a) 目標・教材・指導方法・プログラム全体の方向性などを決定・修正でき(例：指導すべき要素・指導する必要のない要素の特定、教材のレベルの決定)、(b) 使用したテストが適切だったかを確認することができると述べた。

その後、3組のパネリストがそれぞれ所属の教育機関で用いているカリキュラムと言語テスト、そして両者の関係について論じた。李氏は、「言語テストの構成概念的妥当性の考察」をテーマとし、テストの得点解釈を正当化するためには、(a) テストの特徴と (b) カリキュラム上の目標、(c) 授業内で学習する英語の3点の一致を常に確認すべきであると述べた。アジアの中で英語「後進国」になりつつある日本において、国際社会で英語を使える人材を増やすために、学生のニーズを分析し、学生が目標を持ちつつ英語学習を続けられるようなカリキュラムを提供することが求められると主張した。

堀内氏は、金沢工業大学における「英語カリキュラムの工夫とTOEIC Bridge® IPの活用」について紹介した。学生の人生や将来の仕事において成功の糧となる英語コミュニケーションスキルの習得を目標とし、統一のカリキュラム・学習支援計画書(シラバス)を用い、毎回の授業で行う小テストと学生の興味に合わせた多様な指導を行いつつ、総括的に英語力の伸びを知るために外部テストを利用している旨が報告された。

片桐氏・興水氏は、「SELHiにおける諸テストの活用とその成果と今後の課題」を述べた。SELHi 指定時に、インプットからアウトプットにつなげるためのシラバス開発と授業展開を行った結果を、さまざまな英語のテスト(例：GTEC、TOEIC Bridge、語彙サイズテスト、スピーキング・テスト)を用いて検証した結果を報告した。さらに、テスト結果でシラバスや授業内容の改善等に役立った情報(例：GTECの各グレードの結果と定義、他校や過年度生との比較ができたテスト結果)や、役立たなかった・有効に活用できなかった情報(例：語彙サイズテストでの～語レベルという情報は、大学入試との関係が見えにくかったため使いにくかった。スピーキング・テストでの使用語彙数・流暢さ等の詳細な結果は、どの程度だと危機感を抱くべきなどの解釈の基準がなく利用しにくかった)についての言及があった。

質疑応答では、テスト選択の際にテストが測る能力を考慮に入れることの重要性や、英語力に伸びが見られなかった時どう解釈するか、頻繁にテストを使用しつつも長期的視点をどう保ち続けるか、テストが持つ力を有効に生かしたカリキュラム改革のあり方等が話題になった。

小泉利恵 (常磐大学)

講演
多読指導の効果を測定するテストの開発について
野呂忠司 (愛知学院大学)



本講演において野呂先生は、英語による多読活動の効果を測定するテストの開発をする際に考慮すべき点について論じられました。テストに利用するテキストの種類、設問の形式、時間の設定など、過去の研究成果を包括的に網羅し、言語テストを構築するには測定・評価の対象とする能力特性に関する理解が何よりも重要であることを示唆されたと思います。本研究大会で行われた発表者の発表内容を講演の中で言及されるなど、聴衆との距離を近づけようとする野呂先生の姿勢に共感しました。野呂先生素晴らしい講演をありがとうございました。

報告：伊藤彰浩 (愛知学院大学)

第25回研究例会 報告

第25回研究例会は、平成19年5月19日(土)に熊本大学くすの木会館レセプションルームにて開催されました。副会長の木下正義先生(福岡国際大学)の開会の挨拶に引き続き、2名の発表が行なわれ、その後、内容の充実した研究討議のセッションが催された。各発表の詳細については、次の報告をお読みください。

なお、今回のJLTA第25回研究例会の開催にあたりましては、(財)熊本国際観光コンベンション協会より開催補助を賜りました。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。

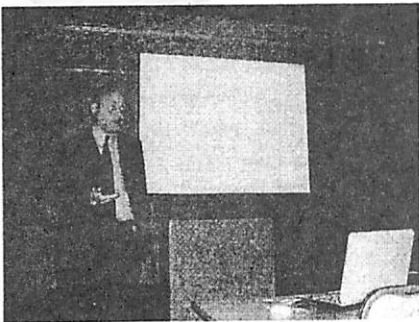
新緑の熊本は爽やかでした。会場校の島谷浩先生(熊本大学)、どうもお世話になりました。

(日本言語テスト学会 事務局)

Workplace Language Benchmarking (What's it really all about?)

Mark Knight (財団法人日本英語検定協会
STEP BULATS 事務局(アドバイザー)/
Hong Kong Polytechnic University)

今回の発表のために遠く Hong Kong からお越しくくださった Mark Knight 氏は、BULATS の Hong Kong での運営・実施・分析・企業へのアドバイス活動等について発表された。BULATS をただ単に受験して評価を受け取って終わりということではなく、相談を受けた企業のある職種の従業員はどんな英語運用の場面や運用能力が求められているのかをまず調査し、その後4技能それぞれのBULATSのテストでどれくらいの評価をとることができる英語力が求められているかという観点から調べ、その調査結果を企業へアドバイスしている Hong Kong での活動例が紹介された。たとえば、「Hong Kong のすぐお隣のマカオで合法化されているギャンブル場において、ディーラーが求められている英語力とは?」「ギャンブル場で楽しんでいる対面のお客さんとの会話程度の英語運用だけではなく、頻繁に本社の上司に電話で報告したり取りする英語運用の場面とその能力が求められている。」といった日本では想定外の仕事の事例なども紹介され、聴衆の興味を引いていた。



なお、BULATS は、ケンブリッジ ESOL が開発したビジネス英語能力テストで、STEP BULATS は、日本のニーズに適したテストとして、BULATS をベースとして、ケンブリッジ ESOL と STEP が共同開発したテストとのこと。日本国内における STEP BULATS の運営・実施・テスト内容等については、<http://www.eiken.or.jp/bulats/>にて詳述されている。

報告: 片桐一彦(専修大学)

第二言語習得過程とパフォーマンステスト(ACE
ライティングテスト1万人分の結果分析から)
佐藤 恭子(英語運用能力評価協会(ELPA)編集部)

佐藤氏は、英語運用能力評価協会で、ACE テストシリーズとライティングテストのコンテンツ・評価基準の作成から、採点システムの設計、採点者トレーニングまで、全開発工程に携われ、2004年末から2005年3月にかけて実施されたモニターテストの分析結果をもとに、2005年10月から始まった ACE ライティングテストの本テストにも関わられた。今回の発表では、全国の高校生、約14,000名を対象としたモニターテストの開発経緯と分析結果が中心に報告されたが、テストの特徴は次の4点である。1)「書く目的」と「書く過程」を重視したテスト設計、2)「書く楽しみ」を提供する内容、3) 質と量の両面をバランスよく評価するシステム、4) 誤りを指摘する添削ではなく、書き直しにつながるヒントを提示するフィードバック。評価方法の特徴としては、コンピュータによるデータ分析結果と複数の採点者による観点別評価(5観点)を総合的に評価したところにあるが、全受験者の英文がコンピュータ入力されていることと、5観点の評価は、5人の採点者が1観点のみを評価する形式をとって、ぶれを防ぐ努力がされているところが驚異的であった。受験者へのスコアレポートに個別のアドバイスが記載されているばかりでなく、学校ごとに誤りの種類・頻度・割合などがレポートにまとめられて報告されており、教育現場に密着したテスト実施姿勢に感銘を受けた。受験者と指導者の両者に非常に貴重な情報をもたらし、まさに現在求められているコミュニケーション能力の指導と評価に有意義なテストであるが、残念ながら、諸般の事情で本テストは今年3月で中断となっているようだ。実施可能性に難があるのであろうが、このような親身なライティングテストが定着するような方策がとられ、早い時期に再開することを心から願っている。

報告: 島谷浩(熊本大学)



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

●JLTAの最初の数年は発表は半分ぐらい、いやそれ以上 OHPで行っていたのではないかな。自分もスライドを次々とめくっていた記憶がある。●今日の学会ではもちろんパワーポイント以外はありません。テストもコンピュータを用いたものが次々と増えている。コンピュータのない生活はもちろん今では考えられないが、ある報告によるとコンピュータのトラブルで失う時間は一人合計数年という。ジャンクメールに費やすのは一人一日 4.4 分だそうだ。くだらないホームページを見る時間に費やされる時間と電力はどれぐらいになるのだろう。●もちろん OHPに戻れといっているのではなく、当たり前ここにコンピュータを少しクールに眺めてみるのもたまにはどうだろうか。何よりコンピュータで見失うものもあるのではないかな。発表ではメッセージが、テストでは問題内容そのものが本質であって、コンピュータで美しく見える、効率よく見えるのは装いにすぎない。テクノロジーに魅せられるのは確かだが、それが覆い隠すものも多い。テレビを見ない日のように、コンピュータを使わない日を設定してみようか。(HS)

日本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>.

