

JLTA Newsletter No. 28

日本語テスト学会

The Japan Language Testing Association

JLTA Newsletter No. 28 発行代表者: 浪田克之介 2008年(平成20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://www.avis.ne.jp/~youichi/JLTA.html>



巻頭言

浪田克之介

日本語テスト学会が設立されて12年が経過した。その間初代会長の友大先生と2代目会長のスラッシャー先生そして両会長を支えた事務局の強力なリーダーシップのもと、本学会はその会員数においても、また活動内容においても大きな発展を遂げてきた。

ところで日本では、昨今言語テストに限らず、広くテストの結果に社会の大きな関心が向けられている。TOEFLにおける日本人受験者の結果が国際比較で低いことは早くから批判の対象となっているが、加えてPISAの順位も落ちてきたなど、話題に事欠かない。さらに国内的には入試センター試験の活用が定着し、文部科学省による全国学力テストが再開されるなど、テストはきわめて日常的なこととなっている。

その一方で、これら大規模テストの実施に、我が国ではどこまでテストングの専門家が関わっているのだろうか。たとえば、入試センターにはテストングの研究者がいることは確かだが、各科目の出題者はそれぞれの分野の専門家ではあっても、そのことだけで適切な出題ができるのか。また、本年度初めて、海外で訓練を受けた看護師や介護士を日本は受け入れようとしている。当然のことながら、日本語の研修にも力を入れると報道されているが、今度は単に日本語の専門家のみでの出題で、それぞれの専門分野に必要な日本語力を的確に測定できるのだろうか。

この点で我々の参考になると思われるのは、すでに海外から多数の移住者を受け入れているカナダやオーストラリアなどで実施されている分野別言語テストである。テスト問題の作成に当たって、テストングと各分野の専門家の綿密な共同作業が見られるのである。

これまで言語テスト学会が積み上げてきた貴重な研究成果を社会に積極的に発信し、専門学会としての貢献をすることも我々に課せられた責務ではないか。

第12回 日本言語テスト学会 全国研究大会 研究発表

於：常盤大学

平成20年9月14日(日)

「英語脳」の持つ能力の特徴とその原因について(GTEC データ分析による考察)

山下 仁司 (株)ベネッセコーポレーション

本発表は、約650名の成人を対象とする調査結果を基に、「英語で発信する際に、ほぼいつも、日本語を介さず、英語でそのまま考えている」または「大体そうしている」と回答した人を「英語脳」を持つ学習者と定義した上で、その人たちの英語能力の特徴及び学習行動を分析することにより、英語学習・教育への示唆を提供した。「英語脳」学習者の英語能力の特徴として、ライティングにおいては文章の構成と綴りや句読点などの書記法、スピーキングにおいては文法・語彙・流暢さの点で、日本語を介在させてしまう学習者と統計的に有意な差が見られた。この点について山下氏は、「英語脳」学習者は、言語知識が自動化されているため、文章の構成全体に注意を向けることができ、正確かつ流暢に英語を使いこなすことができるのかもしれない、と解釈した。発音の正確さは「英語脳」とは関係がないようである、という指摘も興味深かった。「英語脳」学習者は、自分にとって効果的な学習法を工夫するといった自律性があり、英語で日記を書く・母語話者と会話をするなど、自動化・無意識化のための訓練の機会を作ろうとする学習行動の特徴なども示唆に富む調査結果であった。発表後、翻訳者や通訳者のように同時に二言語を操ることのできる人は「英語脳」を持つと言えるのか、「英語脳」というものが本当にあるのか、「英語で考える」とは一体どういうことなのか、といった質問が寄せられた。

小林美代子 (神田外語大学)

コミュニケーションの不確定性について—ルーマンのコミュニケーション論からの説明—

柳瀬 陽介 (広島大学)

言語テストにおける鍵概念の一つ「コミュニケーション」の社会的側面をルーマン(Niklas Luhmann)のコミュニケーション論を導入することによって探究した。意味を二項ではなく、三項(「情報」(Information; information)、「告示」(Mitteilung; utterance/announcement)、「理解」(Verstehen; understanding))で捉える考え方や、コミュニケーションにおける自己準拠自己準(Selbstreferenz; self-reference)と二重の偶発性(Doppelte Kontingenz; double contingency)などの概念、「システムと環境の差異」(Differenz von System und Umwelt; difference between system and environment)が実例とともに、言語テストとの関わりにおいて大変興味深く展開された。コミュニケーションは言語学や言語テストの分野で捉えられていた以上に不確定性の高いものであることが示され、テストはすなわちコミュニケーション、コミュニケーション能力を測定することはできないという点が論証された。言語テストの分野においては全く新しい捉え方が示されたと思うが、平易な発表だったため、単なる理解の確認ではなく、内容について本質的な質問、議論が行われた。

渡部良典 (上智大学)

Language ability structure and Chapelle's (1999) construct definition and interpretation: Findings from the reanalysis of multitrait-multimethod studies

Yo In'nami (Toyohashi University of Technology)

Rie Koizumi (Tokiwa University)

In'nami and Koizumi quantitatively synthesized previous MTMM multitrait-multimethod studies based on Chapelle's (1999) three perspectives toward construct definition and interpretation to determine whether test performance would be attributable to trait only, context only, or both trait and context. They conducted an extensive literature search including

databases, reviewing books and journals, and communicating with other research studies, and carried out confirmatory factor analyses to estimate 10 rival models for each MTMM matrix, each of them being evaluated using the overall fit of the model with the observed data. The presenters summarized the findings in a way in which it was very easy to follow even for novice researchers in the field. Within a limited time, a number of questions were raised, and lively and constructive discussions ensued.

Yoshinori Watanabe (*Sophia University*)

語彙サイズテストの開発と大学英語教育プログラムの教育効果検証の試み - 項目応答理論の大学英語教育への活用 -

齊田智里 (茨城大学)

大学生用の語彙サイズテストを開発し、大学英語教育プログラムの前後に2回実施することで、教育効果の検証を試みた発表であった。

JACET8000の語彙リストから、1版は頻度レベルの低い語を、もう1版は頻度レベルの高い語をより多く抽出し、各100問とした語彙サイズテストを2種類作成した。両者に40項目の共通項目を含めた。抽出する語は、名詞・動詞・形容詞・副詞からとし、一定の頻度の範囲から5語ずつランダムに抽出した。筆頭語に第一義的意味の日本語を与え対象語とし、筆頭語を含む5つの語彙の選択肢から1つを選択させる方法とした。

大学1年生対象の英語教育プログラム開始時に、低頻度語が主な語彙サイズテストを成績上位者961名に、高頻度語が主な語彙サイズテストを成績下位者777名に実施した。項目応答理論を用いて全項目の特性値を共通尺度上で推定した。項目特性値が推定されている項目プールから、難易度の異なる2種類の語彙テスト(各30問)を作成し、1年間の教育プログラム終了後に実施した。プログラム前後の能力推定値を比較検討することによって教育効果の検証を試みることができた。

本研究によって、外部の英語熟達度テストに頼らなくてもより簡便により低コストで教育効果を測定する方法の一例が示された。発表後フロアーからは、語彙テストだけで英語力の伸びを測定することの波及効果について質問が出された。

報告者 片桐一彦 (専修大学)

子どもの英語力を測る: 語彙テスト開発の試み

小林美代子 (神田外語大学)

長谷川信子 (神田外語大学)

町田なほみ (神田外語大学)

2011年からすべての公立小学校の5,6年生へ英語活動が導入されることを念頭においた、小学校英語教育における英語力測定テストの一つとしての英語語彙テストの開発についての発表であった。

子どもの言語発達における語彙の役割の重要性(Cameron, 2001; Ogura, 1999)から、英語力測定テストとして英語語彙テストを採用した。語彙テストは、絵を基にした設問や状況に基づく設問(対話、物語、馴染みのある場面等)が特徴的で、診断的情報も提供できるようにし、一斉テストは次のような構成であった。(1)「音声(聞き取り)」の単語認識(R: 認識)、文理解(R)、談話理解(R)、(2)「アルファベット」の聞き取り・認識(R)、聞き取り・産出(P: 産出)、(3)「文字(読み書き)」の単語認識(R)、文理解(R)、単語綴り並び替え(P)、単語産出(P)。

779名の小学生によるパイロットテストの分析結果を踏まえて、本語彙テストの信頼性と各項目の特性を算出し、英語学習歴や学習期間などの要因とテスト結果の関連を探り、テストの妥当性を検討した。

結論として、信頼性が高く弁別力のある語彙テストであり、子どもたちにほぼ適切な難易度で、学年・学習歴を反映し、語彙知識の発達の程度を仔細に測定し、音声部分のみでも適正な評価が可能との報告がなされた。

報告者 片桐一彦 (専修大学)

就労外国人を対象とする日本語能力の測定 村上 京子 (名古屋大学)

「教室習得環境にある日本語学習者」向けに開発された日本語能力試験や OPI といった既存のテストを「自然習得環境にある就労外国人」の日本語能力の判定に用いることは、読み書き能力の不足や、テスト側からの歩み寄りがないとコミュニケーションが成立しない場合があるために、困難である。そこで村上氏は、自然習得環境にある就労外国人の日本語能力レベルを判定する目的で、Can-do-statements、口答試験、筆記試験を開発した。Can-do statements は CEFR を参照し「聞く」「話す」「やりとり」「読む」「書く」の各領域について 30 項目の自己評価表を作成した。口答試験はインタビューテスト、ロールプレイ、絵の記述テストから成る。筆記試験は名前などのカタカナが書けるか、受験者が日常生活で目にする文字の意味がわかるかなどを問う「読み書き判定シート」を作成した。これらのテスト間の相関係数を算出したところ、Can-do statements の「読む」「書く」と「読み書き判定シート」の結果との間には非常に高い相関が見られた。「話す」「聞く」「やりとり」とはある程度相関はあるものの、個人差が大きく Can-do-statements はガットマンスケールのような順序性を示しにくい状況にあった。日常生活で使用頻度の低い項目についてはできないことが多く、教室習得環境の学習者の結果とはかなり異なる傾向を示した。以上の結果をもとに、村上氏はさらに就労外国人のための日本語能力レベル判定方法を改善し、学習者支援のガイドラインを策定していきたいとしている。

齊田智里 (茨城大学)

Creating valid in-house can-do descriptors for a listening course

Tomoko Fujita (Tokai University)

This research shows the attempt to create can-do descriptors for a listening course at a university, and to conduct validity analyses through the comparison of the can-do results and the data of other paper tests :

General Proficiency Test and Listening Final Test. Based on CEFR, the European Language Portfolio, City & Guilds (International English Qualifications), TOEIC, and STEP, the author develops two sets of 28 can-do descriptors for the basic, intermediate and advanced level students. The first set (Can-do 1) was given in April and the second set (Can-do 2) was given in July, 2008, and the results were analyzed using IRT 1 parameter.

The research questions the author sets for this study are as follows:

- Difference between Can-do 1 & Can-do 2?
 - Relationships between Can-do (1 & 2) and other English paper tests?
 - Differences between Advanced, Intermediate, and Basic level?
 - IRT analyses are useful? → defining level of each can-do descriptor
1. The IRT analysis of the can-do results shows the value of θ of Can-do 2 is better than that of Can-do 1 in all the levels.
 2. Correlation between θ values of Can-do 1&2 and of General Proficiency Test is observed; the value is between .275 and .329. Correlation is a little stronger when Can-do 2 and Listening Test are compared.
 3. As for the difference between the levels, the improvement of Can-do 20 of the basic level is the greatest. In addition, Can-do 2 θ of the basic level correlates with Listening paper tests more strongly than advanced level.
 4. The results of IRT analyses define the difficulty level of each can-do descriptor such as the most difficult item 27, whose ID is analyzed as 2.01231.

Some future implications, such as improvements of wording and inclusion of the experience factor, are also given in the end.

Yukie Koyama (Nagoya Institute of Technology)

慶應義塾日本語版CEFRチェックリストを用いた学習者の英語能力レベルの記述の試み

跡部智 (慶應義塾大学外国語教育研究センター)

境一三 (慶應義塾大学外国語教育研究センター)

野上康子 (株式会社教育測定研究所)

片岡夏子 (株式会社教育測定研究所)

慶應義塾は小学生から大学生までを対象として、言語学習の整合性と質を包括的に高め、複言語複文化能力を備えた独立自尊の先導者の育成を目的として、「行動中心複言語学習プロジェクト」に取り組んでいる。本研究では、このプロジェクト内で年齢の異なる学習者(計3500人)を対象としてCEFR(スイス版)に基づく日本語版のCan-doチェックリストを作成し、調査を行った結果を分析したものである。(ただし、CEFRには初学者レベルに適した項目が無いので、小学生用には英検の4・5級のレベルも補足的に使用した。)この日本語版Can-doリストを小・中・高校用に各3セット作成し、これらを大学生にも配布し調査を実施した。この結果をもとに項目応答理論に基づいて推定した日本語版リストの項目困難度は、CEFRレベルとの整合性が高いものであった。一方、スキル別ではReading, Writing, Strategiesで校種別の差が大きいという結果を得た。しかし、外部評価テスト(CASEC)と自己評価の関連を4セクションで調べたところ、比較的高い相関値が得られた。また、自己評価による学習者の能力値を学校種別間で比較ところ、発達的变化の仕方がスキルによって異なっていることが報告された。

今後の課題としては、個々のstatementの難易度や識別力を検討してさらなる改善を図ること、小学生にも分かりやすい記述の仕方を工夫すること、自己評価—教員評価—客観テストの関連づけを検討することが示されたが、年齢層の幅の広い学習者を対象とし、しかも綿密な分析に基づいて自己評価の評価軸としての可能性を示唆する、意欲的な発表であった。

小山由紀江 (名古屋工業大学)

言語タスクの経験は Can-do アンケートによる自己評価にどのような影響を与えるか - 初・中・上級者のクラス経験と自己評価値との関連の検証—

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 (慶應義塾大学)

本研究は、「言語能力の自己評価を行う際に、評定者の習熟度により、言語タスク経験の影響は変わるか」という研究課題のもと、TOEICを受験した会社員を対象に、評定者が初級レベルの場合(TOEIC400、500)でも、中、上級レベル(600~800)と同様の傾向が見られるか、調査し、経験要因と自己評価の関係を明らかにしようとした。具体的には、400、500、600、700、800(各±10)点の各得点域でのタスク経験率6%以上の項目(400点:34項目、500-800点:各53項目)を抽出し、(1)タスク経験率の特徴の分析、(2)経験者・非経験者の自己評価点の差の検証、(3)自己評価点の差が大きかったタスク、小さかったタスクを難易度、スキル、シチュエーションの観点から、習熟度間、習熟度別に分析が行われた。その結果、発表者らの先行研究の結果と一致した点として、習熟度に関わらず、一般的に、(1)タスクの結果は自己評価点に影響を与える、(2)難しく、専門的、発信に関わるタスクは自己評価点の差が開く。つまり、経験の影響が大きい。(3)易しく、一般的、受信に関わるタスクは自己評価点の差が小さい。つまり、経験の影響が少ない、などが明らかにされた。また、新たに、(1)習熟度が初級レベルの場合(400)は経験要因の影響が少なくなる可能性、(2)習熟度が初級レベルの場合は難易度が中程度でも差が開く可能性、(3)難しい、あるいは発信スキルであっても専門性が低く一般的で経験率の高いタスクは専門的なタスクに比べ差が小さくなる可能性、が示された。そして、最後に結果を踏まえ、Can-do アンケートの留意点について触れた。質疑応答では、TOEICの総合点だけでなく、セクションごとの点数との関係やタスク経験の影響を受けたアンケート結果の解釈について

質問がなされ、いずれも今後の課題とされた。

滝沢雄一 (福島大学)

Estimating reading test task difficulty: Will a check sheet help?

Randy Thrasher (*Okinawa Christian University*)

Prof. Thrasher discussed the feasibility of using a check sheet to estimate the difficulty of a reading passage together with the questions used to determine the level of comprehension of the passage. In the light of variables that affect the nature of reading, in order to improve difficulty estimates, we need to include what Alderson calls reader variables as well as what he calls text variables in our estimates and consider both the text and the devices used to measure comprehension in our calculation. Prof. Thrasher also focused on a Relevance Theoretic perspective on reading task difficulty and claimed that instead of dividing text variables from reader variables, as Alderson does, we can look at language variables and the variables involved in the inferential aspect of communication. He presented the features of both the text and the questions he had asked his English-teaching colleagues and graduate students about. What he has learned so far is that (1) Estimating difficulty is far more complicated than he realized. (2) The attempt to develop a way to estimate difficulty is a very useful exercise. (3) Training is required to use the check sheet (probably any check sheet) well. (4) No check sheet will be able to overcome the problem of individual differences in what test takers bring to the test task.

滝沢雄一 (福島大学)

Comparison of Keyword-assisted, Practiced, and Impromptu Performances on an Oral Summary Classroom Test: A Pilot Study

Hidetoshi Saito (*Ibaraki University*)

The purpose of the present study is to compare student oral summary (story retelling) performance under three conditions, keyword-assisted, practiced,

and impromptu, and to examine the role of practice and the issue of transferability in classroom test performance. Thirty-one university students in an EFL classroom participated in the study. All of them took a standardized speaking test (TSST) before the classroom test. The participants were assigned two passages randomly chosen from three prepared passages. They were instructed to give an oral summary for one passage with keywords that they had chosen, but were required to practice for both passages. They were also given an additional passage to summarize on the spot. During the test session, the students orally summarized these three passages: the first passage, randomly chosen on the spot, with their prepared keywords at hand; the second passage they had practiced, whose prepared keyword list was withheld at the time of testing; and the third passage, given on the spot with no keyword assistance allowed. The results were analyzed through Rasch analyses for estimating performance and an ANOVA (condition x ability (TSST)) for testing differences in performance. Three trained raters rated the performances using a three-item rating scale (fluency, accuracy, and content). The results indicated that the three performances differed, but the difference between the keyword-assisted and practiced conditions was much smaller than the differences between these conditions and the impromptu condition.

A comparison of speech samples of monologic tasks in speaking tests between computer-delivered and face-to-face modes

Yujia Zhou (*Graduate School of Tokyo
University of Foreign Studies*)

The last few years have witnessed a growing interest in applying computer technology to the delivery of speaking tests. Despite concerns over delivery mode of computer on examinees' performance (e.g., Chapelle, 2003; Alderson, 2004; Chapelle & Dougals, 2006), little is yet known about whether and in what way examinees' oral performance differs when talking to a

computer from that when facing an interviewer. The present study thus addressed this issue by comparing speech samples in the computer-delivered and face-to-face modes with a focus on linguistic performance of monologic tasks. This study also examined the interactions between test modes and examinees' proficiency level. Seventy-nine Japanese EFL students took two speaking tests across modes. Examinees' speech samples were then compared on a range of measures for fluency, accuracy, and complexity. Results indicated a significant but mixed effect of test mode only on fluency: examinees used more dysfluent words in face-to-face mode but more fillers in computer mode. Further, the effects of test mode on oral performance were not found to relate to examinees' oral proficiency. Implications for language assessment and second language acquisition research will be discussed, and directions for future studies will be offered in the presentation.

日本人高校生のためのコンピュータを利用した半直接面接テストの有用性検証

米野和徳 (山形県教育県庁高校教育課)
米野先生の研究の目的は、コンピュータを利用した半直接面接スピーキングテストの有用性検証である。学習指導要領でオーラルコミュニケーション能力の重要性は強調されている。しかし学校現場では、テストはあまり行われていないのが現状である。その主な理由は実行可能性(テストの実施時間や採点時間)の問題である。その欠点を補うために、コンピュータを利用した半直接テストでのスピーキングテスト(the computer-mediated semi-direct interview test)を用いて有用性を検証した。テストの内容は、高校生160名、ALTとJTE二人の評価者、評価項目は1)適切さ1(応答内容)、2)適切さ2(発話の速さ)、3)正確さの3つを用いた。検証項目は妥当性、信頼性、実用性である。妥当性の検証はmany-faceted Rasch measurementを用いて項目分析(モデルとの適合性)や受験者からフィードバックをもとに表面妥当性につ

いて検証した。信頼性はintra & inter rater reliabilityを検証した。結果の概要は、評価項目はモデルと適合し、信頼性のある結果が得られた。コンピュータを利用した半直接面接テストはテスト作成から評価を終えるまでの時間等を考慮すると実用的であることがわかった。ただ、敢えて申し上げれば有用性(usefulness)の検証に authenticity, interactiveness & impactらの項目が残念なことになってなかったもので、この3つの要素をいれるとさらに有用性検証が深まる研究になっていくと思われる。今後の米野先生に期待していきたい。そして、その結果が他の高校にも波及することを願って止まない。

秋山朝康 (文教大学)

グループ・オーラルテストにおける評価改善の研究

曲明(東京外国語大学院)

個人やペアのオーラルテストと比較してグループ・オーラルテストはあまり研究されていない分野である。例えばある程度の高い信頼性を得るためには何人の評価者が必要であるかなど、興味あるトピックがグループ・オーラルテストには多い。この研究の主な目的は教育現場で行われるグループテストの信頼性と実用性を検証することである。テスト内容は中国語を学ぶ3人~5人の高校生グループで、授業で学んだことを評価する達成度テストである。評価項目は、1)語彙、2)文法、3)発音、4)流暢さ、5)自発性、6)回答の適切さ、であった。評価者3人で、評価者訓練を施したあとで評価した。テスト結果は一般可能性理論を用いて分析した。その結果、信頼性を示すG係数はある程度高い係数が得られた。そして、引き続きD研究をした。その結果によると、G係数を維持するためには、評価項目3つと2人の評価者数が最低必要であることが判明した。質疑応答では、テストで扱うトピックの種類やグループ内での生徒の組み合わせ等(男女、個人の特質、英語力の違い等)、また評価者が教師一人の場合は、どのようにG係数が変化するのかこれ

からの課題として提起された。また、グループ内での生徒の組み合わせ等により、どのように会話の内容が変化するかなどはこれから興味深い研究テーマであると言えよう。今後の研究が非常に楽しみな発表であった。

秋山朝康 (文教大学)

Examining an EFL writing rubric for classroom use

Masaki Michiko (Osaka International University)

Otoshi Junko (Gunma Prefectural Women's University)

Kuru Yukiko (Aichi Medical University)

While such writing rubrics as the ESL Composition Profile and the TOEFL writing scoring guide are well-known, rubrics still do not seem popular for classroom use in Japan. Recognizing the need for a user-friendly rubric designed to work at higher education classroom in Japan, we developed a rubric of our own, which consists of five dimensions and four levels and has descriptors for each level of each dimension both in English and Japanese (Nishijima, Hayashi, Masaki, Kinshi, & Kuru, 2007). Using threshold loss agreement indices, we verified the reliability of the rubric for small samples of college student writings (Kuru, Masaki, & Kinshi, 2007). In this study, we first examine the reliability of our rubric for larger samples of college student writings, using the data from writing classes at a public university in Kanto region, where eighty-three international communication majored students wrote timed essays on a TOEFL Writing Essay prompt. Two instructors and the students rated the essays using our writing rubric and the TOEFL scoring guide, respectively. To see the attitudes toward our rubric, we also conducted a questionnaire to the students. Using the same data, then we see the correlation between the assessments with our rubric and those with the TOEFL scoring guide as well as the correlation between the student self-assessments and the instructor's assessments. Based on the features of our rubric that such analyses will clarify, we finally discuss the

effective use of our rubric as well as its revision ideas.

エッセイライティング分析的評価の修辞面構成要素に関する考察

宮崎 啓 (慶應義塾高等学校)

ライティング能力分析的評価の、修辞面構成要素のうち、特に、content, organization, cohesion, voice の構成概念妥当性の検証結果が報告された。発表者の研究目的は、content と他の構成要素との重複を予想し、どの構成要素が content の強い predictor となりうるかを見出すことにあった。報告された実験データは、高校2年生74名に与えられた説得文エッセイの採点データ(有効データは70名)で、日本人英語教員1名と Native の大学英語教員1名によって採点されたものであった。重回帰分析の結果、content と voice、organization と cohesion 間に高い相関があったことが報告された。この分析結果は、修辞面の評価は、content と organization の2点のみで十分であることを示唆し、修辞面を content と organization のみで分析評価している ESL Composition Profile (Jacobs et al., 1981) や Schoonen's analytic rating scale (Schoonen, 2005) を支持するものとなった。フローアからは、報告された評価者間信頼性は高いとは言えず、rater training が必要であるという指摘があった。実際、採点データにおいて、coherence のスコアが評価者間で差が大きく、分析対象から除外されたこと、そして、評価基準の記述子解釈の相違がその一因となったと報告されていた。本発表は、既存の評価基準の記述子にもあいまいな要素があり、評価に影響を与えている可能性が大きいことを示唆した点においても興味深かった。ライティングテストの信頼性を高めるために、明解な評価基準はますます重要とならずで、今後の研究の進展に期待したい。

島谷 浩 (熊本大学)

教員によるスピーチ評価と学生のピア評価との関係

奥田 利栄子 (広島大学)

大津 理香 (茨城大学)

教員による評価(TA)と学生間のピア評価(PA)との一致の度合いを探究した研究で、1) 評価基準についての説明の違いがTA-PAの相関に影響を与えるか、2) 学生は、教師の評価に匹敵する適切な評価を行うことができるか、3) 学生にとって正しく評価できる項目とそうでない項目があるかに関する調査結果が報告された。1)については、実演を交えて説明されたグループとそうでないグループの間に有意差が報告された。2)については、実演を交えて説明されたグループでは、5つの評価項目(Pronunciation, Eye contact, Fluency, Content, Overall)においてTAとPAは高い相関があったことが報告された。3)については、Accuracy(文法的な正確さ)に関する評価においては、TAとPAの相関が低かったことが報告された。つまり、評価基準について学生の理解度が十分な場合には、PAはTAに匹敵するレベルで、一部は最終評価に組み入れ可能なこと、ただしAccuracyの評価は除外する必要があることが示唆された。学習者の相互評価は、有意義なものとして定着しつつあるが、その信頼性について疑問の声があるのは事実だが、本研究結果は、学習者の相互評価の信頼性を、一部であるが立証しており興味深いものであった。研究対象となった学生の熟達度がある一定範囲であったこと、また評価者のサンプルが少ないなどの限界が発表者より示されたが、教育現場への示唆に富む研究内容で、フロアーからも好意的に受け止められていた。

島谷 浩 (熊本大学)

スピーキングにおける生徒相互評価と自己評価の妥当性:日本人高校生の場合

深澤 真(茨城県立竹園高等学校)

深澤氏はスピーチの評価方法としての生徒相互評価と自己評価の妥当性についての研究結果について発表を行った。同氏

の研究の目的は、(1)4つの下位項目における評価項目(発音,文法,流暢さ,発表態度)について、教員による評価と比べた生徒相互評価の妥当性に差があるかの検証と、(2)総合評価も含めた生徒相互評価と自己評価の妥当性に差があるかの検証の2点であった。研究は、日本人高校生79名(有効データ数62)を対象に実施され、主に併存的妥当性の観点から生徒相互評価と自己評価の妥当性の検証及び比較を行なわれた。研究の結果として、次の2点が明らかになった。第1に、生徒相互評価において、4つの評価基準のうち、流暢さ、発表態度については教員による評価と比べても十分な妥当性が認められたが、文法、発音に関しては部分的にしか十分な妥当性が認められなかった。第2に、総合評価を含めた5つの評価項目について生徒相互評価と自己評価を比較すると、それら2つの評価の妥当性には差があり、特に総合評価、流暢さ、発表態度で妥当性の差が大きいことがわかった。研究発表後活発な質疑応答があり、特に自己評価者1名に対して相互評価者の人数をどのように取り扱うべきかについて議論がなされた。

中村 優治(慶應義塾大学)

スピーチ・プレゼンテーションの評価_教師の総合評価と客観的測定値および生徒のアンケート結果との相関について

藤森 千尋(東京大学大学院)

藤森氏は、高校1年生、オーラルコミュニケーションの授業において行われたShow & Tellのスピーチ・プレゼンテーションを、正確さ・流暢さ・複雑さの3つの観点から、それぞれに適した客観的測定方法によって量的に分析し、日本人教師(JT)と外国人講師(ALT)の各総合評価との相関を調べたものに関して発表を行った。同氏はまた、聴衆であるクラス生徒たちに、一人一人のスピーチの印象に残っている項目を選択するアンケートを実施し、その結果と教師の総合評価との相関についても考察を行った。結果として次のようなことが明らかになった。まずJTとALTの総合評

価には非常に高い相関が見られたが、総合評価と正確さ、流暢さ、複雑さに関する、いくつかの客観的測定値との間に相関は見られなかった。総合評価と相関が見られたのはスピーチの総語数と総節数であった。またアンケート結果との相関については、5項目---流暢さ、態度・話し方、テーマ、内容の豊かさ(複雑さ)、英文の分かりやすさ(正確さ、簡潔さ)のうち、態度や話し方がJTやALTの総合評価と高い相関を示した。流暢さに関しては、JTの評価と相関が見られ、英文の分かりやすさについては、ALTの評価と相関が見られた。5項目合計についてはALT評価とJT評価ともに相関が見られた。発表後の質疑応答ではspeakingとwritingの両方に関係するような項目の検討や、speaking特有項目の量的追求に関して議論が行われた。

中村 優治(慶應義塾大学)

Text length and language for recall: Their influence on L2 listening test-taking strategies

Hideki Sakai (Shinshu University)

Professor Sakai's presentation was on the effects of language of recall (L1 in the free written recall test versus L2 in the dictation test) and text length (text with short segments versus text with long segments being played at one time) on test takers' L2 listening strategies. Thirty-eight Japanese EFL students were made to listen to a text with long segments and answer a 26-item, 6-point scale questionnaire designed to measure listening test-taking strategies. They were also required to listen to a text with short segments and answer the same questionnaire. Half of the participants were requested to respond in their L1, and the other half were requested to respond in their L2. The results showed that text length exerted a greater influence on listening test-taking strategies than did language in which material was recalled. Further, the dictation that had a text with long segments elicited a wide range of strategies.

The audience had several questions and suggestions. One suggestion was to group

the responses to questionnaire items designed to measure the same test-taking strategy (e.g., guessing, planning) rather than interpret the responses to each questionnaire item. The presenter agreed with this suggestion. A question that was asked pertained to whether the questionnaire items were based on the literature on test-taking strategies in general or on test-taking strategies in listening. The presenter's response was the latter. Another question concerned whether there were differences in strategy use between test-takers who performed well and those who performed poorly in the test(s). The presenter stated that this would be a direction for future research to pursue. By and large, the presentation was very well received, and the question and answer session proved insightful.

Yo In'nami (Toyohashi University of Technology)

大学入試センター英語入試リスニング試験の妥当性検証: 社会・認知妥当性の枠組みから

柳川浩三(神奈川県立大和西高校)

本発表では、大学入試センターリスニング試験の妥当性を、社会・認知妥当性の枠組みを用いての検証の結果が論じられた。社会・認知妥当性とは、実際の言語使用における社会・言語的コンテキストと言語使用者の認知処理とが、テストタスクにどの程度反映しているかを問うものである。具体的な検証の観点とは、以下の3点である。1) 学習指導要領は、実際の言語使用時の認知処理と言語社会的コンテキストをどの程度反映しているか。2) 大学入試センターリスニング試験は、言語・社会的コンテキストをどの程度実現しているか。3) 大学入試センターリスニング試験は、学習指導要領をどの程度反映しているか。

検証の結果を踏まえて、次のことが指摘された。1) authenticity に欠ける点がある(例えば、2度聴けること、話される英語に fillers や hesitation marks がないことなどがあげられる)。2) 使用語彙については、学習指導要領を反映している。3) 討論のよう

な場面が使われていない、英語の varieties が反映されていない、という点では学習指導要領を反映していない。4) pragmatic competence を要求する問題が少ない、など。

塩川春彦(北海学園大学)

「見えないもの」を測る：日本語談話の省略要素の解釈について

森谷浩士(神田外語大学)

本発表は、日本語の談話の特徴である「省略」に着目し、談話処理方略の転移という観点から、第二言語としての日本語学習者の理解のプロセスを探ったものである。発表者は、英語が日本語や中国語ほど省略を許さない言語であることに着目し、中国語母語話者と英語母語話者の間で、母語背景の違いが、省略の多い日本語の談話の解釈に影響を及ぼすか否かを考察した。

「解釈」に焦点を当てる理由として、まず先行研究が言及された。先行研究では省略に関しても産出データを扱うものが中心であることが指摘され、その上で、省略部分が産出データに現れないことは習得の証明にならない、と論じられた。

検証においては、省略のある自然な文章と、省略要素を復元した文書を読んで、読みやすさに違いがあるかを判定するタスクと、省略を含む文章を読んで、省略要素を特定するタスクが実施された。

結果として以下のことが指摘された。1) 中国語母語話者は復元形を読みにくいと評価した。2) 英語母語話者は、先行隣接文の要素を省略要素と見る傾向があり、談話始発文内の要素を省略要素と見る傾向があり、省略要素を特定するタスクの正答率において中国語母語話者の方が高かったことから、省略要素の解釈に違いがあった。

塩川春彦(北海学園大学)

ESL vs. EFL in terms of learning difficulty of implicit and explicit knowledge

Kazuyuki Shite (Tokyo University of Social Welfare)

The paper aims to investigate whether some grammatical structures are easy in terms of implicit knowledge but difficult in terms of explicit knowledge or vice versa for Japanese EFL learners, in comparison with Ellis' (2006) findings based on an ESL setting. A questionnaire and four tests were administered to 57 Japanese learners of English from two universities. The tests employed were timed GJT (grammatical judgment test), untimed GJT, META (metalinguistic knowledge test) and OIT (oral imitation test). A principal component analysis was performed and two components were extracted, as in Ellis (ibid): one (implicit knowledge) for OIT and timed GJT and the other (explicit knowledge) for untimed GJT and META. The difference in mean percentage score between the tests for the first component and those for the second component was computed. Ten structures (out of seventeen) were found easy in explicit knowledge and difficult in implicit knowledge, with a 20 percentage point difference or more, two others showing the opposite relationship. It was concluded that the tests measured the learners' implicit and explicit knowledge separately regardless of the ESL/EFL difference while Japanese EFL learners seemed to have developed explicit knowledge more than implicit knowledge. A major pedagogical implication is that teachers may be able to pinpoint students' strengths and weakness in terms of implicit and explicit knowledge of grammatical structures by using such tests before and after classroom instruction. Some questions from the floor address the validity of PCA in the confirmatory analysis of two pre-positing constructs and no others.

Ken Norizuki (Shizuoka Sangyo University)

明示的知識と暗示的知識：時間制約の有無、文法的適格性 / 非適格性、産出 / 判断の関係

島田 勝正 (桃山学院大学)

Ellis (2005) の研究結果に反して、島田 (2007) は、文法性判断テスト (GJT) の提示時間制約の有無 (timed (TM) / untimed (UT)) が、明示的知識 (EK) と暗示的知識 (IK) の測定に必ずしも影響しないことを示した。今研究は、EK / IK の測定と TM / UT との関係をさらに追及し、項目の文法的適格性 (grammatical: GR; ungrammatical: UG)、処理過程 (産出 / 判断) との関係も検証することを目的とした。TM / UT 両モードの GJT の GR / UG 項目、oral imitation test (OIT) の GR / UG 項目、error correction test の UG 項目を対象に、探索的因子分析を行うと、2 因子解では、F1 が UG 項目、F2 が GR 項目に、3 因子解では、F1 が UG 項目、F2 が GR 項目、F3 は OIT 項目に負荷が高かった。モデルの適合度により、2 因子解は棄却され、3 因子解が妥当であることが確認された。今研究においても TM / UT は「判断」に関与せず、UG 性の F1 が EK、GR 性の F2 が IK、OIT 項目の F3 は、「産出」に関係する EK / IK とは独立した因子を表していると考えられ、EK と IK の関係を 2 項対立ではなく、連続体の中で解釈する有効性が示唆された。将来、Amos 等を用いて、強力な仮説に基づく確証的因子分析を行う必要性が指摘された。

法月 健 (静岡産業大学)

音韻認識とスペリングの関係

高波 幸代 (筑波大学大学院)

まずは、小生が発表時間の司会を間違えて数分遅れたことに、発表者に陳謝したい。

発表者の高波先生はパワーポイントを使用して、次の項目に関しての発表をされた。

1. 音と文字 (個々の音と記号の関連性と蜜接さ・英語のつづり覚えることは音韻処理との関係) 2. 音韻認識とは 2.1 音韻認

識とスペリングの関係 2.2 音韻認識課題に関する詳細な説明。3. 研究の目的と調査質問 3.1 研究の目的 スペリング = 音素と書素 = 音と文字 3.2 調査質問 4. 調査方法 4.1 参加者 4.2 調査手順 4.3 材料 4.4 学習方法 1) P 文字は書かない 2) W 発音はしない 3) PW 発音練習 + 文字練習 5. 調査で使った単語 5.2 単語練習 5.3 単語の文字数 6.1 アンケート結果 6.1.2 学習方法に関する、事前アンケートでは英語のつづりを覚える一番よい方法だと思うものを 1 つ選びなさい P 4%・W 38% PW 58% 6.1.3 事後アンケート 最も効果的だったと思う方法を選択しなさい P 3% W 41% PW 58% 6.3 分析結果 グループと学習方法では グループと学習方法の交互作用なし 単語の文字数による比較 では単語の文字数と学習方法には関連性が見られ 6.4 考察 学習方法 P において、文字数 10 文字になると、得点が下がり始める PW は文字数による影響は受けない 結論として音 + 文字で学習する必要性があると説明した。司会者からコメントを述べた。参加者 4 名。

木下 正義 (福岡国際大学)

Rasch モデルに基づくアダプティブテストシステム開発の試み

秋山 實 ((株)e-ラーニングサービス)

テストの結果に客観性を重視する立場でオンラインプレースメントテストシステムを構築しようと試みている。客観性を重視するため、Rasch モデルを基礎にする必要がある。オンラインプレースメントテストは、確実に実施でき、かつ、短時間に実施できなければならない。確実に実施するためには日常使用しているツールを用いることが一つの方策であり、オープンソースの LMS である Moodle を採用した。短時間に実施するためには、アダプティブテスト方式を採用するのが最善である。秋山・今井は、J-CAT を開発し山口大学の留学生を対象に日本語クラスのプレースメントテストで使用して実用性を確認した。しかし、この方

式の理論的な解析や限界については明らかにされていない。秋山は、新たに Rasch モデルを基礎としたテストレット方式のアダプティブテストシステムのプロトタイプを Moodle のプラグインとして開発した。J-CAT と同じくテストレット方式を採用した。テストレットを構成する個々のアイテムの回答パターンから受験者の能力値を推定し、次に受験すべきテストレットを決定し受験させる。さらに受験したすべてのテストレットの回答パターンから更に受験者の能力値を推定する。受験すべきテストレットが無くなった時点でテストが終了する。本方式により短時間にテストを収斂させることができる。

言語テストの実施形態に関わる意思決定 - AHP (階層分析法)による支援の試み 劉 東岳 (プロメトリック株式会社)

近年の技術発展に伴い言語テストの様々な実施形態が可能になっており、測定理論の観点からは統計的な方法論による比較評価の研究(Comparability Study)が活発になされている。本研究では、幾つかの典型的な言語テストの実施形態の中から1つの実施形態を選出するという状況において、主催者の業務負荷や受験者の利便性などの「得てして主観的になりがちな評価軸」を含めた意思決定の過程を合理的に支援する目的で、複数のシナリオに基づいて AHP (Analytic Hierarchy Process, 階層分析法)の適用を検討した。確かに言語テストは「言語能力を測定する仕組み」ではあるが、世の中では「言語能力を証明する機会」としての役割も期待されることが少なくない。その意味で、言語テストの実施形態を比較し選択する場面では、「測定する仕組みとしての品質」だけでなく、「証明する機会としての品質」も大切な評価軸となるはずである。言語テストの実施形態を選択する意思決定の現場に、本研究の試みが貢献することが期待される。

Moodle 版 TDAP3.0 の開発

中原 敬広 (e ラーニングサービス)

大友、中村、秋山が開発したテストデータ分析プログラム TDAP2.02 を進化させ、オープンソースソフトウェアの LMS である Moodle のプラグインとして tdap モジュールを開発し、受験データをエクスポートせずにテストデータを分析し、その結果をアイテムバンクに登録できるように改善し、さらに今回はそれを最新版の Moodle1.9 に対応させた。その結果、以下のような機能が追加されることとなった：

- ・語彙レベル判定機能付アイテムエディタ
- ・項目情報関数やテスト情報関数のグラフ化
- ・ABC ファイルのインポート
- ・テストレットの作成
- ・受験者シミュレーター (= 受験結果を確認できるシミュレーション機能)
- ・CAT モジュールとの連携

今後の課題としては以下の4点があげられるとのことである：

- ・多肢選択問題以外に対応させる
 - ・アイテムバンク機能の強化
 - ・テスト作成機能の強化
 - ・2パラメータによる分析ツールの開発
- 大変嬉しいことに、このプログラムは e ラーニングサービスのホームページ (<http://www.e-learning-service.co.jp/index.html>) から簡単な利用登録をすれば無料で配布してもらえる。

大津 敦史 (福岡大学)

習熟度別クラス編成のための英語基礎力 判定標準化テスト作成の試み

木村 哲夫 (新潟青陵大学)

近年の大学入学者の英語力が多様化していることに鑑み、習熟度別クラス編成を導入する大学が増えてきている。本研究は、オープンソースウエアであり、且つ充実した機能を有する LMS として定評のある Moodle を用い、英語基礎力判定テストを作成・実施し、そのデータを管理・分析し、標準化テストとしてクラス編成の基礎資料

を提供する試みを行ったものである。テスト項目は自作のものと英検の過去問を利用し、分析には1パラメータと2パラメータのロジスティック・モデルを試行した。その結果、事前に misfit をうまく除去して、1パラメータ・ロジスティック・モデルまたは NTT (Neural Test Theory; Shojima, 2008) を用いて分析した方がうまく行くように思われた。特に、後者の NTT を利用すれば、解釈や判断がより容易に感じられた。また、本研究のテストで得られた学習者の能力推定値と他の外部テスト (TOEIC Bridge, CASEC) で得られたスコアとの相関分析も実施し、その妥当性の検証も行った。今後の課題としては、以下の4点があげられるとのことである:

- ・どのような項目がよりよい項目なのか精査する必要がある
- ・Misfit をもっとうまく取り除く方法を抽出する
- ・アイテムバンクを構築する
- ・CAT に対応させる

大津 敦史 (福岡大学)

ワークショップ

「SPSSを使って分散分析をやってみよう」
平井明代 (筑波大学)

長橋雅俊 (筑波大学大学院)

準備して頂いた24ページにわたる詳細なハンドアウトを元に、2時間のワークショップを理論編、実践編と2部構成に分け、初めに平井先生が、分散分析とは何か、そしてその理論を詳しく説明された。3つのグループデータを用い、全体の平方和、群間の平方和、群内の平方和のそれぞれの計算の仕方、計算式を明記し、パソコンで簡単に行えてしまう分散分析の中身を噛み砕いていただいた。理論を理解した後、後半は長橋さんが担当し、実際に参加者とパソコン教室にて、一元配置の分散分析を行い、質問を受け付けつつ進んでいった。まずエクセルファイルにデータをどのように作成するかを説明し、そのファイルを実際にSPSSで開いていく作業から始めた。SPSSでは何をどこに入れ、どこをクリックすれ

ばいいのかが分からなくなりやすいが、その点も実際のスクリーンをハンドアウトに示しているので、説明を聞きながら参加者も問題なく進んでいた。一元配置の後には、二元配置分散分析 (対応なし×対応なし)、二元配置分散分析 (対応なし×対応あり)、下位検定についてと話が進んでいった。最後には効果量の話があり、SPSSで求められるイータ二乗、偏イータ二乗などについての話もあった。参加者からも多くの質問があり、分散分析を行う際の前提条件、サンプルサイズ、球面性とは何か、多重比較を行う際の p 値の調整、など質問も多岐にわたり、活発な意見交換があり大盛況のなか終了した。

中西貴行 (常磐大学)

Plenary Speech: Defining, Setting and Validating Standards in Language Testing.

Tony Green (University of Bedfordshire)

Dr Green outlined some of the ways in which the word 'standard' is used by language testers. These include the process of determining cut scores through standard setting, quality standards (such as the JLTA code of practice) and the notion of a standard language as a target for learning. Dr Green focussed on the implications of these interpretations of 'standard' for the use of frameworks such as the Common European Framework of Reference for Languages (CEFR). He described how, in the interests of the standardisation of qualifications across borders and across settings, organisations are coming under increasing pressure to relate local schemes to such frameworks. However, relating tests to a CEFR level may serve to obscure important differences between tests: differences of purpose, content and quality. He argued that the key benefits of using a framework such as the CEFR may derive from engaging in the suggested processes of specification rather than in making claims of equivalence. These processes encourage test developers to reflect on test purpose and to communicate more

effectively with test takers and score users. He described how a collaborative programme of research – the English Profile – aims to provide reference level descriptions for English that will facilitate such processes for users of the CEFR.

Dr Green's speech touched upon the substance of CEFR and served to educate the audience about its aims. This was particularly useful since Dr. Green is obviously well versed in the goals of CEFR, having carried out significant research in the field for some time.

Tomoko Fujita (*Tokai University*)



Symposium

Setting and Validating Standards on Language Tests

Coordinator:

Tomoko Fujita (*Tokai University*)

Panelist:

Michiko Nakano (*Waseda University*)

Discussant:

Tony Green (*University of Bedfordshire*)

Tomoko Fujita (*Tokai University*)

The symposium was concentrated mainly on a practical use of CEFR and Can-Do statements. First, Prof. Michiko Nakano introduced the Tutorial English Program of Waseda University with the application of CEFR levels to the curriculum. This trial was based on the principle that educational reforms change simultaneously with digitalization and globalization. The Tutorial Program therefore aims at the important educational

achievement in creating independent learners, problem solvers, and practical language users in the IT and global society. The accurate survey and the IRT analysis were conducted to establish the Tutorial Program. What can be seen in the experiment is that there has been gradual appearance of learners' needs to have the learning environment for self-reflection as well as to improve learners' communicative competence in particular contexts. In this sense, as Prof. Nakano described, CEFR could be a helpful scheme to provide both learners and teachers with practical guidelines for what level the learners pass through and what to do next.

Next, Prof. Tomoko Fujita's presentation addressed the validation of placement test (PLT) for Can-Do statements (CDS). In order to make the placement system meaningful, Prof. Fujita laid particular stress on the content analysis, which is relevant to course objectives and learners' achievements of goals. Prof. Fujita also emphasized that learning is embodied in frameworks that relate constructs to contexts, and operationalized in test specifications that articulate purpose in practice. The quantitative analyses were performed to examine the extent to which the learners show their abilities in the contents based on the in-house CDS, showing that there is sufficient validity of developing the relationship between PLT and CDS. Fitting the appropriate test items into each level of the CDS may not be an easy task, but Prof. Fujita suggests that setting content standards for learners constructs an essential step to meet the future needs of language learning.

Last, Prof. Tony Green highlighted the important points of CEFR and CDS. Prof. Green pointed out that CEFR and CDS should be seen as a series of

guidelines from which tests or teaching methods and materials can be built to suit specific contextualized needs. Regarding the relations with tests, in the CEFR, there is less concrete suggestion of linkage between particular task types and testing situations; yet the CEFR can still be useful as a heuristic that allows for creative development of test specifications. Prof. Green laid emphasis on the belief that the focus of CEFR and CDS is largely on matters of process, such as meaningful learner development or diagnostic application, rather than those of product.

After the presentations, through the active discussions with the audience, all the three presenters clearly expressed their common idea that CEFR profile becomes an effective indicator of developing better learning and teaching guidelines. The discussions made us realize the importance of “standard setting,” by which we need to make decisions about whether learners have the language and communication abilities necessary to undertake particular activities. Kei Miyazaki (*Keio Gijuku High School*)

第 27 回研究例会 報告

第 27 回研究例会は、平成 20 年 7 月 12 日(土)に桃山学院大学 2 号館 CALL 教室にて開催されました。今回は、第 1 回中部地区英語教育学会近畿地区研究会との共催ということで、中部地区英語教育学会会員の方のご参加もございました。

青木昭六先生の講演に、島田勝正先生のワークショップ、伊藤潤先生の研究発表と、例会ながらも 3 つの種類を聴くことができました。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 27 回研究例会の開催にあたりましては、桃山学院大学より会場のご提供を賜りました。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。

大阪の 7 月は暑かったです。当日は、まだ梅雨明け宣言は出ていませんでしたが、結果的にすでに梅雨明けしていたようです。でも会場内はとても涼しく、また綺麗でモダンな校舎で快適な半日でした。

会場校の島田勝正先生(桃山学院大学)、どうもお世話になりました。

(日本言語テスト学会 事務局)



講演

「コミュニケーションのための『基礎体力』」 青木昭六(兵庫教育大学名誉教授)

まず始めに、Communicative Language Teaching (CLT) で育成すべき「基礎体力(vital force)」とは何かについて、お話がなされた。「コミュニケーションのための基礎体力」について Canale (1983), Cummins (1991), Widdowson (1978, 1983, 1990, 1992)を始め多数の文献を引用しつつ、多角的にまた分析的に説明がなされました。その際、具体的例文に関しても、Shakespeare の *Richard III*、白楽天の『長恨歌』、川端康成の『雪国』をはじめ、古今東西の様々な著名な文学作品からたくさん例示がなされた。

その後、「ではどのようにして、そのコミュニケーションのための基礎体力を育成していくのか?」といった話題に移った。指導力の構成、発問のタイプの種類、「指導の質」の検討、「有意味」の発問タイプ、発問の適切性、推論能力のための発問のあり方等について、中学校・高校の英語教育

で考慮に入れていくとよいと思われる点について、先ほどの学術的分析結果から得られた知見を元に提案された。同時に、そういった指導を行なった場合の効果に関する、Shimada (1992)等の実証研究結果についても紹介がなされた。

なお、上記の内容は、次の本に整理されて詳述されているとのことであった。青木昭六 2008.『コミュニケーション推進力としての推論能力』村田久美子・原田哲男編著『コミュニケーション能力育成再考』22-51. ひつじ書房。

細かい厳密な分析が多角的に多くなされている一方で、全体像を見失うことなく包括的に整理整頓がなされてまとめられており、また、古今東西の様々な文学作品から例文を採用するなど、青木先生の博覧強記の教養に感銘を受けた講演であった。

片桐一彦(専修大学)

ワークショップ

「エクセルを使った言語テストデータ処理」 島田勝正(桃山学院大学)

「選択肢(ABCD)を1,0 データに変換する」、「素点をtスコア(偏差値)に変換する」というエクセルを使った基礎的な言語テストデータ処理の方法に関して、参加者全員が実際にエクセルを動かしながら、ワークショップが行われた。

今回は、学生さんの参加が多かったということもあり、言語テストデータ処理に馴染みがなかった参加者の方にとっては、「言語テストデータ処理は取っ付きにくい」という壁が少しでも払拭されたようであった。

片桐一彦(専修大学)

研究発表

「L2 聴解テストにおける構成概念妥当性検証」

伊藤潤(広島学院高等学校)

L2 聴解テストにおいて、測定される言語技能(設問レベル)の心理学的次元は、低次元の設問と高次元の設問の2つに大きく分類されると考えられる。低次元の設問とは、textually-overt であり language-bound

なので、明示的な情報を聴き取ったり命題的な意味を理解する能力を測定する。一方で、高次元の設問とは、textually-covert であり cognition-bound なので、明示的な情報から類推したり発語内行為的な意味を解釈したりする能力を測定する。

心理学的に異なる2つの言語特性を測定するとして区別される低次元の設問と高次元の設問の2種の設問が、計量心理学的にも別次元の技能を測定する設問として分類できるかどうかを調べることを目的として、MTMM(多特性・他方法)による実験・分析をおこなった。

結果は、計量心理学的には2次元だと結論付けても問題はないとのことであったが、MTMMで5つの条件のうち2つが満たされなかった。MTMMで5つの条件のうち2つが満たされなかったということで、この点がどうしてなのか、質疑応答においてフロアーからも意見が出され、リーディングテストの場合の内容構成と聴解テストの場合の違いを対比するなどさらに突っ込んだ議論がなされた。

記録：片桐一彦(専修大学)

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



●今年も新しい新入生を迎え、感じるのは Speaking の能力の向上である。もちろん何らかのテストで継続的にデータを取らなければ、この主観的な「感じ」をきちんと説明はできないのだが、英語を堂々と話す学生の増加を感じるのは私だけだろうか？これがセンター試験によるリスニング導入がもたらした Washback なのか、それともこの時代（多量の英語メディアとの接触や外国人の流入）が少しずつ影響を与えているのか。やはりこれもデータをとらなければわからないのだが、嬉しいことである思いたい。(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>

