

日 本 言 語 テ ス ト 学 会 (JLTA)

第 13 回 全 国 研 究 大 会 (2009 年 度) プ ロ グ ラ ム

**Handbook
of
the 13th Annual Conference
of
The Japan Language Testing Association**

大会テーマ：History of Language Testing in Japan

日時：2009 年 9 月 7 日 (月) 8:30 ~ 17:35

会場：北海学園大学7号館
(〒062-8605 札幌市豊平区旭町 4-1-40, TEL (代表) 011-841-1161)

主催：日本言語テスト学会 (JLTA)

日本言語テスト学会 (JLTA)
The Japan Language Testing Association (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>

全 国 研 究 大 会 本 部 委 員

浪田 克之介	(北海道大学名誉教授)
中村 優治	(慶應義塾大学)
中村 洋一	(清泉女学院短期大学)
小泉 利恵	(常磐大学)
塩川春彦	(北海学園大学)

全 国 研 究 大 会 運 営 委 員

伊藤 彰浩	(西南学院大学)
法月 健	(静岡産業大学)
島谷 浩	(熊本大学)
中村 優治	(慶應義塾大学)
塩川 春彦	(北海学園大学)
藤田 智子	(東海大学)
Soo-im Lee	(龍谷大学)
島田 勝正	(桃山学院大学)
小泉 利恵	(常磐大学)
前田 啓朗	(広島大学)

全 国 研 究 大 会 実 行 委 員

塩川 春彦	(北海学園大学)
竹村 雅史	(北星学園大学短期大学部)

研 究 発 表 審 査 委 員

塩川 春彦	(北海学園大学)
伊藤 彰浩	(西南学院大学)
法月 健	(静岡産業大学)

The 13th JLTA Annual Conference

Place: Hokkai-gakuen University Toyohira Campus No. 7 Building

Date: September 7, 2009 (Monday)

Timetable

8:30 ~	Reception (No. 7 Building 4 th Floor)	
9:00 ~ 9:15	Opening Ceremony (Room D30)	
9:20 ~ 10:00	Presentation Part I	Workshop (Room D502) 9:20 ~ 11:20
10:05 ~ 10:45	Presentation Part II	
10:50 ~ 11:30	Presentation Part III	
11:30 ~ 12:30	Lunch Break (JLTA Business Meeting: Room D404)	
12:30 ~ 13:10	Presentation Part IV	
13:15 ~ 13:55	Presentation Part V	
14:00 ~ 15:10	Plenary Speech “History of English Language Testing in Japan” Masamichi Tanaka (Emeritus Professor, Hiroshima University) (Room D30)	
15:20 ~ 17:00	Symposium Theme: “Studies on Language Ability and Assessment” Coordinator & Panelist Miyoko Kobayashi (Kanda University of International Studies) Panelist Miyuki Sasaki (Nagoya Gakuin University) Sukero Ito (Tokyo University of Foreign Studies) (Room D30)	
17:05 ~ 17:25	General Business Meeting (Room D30)	
17:25 ~ 17:35	Closing Ceremony (Room D30)	
18:00 ~ 19:30	Banquet (Restaurant Cosmos)	

Exhibition: Room D401 Headquarter: Room D404 Lounge for Participants: Room D402

大会日程表

8:30 ~	受付 (7号館4階)	
9:00 ~ 9:15	開会行事 (D30 教室)	
9:20 ~ 10:00	研究発表 Part I	ワークショップ (D502 教室) 9:20 ~ 11:20
10:05 ~ 10:45	研究発表 Part II	
10:50 ~ 11:30	研究発表 Part III	
11:30 ~ 12:30	昼食 (JLTA 役員会 : D404 教室)	
12:30 ~ 13:10	研究発表 Part IV	
13:15 ~ 13:55	研究発表 Part V	
14:00 ~ 15:10	講演 (D30 教室)	
15:20 ~ 17:00	シンポジウム (D30 教室)	
17:05 ~ 17:25	総会 (D30 教室)	
17:25 ~ 17:35	閉会行事 (D30 教室)	
18:00 ~ 19:30	懇親会 (レストラン・コスモス)	

業者展示: D401 教室 大会本部: D404 教室 参加者休憩室: D402 教室

昼食は、受付時に受付で弁当引き換え券を購入していただき、昼食時にレストラン・コスモスで弁当と引き換えてください。

A lunch ticket can be purchased at the reception; it will be exchanged for a lunch box at the student restaurant in Hokkai Gakuen University (Coop Cafeteria, Restaurant Cosmos).

日本言語テスト学会第13回全国大会プログラム

9月7日(月)

8:30~

受付(7号館4階)

(PC利用発表者:発表教室で機器接続確認)

参加費:会員1,000円、学生1,000円、一般3,000円

9:00~9:15

開会行事(7号館3階 D30教室)

総合司会 塩川 春彦(研究大会実行委員長・北海学園大学)

挨拶 浪田 克之介(北海道大学名誉教授)

9:20~11:20

ワークショップ(7号館5階 D502教室)

テーマ: TDAPを使用するテストデータ分析

講師: 中村 洋一(清泉女学院短期大学)

9:20~11:30

研究発表1(発表30分, 質疑応答10分)

Part I 9:20~10:00

Part II 10:05~10:45

Part III 10:50~11:30

11:30~12:30

昼食

(役員会:7号館4階 D404教室 休憩室:7号館4階 D402教室)

12:30~13:55

研究発表2(発表30分, 質疑応答10分)

Part IV 12:30~13:10

Part V 13:15~13:55

14:00~15:10

講演(7号館3階 D30教室)

司会 塩川 春彦(研究大会実行委員長・北海学園大学)

紹介 浪田 克之介(北海道大学名誉教授)

演題: 日本の英語学力評価の歴史

講師: 田中 正道(広島大学名誉教授)

15:10~15:20

休憩

15:20~17:00

シンポジウム(7号館3階 D30教室)

テーマ: 言語能力と評価に関する研究:現在・過去・未来

コーディネーター兼パネリスト: 小林 美代子(神田外語大学)

パネリスト: 佐々木 みゆき(名古屋学院大学)

伊東 祐郎(東京外国語大学)

17:05~17:25

総会(7号館3階 D30教室)

議長選出

報告 中村 洋一(JLTA事務局長・清泉女学院短期大学)

17:25~17:35

閉会行事(7号館3階 D30教室)

18:00~19:30

懇親会(レストラン・コスモス)

司会 竹村 雅史(北星学園大学短期大学部)

研究発表・ワークショップ一覧(概略)

Part	Time	第1室 (D40)	第2室 (D41)	第3室 (D42)	第4室 (D50)	第5室 (D502)
1	9:20 ~ 10:00	上村	In'nami, Koizumi	Kikuchi	島田	ワークショップ 中村 9:20 ~ 11:20
2	10:05 ~ 10:45	秋山	Zhou	Nakamura, Mitsunaga	木村	
3	10:50 ~ 11:30	中原	Fujita	Jin	(空き)	
4	12:30 ~ 13:10	石川慎・石川有・ 水本・三橋	Sato	Song	楊	(空き)
5	13:15 ~ 13:55	法月・伊藤・ 島谷・木下	Thrasher	Kwon	森谷・ 小林	(空き)

研究発表一覧(詳細)

第1室(7号館4階 D40 教室)

司会 Part I 藤田 智子 (東海大学)
 Part II & III 滝沢 雄一 (福島大学)
 Part IV & V 島田 勝正 (桃山学院大学)

Part	発表者氏名(所属)	発表題目
I	上村 武司(株式会社ピアソン・エデュケーション)	VERSANT™ スピーキング&リスニング試験活用におけるコミュニケーションデザインの構築
II	秋山 實(東北大学大学院 院生)	オンラインテストの回答時間で習得度を測る試み - 英語の語彙・文法に関するテストの回答時間の場合 -
III	中原 敬広(株式会社eラーニングサービス)	「Moodle で簡単 e テスティング」を実現するテストサイトの開発
IV	石川 慎一郎(神戸大学) 石川 有香(名古屋工業大学) 水本 篤(流通科学大学) 三橋 峰夫(国際ビジネスコミュニケーション協会)	よく読める学習者はよく話せるのか: TOEIC テストおよび TOEIC SW テストを用いた日本人英語学習者の4技能能力値の関係
V	法月 健(静岡産業大学) 伊藤 彰浩(西南学院大学) 島谷 浩(熊本大学) 木下 正義(西南学院大学非常勤)	錯乱肢とテキストの分析による新旧 TOEIC の比較検証

第2室(7号館4階 D41 教室)

司会 Part I 法月 健 (静岡産業大学)
 Part II & III 齋藤 英敏 (茨城大学)
 Part IV & V 印南 洋 (豊橋技術科学大学)

Part	発表者氏名(所属)	発表題目
I	Yo In'nami (Toyohashi University of Technology) Rie Koizumi (Tokiwa University)	Establishing guidelines on how to select databases for meta-analysis in applied linguistics
II	Yujia Zhou (Tokyo University of Foreign Studies)	Comparing ratings of a face-to-face test and a telephone-mediated speaking test
III	Tomoko Fujita (Tokai University)	Does awareness of listening course objectives change students' self- can-do assessments?
IV	Rintaro Sato (Nara University of Education)	Reconsidering the effectiveness of TBL in comparison with PPP in the Japanese EFL Classroom
V	Randy Thrasher (Okinawa Christian University)	Challenges to Language Testing Posed by Task-based Language Teaching

第3室(7号館4階 D42教室)

司会 Part I 島谷 浩 (熊本大学)
 Part II & III 渡部 良典 (上智大学)
 Part IV & V 中村 優治 (慶応義塾大学)

Part	発表者氏名(所属)	発表題目
I	Keita Kikuchi (Tokai University)	Constructing and validating a questionnaire: Using Rasch PCA and Confirmatory Factor Analysis
II	Yuji Nakamura (Keio University), Haruhiko Mitsunaga (Tokyo Institute of Technology)	Developing an in-house placement test for measuring students' English proficiency
III	Kyung-Ae Jin (Korea Institute for Curriculum & Evaluation)	Developing Diagnostic Test for Teaching English in English
IV	Min-young Song (Korea Institute for Curriculum and Evaluation)	Developing the National Assessment of Educational Achievement (NAEA) in Korea
V	Oryang Kwon (Seoul National University)	Developing a Framework and Prototype of the Korean National English Proficiency Test for Secondary School

第4室(7号館5階 D50教室)

司会 Part I & II 竹村 雅史 (北星学園大学短期大学部)
 Part IV & V 前田 啓朗 (広島大学)

Part	発表者氏名(所属)	発表題目
I	島田 勝正(桃山学院大学)	明示的知識・暗示的知識を測定するテストバッテリーの可能性と問題点
II	木村 哲夫(新潟青陵大学)	言語テストにおける段階評価の実際：入試とプレイスメントテストのデータ処理
III		
IV	楊 元(筑波大学大学院人文社会科学 科学研究科国際日本研究専攻博士 後期課程)	日本語能力試験(聴解)の国内外の正答率の差の大きい問題から見えるもの 日本滞在経験の有無を中心に
V	森谷 浩士(神田外語大学) 小林 美代子(神田外語大学)	日本語能力評価を考える：質的視点より

Plenary Speech (講演)

日本の英語学力評価の歴史 田中 正道 (広島大学名誉教授) (D30 教室にて)

日本の英語学力評価の歴史を記述しようとする時、そもそも何を記述すれば歴史の記録として価値があるのか、出発点のところで意外と悩まされるものである。かなりの期間思いをめぐらしたが、英語教育関係者のこの分野・領域の主要関心事はおそらく以下のような観点ではなかろうかと考えた。

- 1) 「英語学力」概念の変遷
- 2) 測定・評価のツール(テスト、試験など)の発達、充実のプロセス
- 3) 入学試験に係る古今東西の論議の推移

これら以外にも記述に値する観点が考えられる。例えば、言語テストの開発・改善・普及に貢献のあった人物、古くからある各種の検定試験(「専検」、「高検」等)の調査・紹介などもある。今回、とりあえずの作業としては上記3つの観点到しぼることとする。

歴史記述をするにあたって、いま一つ決断しなくてはならないことがある。それはいつからを記述の起点とするかということである。日本人の英語学習の歴史はその気になればフェートン号事件(1808)まで遡れる。けれども江戸時代に英語に接することが出来た人は極めて限定されており、国民的レベルの英語学習ではなかった。学校教育が一応整備され、組織的に英語教育が開始されたのは明治時代である。(これとて経済的にゆとりのある一部のエリートに限られていた。)ともあれ、法制が整い、教育の体制もどうにか準備が出来た明治期以降を起点として記述することとする。

学力評価の歴史の記述で使用可能な学術資料は極めて少ない。それらのうち、上記3つの観点から epoch-making なものを厳選し、本講演で chronological に紹介していこうと考えている。21 世紀の日本の言語テストの研究・開発の推進役が期待される次世代の方々の参考になればと願っている。

Symposium (シンポジウム)

テーマ： 言語能力と評価に関する研究：現在・過去・未来

コーディネーター兼パネリスト

小林 美代子 (神田外語大学)

パネリスト

佐々木 みゆき (名古屋学院大学)

伊東 祐郎 (東京外国語大学)

(D30 教室にて)

ヨーロッパの言語評価研究の動向: 日本への示唆

小林 美代子 (神田外語大学)

現在、ヨーロッパでは、EU に代表されるように国境を越えて人々が自由に行き来することが可能となり、言語や文化の交流も活発に行われている。ヨーロッパ市民の一員として個々人が複数の言語を使えるようになることを目標に掲げ、Plurilingualism(複言語主義)という考え方も提唱されている。そのような中で、2001 年に出版された欧州外国語共通参照枠 (CEFR) が、言語政策、外国語教育・評価の枠組みとして活用され、言語評価実践にも大きな影響を与えている。CEFR の特徴である言語能力記述(Can-do statements)やポートフォリオが注目を集めているが、同時に CEFR の基本理念や社会的・歴史的背景に対する十分な理解がないまま評価指標として一人歩きする風潮があることには懸念が持たれている。

国連国際言語年の 2008 年には、The Social and Educational Impact of Language Assessment というテーマの下、ヨーロッパ言語評価者会議の国際大会 (ALTE2008) が英国ケンブリッジ大学において開催された。この大会テーマからもわかる通り、現在のヨーロッパでは、言語評価が社会や教育に与える影響、評価の使用目的や結果の解釈の持つ意味、評価の倫理や説明責任といった事柄が大きな関心事となっており、まさに Messick(1989)の提唱する結果的妥当性が議論の焦点となっていることが窺われる。大会の研究発表タイトルを概観してみると、当然のことながら大会テーマに沿った発表が多く、impact、washback、effect などといった語が並び、また学習、学習者、教師、指導、学校といった語も頻繁に現れている。本発表では、この大会の研究発表概要、特に基調講演の概要に触れながら、ヨーロッパの言語評価研究の一端を紹介し、日本の言語評価研究及び実践への示唆を探求する。

「第2言語ライティング力研究」の過去、現在、未来

佐々木 みゆき (名古屋学院大学)

本発表では、過去20年のL2 Writing Abilityの研究を「研究対象としてのWhat, How, Why」の観点から概観する。特に、「What->How->Why」と変遷してきた自分自身の研究の歴史を具体例として、「能力とは何か」、「能力を測るとはどういうことか」といった問題についてもヒントを探してみたい。Whatの研究には、学習者が書いたテキストの語や構造を習熟度別で比較したものや、作文の質に影響を与える要因を調査したものがある。どちらも「能力」や「評価」の問題に直接関わっている。

Howの研究の対象は、「学習者が書くプロセス」である。第一言語と第二言語の比較や、expertsとnovicesの比較などがある。このように作文完成までのプロセスを扱うHowの研究は、完成品のみを評価することの多い評価の研究とはあまり縁がないような気がするが、プロセス自体を評価するportfolio assessmentなどに影響を与えている。また、作文力の低い学習者に、作文力が高い学習者の方略である「事前に全体の枠組みを計画する」などを強制すると作文の質があがるのか、もしあがるなら、それは、その学習者の「作文力」と言えるのか、などの興味深い問題も提起する。

Whyの研究は、「良いものを書きたい、書けるようになりたい」と思わせる要因は何か、に関わっている。長期的に学習者の成長を追うケーススタディで扱われることが多く、学習者を取り巻く社会的文化的要因を考慮に入れたものがほとんどである。この種の研究も評価の分野とは縁がうすいように見えるが、同じ作文の課題に取り組んでいても、学習者が、「ただ課題をこなしている」と思っているか、「より良い作品を生み出したい」と思っているかで、作文の質は変わってくるのではないだろうか。その「質の違い」は「実力の違い」と言えるのか？などの問題は、評価論でも議論の対象になり得るのではないだろうか。

日本語教育における言語能力記述文 (Can-Do Statements) とテストニング

伊東 祐郎 (東京外国語大学)

日本語教師がテストを作成する際に、何をよりどころとして問題項目を作成しているのだろうか。日本語を教えられる教師であれば、経験と勘に基づいてテスト作成も難なくできるだろうと考えられている、と言っても過言ではない。

テスト結果から得られる得点は、測定誤差が最小限に抑えられていて、信頼に値するものであることが望ましい。それとともに、テストの得点の解釈や利用が妥当であるかどうかに関心が高まるべきだとして、妥当性の重要さも無視できない。

最近、外国語教育の分野では、テスト結果から得られる得点を具体的な能力の解釈として活用できるよう、得点に対する意義付けをこれまで以上に重要視するようになってきている。その結果、能力基準を設け、それに対する能力記述文が記されるようになってきた。このような背景の中で、日本の大学学部進学を目指す留学生に対する1年間の日本語コースでは、「言語能力記述文 (Can-Do Statements)」の構築に取り組んでいる。主な目的は、大学進学に必要な日本語運用力の最終目標を明確にし、習得すべき運用力とそれを下支えする言語スキルを捉え直すことにある。また、カリキュラムと評価のあり方に整合性をもたせることもねらいとしている。

「言語能力記述文」は大規模試験であるTOEFLやTOEIC、英検などにおいては既に紹介され、活用されている。本シンポジウムでは、日本語コースにおける「言語能力記述文」の構築の経緯を紹介しながら、今後の言語能力の評価と日本語テスト作成のあり方について考察してみたい。テストが、教師の一方的な情報入手手段であった時代から、受験者や第三者に対しても様々な影響力を及ぼす存在となってきた以上、テストに対する説明責任はおのずと出てくることになる。これからは日本語教師の評価リテラシーと責任ある対応が一層求められる時代になってくるだろう。

Paper Presentation (研究発表)

第1室 (D40) Part I

『VERSANT™ スピーキング&リスニング試験活用におけるコミュニケーションデザインの構築』
上村 武司 (VERSANT 事務局、株式会社ピアソン・エデュケーション)

VERSANT™は、英語のスピーキング及びリスニング能力を自動的に判定するテストです。米国のオーディネート社の特許技術を採用した高度な音声認識技術を利用して、個人の口頭英語表現力を迅速かつ客観的に電話で測定する唯一のテストです。VERSANT™はリスニング、スピーキングなどのコミュニケーションを重視しているため、受験者がどれだけ英語を知っているかよりも、どれだけ英語を実際に話せるかを判定するようになっています。また、スコアも即時にわかるため、日本を含め世界中の企業や教育機関などで採用されています。この発表では、現在の英語教育下におけるコミュニケーション活動とその評価方法、そしてコミュニケーション活動を引き出すモチベーションについて、VERSANT™を採用している企業・大学のデータを元に紹介いたします。

第1室 (D40) Part II

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

秋山 實 (東北大学大学院 院生 / 株式会社eラーニングサービス)

eラーニングシステムの普及に伴い、オンラインテストの利用も拡大してきている。しかし、テストの回答を評価するときに利用する情報は、従来の紙と筆記具を使ったテストと何ら変わっていない。オンラインテストでは、テスト項目への反応情報として、受験者が回答を選択するまでにかかった時間、回答を確定するまでにかかった時間、回答の書き直しがあった場合はその回答と時間、といったデータがあり、自動的に蓄積することが可能であるが、まだ標準的な機能として実現されていない。

本発表では、2006年12月に東北工業大学の1年生、2009年6月と7月に名古屋工業大学の1年生に実施した英語のプレースメントテストの回答時間データを用いて習得度を測る試みを紹介する。受験者は287名で、テストは4つのセクションから構成されている。そのうちの語彙・文法のテストセクションであるPart3のデータについて検討する。

回答時間の頻度分布を概観すると、項目困難度が高いテスト項目ほど分散が小さく、項目困難度が低いテスト項目では分散が大きくなる。別の見方をすれば、そのテスト項目で問われている内容を習得している受験者集団の回答時間の頻度分布の分散は小さく、習得のプロセスの途上にいる受験者集団の回答時間の頻度分布の分散は大きいということが言えそうである。

そこで、受験者を成績上位、中位、下位の3つの集団に分け、それぞれの回答時間の頻度分布を比較すると、成績上位、中位、下位の順で分散が大きくなり、平均値や中央値も大きくなることが分かった。このことから、成績上位群の回答時間分布を習得が進んだ場合の受験者の回答時間分布と考えて、これを基準に、たとえば、成績上位群の回答時間分布の下側2.5%値と上側97.5%値を回答時間の下限と上限とし、回答時間がこの間に収まり、かつ、正答した場合に、そのテスト項目の内容を習得していると考え、このようにして習得度を測る試みについて議論する。

第1室(D40) Part III

「Moodleで簡単eテストング」を実現するテストスイートの開発

中原 敬広(株式会社eラーニングサービス)

秋山・中原はこれまでに、教師の負担を軽減しながらeテストングを実現するために、1)大友・中村・秋山が開発したテスト分析プログラム TDAP2.0 for Windows を移植、2)アダプティブテストを実現する CAT モジュール、3)語彙レベルが制御されたアイテムを作成する E-Edit および J-Edit アイテムエディタ、4)多肢選択形式のアイテムを一括作成・登録する MultiChoiceMaker、5)Cloze 形式のアイテムを簡単に作成する Cloze エディタ、6)繰り返しテストを受験させて学生・教師に習得状況をフィードバックする「暗記カード」、7)回答時間を記録する RT 形式のアイテム、などをオープンソースのeラーニングソフトウェア Moodle のプラグイン、あるいは EXCEL のマクロとして開発し公開してきた。TDAP は、受験者数が少ない場合に分析できないという方式上の問題があった。複数の同一アイテム構成のテストを一括分析できるように改善し、さらに回答時間を考慮して分析する機能を付加した。さらに、多肢選択形式だけでなく複数選択形式、記述形式、回答欄が単一の Cloze 形式でも分析できるように改善した。CAT モジュールは、正誤だけでなく回答時間を考慮した判定ができるように改善した。アイテムの回答時間は RT 形式のアイテムを使用するか、TDAP でインポートすることによりアイテムバンクに蓄積できる。回答時間を考慮することにより、早すぎる回答や遅すぎる回答については、その問題が未習得であると判断する。受験結果の表示については、アイテム毎に必要な情報をトレース表示できるように改善した。

語彙などの自学自習用に開発した「暗記カード」モジュールは、1問ずつ回答し、結果を習得度別の色で示すことによって視覚的に受験者の習熟度を受験者自身も教員も容易に把握でき、学習指導に生かすことができる。

第1室(D40) Part IV

「よく読める学習者はよく話せるのか：TOEIC テストおよびTOEIC SW テストを用いた日本人英語学習者の4技能能力値の関係」

石川 慎一郎(神戸大学) 石川 有香(名古屋工業大学)

水本 篤(流通科学大学) 三橋 峰夫(国際ビジネスコミュニケーション協会)

1. 概要

本研究では、日本人英語学習者に見られる4技能の能力間の関係、および学習者属性の影響を検討する。本研究は100名以上の日本人英語学習者のデータに基づく大規模なもので、研究で得られた知見は、英語教育におけるカリキュラムの評価・開発、および外部テストの活用法の検討に有益な示唆を与えるものとなる。

2. RQ

リーディング力とリスニング力(喜田, 2007), 学習態度と英語力(川尻他, 2005), 習熟度とスコア(Childes, 2002), 語彙力と英語力(竹蓋, 2006; 石川, 2008)などの相関に関する先行研究をふまえて、本研究では(1)4技能の相関はどの程度か、(2)TOEIC スコアはどの程度発信能力スコアを予測するか、(3)語彙力と4技能の相関はどの程度か、(4)習熟度は4技能にいかなる影響を及ぼすか、の4点を検討する。

3. データと分析手法

被験者(日本人大学生・大学院生)はTOEIC テスト、TOEIC SW テストの2種類を一定期間内に連続受験し、あわせて、語彙力テスト・意欲態度アンケート・英語使用経験アンケートなどに回答した。データ分析には相関・回帰分析・クラスター分析などの統計手法が使用された。

4. 結果と考察

本研究のデータは現在分析中であるが、先行実施された30人規模(TOEIC スコアは $M=748.2$, $SD=78.9$)のパイロットスタディ(石川, 2009)からは、スピーキングを中心に考えた場合、相関係数はリスニング($r=0.64$) > リーディング($r=0.47$) > ライティング($r=0.22$)の順となること、語彙力はリーディング以外には十分な相関を持たないことなどが明らかになっている。より幅広い習熟度の学習者を対象とした本研究のデータ分析結果は、当日の発表において詳細に紹介する。

注: 本発表は国際ビジネスコミュニケーション協会と神戸大学石川研究室による共同委託研究成果の一部である。共同研究には、連名発表者のほか、小泉利恵氏(常磐大学)が加わっている。

第1室(D40) Part V

「錯乱肢とテキストの分析による新旧 TOEIC の比較検証」

法月 健(静岡産業大学) 伊藤 彰浩(西南学院大学)
島谷 浩(熊本大学) 木下 正義(西南学院大学非常勤)

2006年5月に Test of English for International Communication (以下、TOEIC) は、より現実的なコミュニケーション能力の測定と評価を標榜し、大きく改訂された。新 TOEIC(R)の問題様式に見られる特徴は、1) 一部の問題におけるテキストの長文化及び一問題あたりの問題数の増加、2) 聴解セクションの音声に含まれる英語アクセントの多様化、3) 読解セクションにおいて誤文訂正問題が廃止され、代わりに文章の空所補充問題が追加されたことの3点に集約することができる。このような変更により、改定前と改定後の TOEIC のテスト及びテキスト特性、それに対応する受験者の応答にどのような影響が及んだのだろうか。テスト全体や聴解・読解セクション別の特徴、個別項目の特性、受験者の応答様式等を比較分析するため、日本の大学に学ぶ136人の学生に1~2週間の間隔において、新旧2つの様式の TOEIC テストを実施した。本発表では、主として、(1) 分析に使用した新旧 TOEIC 全体の平均点に有意な差は見られるか、(2) 新旧 TOEIC 聴解、読解セクションのそれぞれの平均点間に有意な差は見られるか、(3) 200問の項目難易度と136人の受験者能力は、同一の間隔尺度上にどのように位置づけられるか、(4) 項目難易度、項目弁別力、項目適合度、選択肢適切度の観点から、十分に機能していない項目は存在しないか、(5) 受験者適合度や受験者能力値の変動性の観点から、テストに順応していない可能性のある受験者はいないか、(6) 項目難易度とテキストのリーダビリティや使用語彙の特性には何らかの関係は見られるか、(7) 上記の項目及びテキスト特性と受験者の応答様式に何らかの質的な関係を見出すことは可能かについて議論し、分析結果から、教育的示唆と今後の研究の方向性を考察する。

第2室 (D41) Part I

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

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

We examined the combination of databases that has been used in meta-analysis and that provides a comprehensive coverage of applied linguistics journals in order to provide database selection guidelines for conducting a meta-analysis in applied linguistics. A literature search using all or as many databases as possible would be ideal for a comprehensive coverage of previous studies. However, there are cases where database selection is necessary. The results for Research Questions 1 and 2 (Which databases are used in meta-analysis in applied linguistics? and What combination of databases is used in applied linguistics?) indicated that ERIC, LLBA, ProQuest Dissertations and Theses, and PsycINFO and combinations of these four databases were frequently used in previous meta-analyses and are recommended for meta-analysts. The result for Research Question 3 (Which databases provides a comprehensive coverage of journals in applied linguistics?) revealed that LLBA, ERIC, MLA, Linguistics Abstracts, and Scopus covered journals well and are recommended for meta-analysts. LLBA includes all the 24 journals surveyed in this paper, and these 24 journals can also be covered using a combination of the other four databases. However, due to differences in journal coverage periods and potentially missing volumes and articles, using only LLBA or any single database is quite likely to lead to overlooking relevant articles. Based on our findings, we recommend using as many databases as possible or selecting at least two from the seven databases above according to one's research purposes and database characteristics such as journal coverage (periods). In particular, we argue that ERIC and LLBA are essential for meta-analysis in applied linguistics. Furthermore, since specific issues of a journal are often not included, use of a search engine available on the journal homepage or a CD-ROM version of the journal is recommended. Moreover, even if journal coverage (periods) overlaps between databases, using multiple databases is strongly advised as this would preclude any failure to identify a study due to a meta-analyst's search fatigue or the different frequencies at which databases are updated.

第2室 (D41) Part II

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

Yujia ZHOU (Tokyo University of Foreign Studies)

Recently, in order to increase administration efficiency, speaking tests using technology have been administered as substitutes for face-to-face tests. In Japan, a telephone-mediated speaking test, Telephone Standard Spoken Test (T-SST), has been developed as an alternative for Standard Spoken Test (SST). SST was developed based on Oral proficiency Interview (OPI), aiming to discriminate spoken English of Japanese learners. In order to explore the appropriateness of test substitution, this study addressed the comparability of the two tests based on the data of 83 examinees from two perspectives: (1) scoring quality and (2) examinee performance. Scoring quality was examined in terms of level of rater agreement and generalizability of the ratings using generalizability theory. To compare examinee performance elicited by each test, absolute agreement of ratings, Spearman rank order correlations and Pearson product-moment correlations were calculated and t test on the mean scores was also performed. The results of the analyses will be reported in the presentation.

第2室 (D41) Part III

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

Tomoko Fujita (Tokai University)

The purpose of the study is to examine how the students' awareness to the course objective changes their responses to can-do self-checklists for a new listening course at a university language programme. About three thousand Japanese college students had been divided into three different levels, advanced, intermediate, and basic, according to the results of an in-house placement test. The teachers in the program created six different types of original can-do statements (CDS) for the listening course based on the Common European Frame of Reference (CEFR): 1. speed and content, 2. announcements, 3. audio materials, 4. conversations, 5. visual materials, and 6. vocabulary and complexity of sentences. Two CDS (3. audio materials and 4. conversations) were chosen as objectives for four 90 minutes special lessons. Students' responses to these CDS before and after the four special lessons were compared. However, each CDS has different descriptions for each of the three levels. The description for the intermediate level of the audio materials (CDS3) is "I can understand the main points of short radio news and simpler recorded material about familiar subjects" and the description for the same level of the conversations (CDS4) is "I can catch the main points of a conversation on familiar topics." Therefore, 18 CDS on the can-do self-checklists were answered by 230 randomly-selected students from the three course levels. Four-point scales were used for the self-checklist. After the lessons, students again completed the same checklists.

Data analysis involved comparing differences in student responses before and after the lessons. Students' average post- checklist total scores increased ($M = 1.36$) from the pre-checklists. The increase was the most for CDS4: conversations ($M = 0.16$) and the average increase for CDS3:audio materials was 0.07. Secondly, the results were broken down by student proficiency levels. Generally, advanced students showed higher gains than students in the other two levels. The average difference between pre and post total scores for advanced students was 2.79, while the other two levels showed gains of 0.80 and 0.63. Concerning CDS3 and 4, advanced students showed the largest increases (0.278) for CDS4 and 0.235 for CDS3. However, the increases by students in the other two levels for CDS3 were not quite as large as those for CDS4. The effect of the four special lessons seems to be clear for conversations, but less so for audio materials.

第2室 (D41) Part IV

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

SATO Rintaro (Nara University of Education)

Since one of the overall objectives in English Education both in junior and senior high school is to develop students' practical communication abilities, the utilization of task-based language learning (TBL), which is a logical development of communicative language teaching (Willis, 1996), seems to be attractive. In fact, the use of TBL in English teaching in Japan has been gaining great attention. The belief that practice in production has a crucial role for developing L2 proficiency is reflected in English lesson plans based on common and traditional teaching methodology using the Presentation-Practice-Production (PPP) model. This long-established traditional teaching methodology based on PPP is now being replaced by TBL in SLA (e.g., Long and Crookes, 1991; Skehan, 1996, 1998; White, 1988; Willis, 1996, 2004). However, in the Japanese EFL learning environment where there is no actual need to use English for communication outside the classroom, we might be skeptical of the effectiveness of TBL. To examine whether learners would use target grammatical structures or not, I conducted a brief survey with twenty-one university students majoring in English Education, and found that it is unlikely that learners learn target items in a task effectively. In the presentation I will discuss: the suitability and problems of TBL in the Japanese secondary school context; the effectiveness of PPP from the point of view of skill acquisition theories; possible problems about PPP. In addition, suggestions for effective teaching procedures will also be discussed. I believe that investigation into the suitability of TBL and PPP in the Japanese EFL classroom is crucial, as we have to choose methodologies that can effectively develop learners' English proficiency. Only then will we be able to improve both the theoretical understanding and practice of measurement and evaluation in English education in Japan.

References

- Long, M. & G Crookes (1991). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26, 27-55.
- Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp.17-30). Heinemann: Oxford University Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Heinemann: Oxford University Press.
- White, R. (1988). *The ELT Curriculum*. Oxford: Blackwell.
- Willis, J. (1996). *A Framework for Task-based Learning*. Harlow, UK: Longman.
- Willis, J. (2004). Perspectives on task-based instruction: Understanding our practices, acknowledging our different practitioners. In B. L. Leaver & J. Willis (Eds.), *Task-based instruction in foreign language education* (pp. 3-44). Washington, D.C.: Georgetown University Press.

第2室 (D41) Part V

Challenges to Language Testing Posed by Task-based Language Teaching

Randy Thrasher (Okinawa Christian University)

In the last 20 years or so Task-Based Language Teaching (TBLT) has become an increasing popular classroom practice. This paper introduces TBLT and, building on the pioneering work of Long & Norris, Brindley, Norris et al, and Byrnes; discusses the issues it raises for language testing. It also examines ways in which such classroom practice might be appropriately measured. The paper argues that, although task-based teaching fits well with Bachman and Palmer's demand for task tasks that match real-world language activities, this teaching approach requires some departure from traditional testing practice if what happens in a TBLT classroom is to be appropriately mirrored in test results. It also raises the question of whether it is possible to adequately measure the process approach advocated by some TBLT practitioners. The paper closes with a discussion of how language testing might help resolve the problem of the lack of accuracy of many students in 'focus on meaning' classrooms.

References

- Bachman, Lyle and A Palmer Language Testing in Practice Oxford University Press
- Brindley, Geoff Task-centered language assessment in language learning: The promise and the challenge in Van den Branden et al.
- Byrnes, Heidi The role of task and task-based assessment in a content-oriented collegiate FL curriculum in Van den Branden et al.
- Long Michael & John Norris Task-based teaching and assessment in Van den Branden et al.
- Norris, John M. et al Examinee abilities and task difficulty in task-based L2 performance assessment in Van den Branden et al.
- Van den Branden, Kris, Martin Bygate & John M. Norris (editors) Task-Based Language Teaching: A Reader John Benjamins

第3室 (D42) Part I

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

Keita Kikuchi (Tokai University)

In this presentation, the presenter reports the result of the both Rasch PCA and confirmatory factor analysis (CFA) of the survey on demotivation among Japanese high school learners of English. Dörnyei (2001) defined the demotivation as “specific external forces that reduce or diminish the motivational basis of a behavioral intention or an ongoing action” (p. 143) in the context of second language learning. In contrast to amotivation, which refers to a lack of motivation and is most associated with self-determination theory (Ryan & Deci, 2002), demotivation is considered to be related to specific external forces, “demotivators,” which causes the reduction of motivation.

Following Dörnyei’s definition of demotivation, this paper focused on specific external forces, so-called, demotivators, that Japanese high-school students reported which can cause their motivation to be reduced or diminished. On the basis of the Kikuchi’s (in-press) qualitative study and several other studies of demotivation (e.g., Kikuchi and Sakai, in press; Sakai and Kikuchi, 2009), a 40-item questionnaire was developed to gather survey data from over 1,200 high school students. Using the conventional factor analysis and Rasch factor analysis of PCA residual (Linacre, 2004), six demotivators (teachers, characteristics of classes, experiences of failure, class environment, class materials, and learners’ interest) originally hypothesized were checked and revised five constructs were confirmed. The process of constructing measures on demotivators guided by Bond and Fox (2007) and Linacre (2004) is presented. In addition, the strengths of using Rasch PCA residual to identify common factors in the survey data are discussed.

Key References

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.
- Kikuchi, K. (in-press). Listening to our Learners’ Voices: What Demotivates EFL High School Students? *Language Teaching Research*, 13.
- Kikuchi, K., & Sakai, H. (in press). Japanese learners’ demotivation to study English: A survey study. *JALT Journal*.
- Linacre, J. M. (2004). *A users guide to Winsteps, Version 3.5*.

第3室 (D42) Part II

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

Yuji Nakamura (Keio University), Haruhiko Mitsunaga (Tokyo Institute of Technology)

Placement tests are one of the most widely used types of tests within educational institutions. Although some schools still choose commercially-produced proficiency tests, a school's placement test should be relevant to its curriculum, and therefore, should ideally be developed by the test administrators themselves. This paper argues in favor of in-house placement tests in order to meet the needs of the institutions and discusses the possibility of item banking. Keio University (the faculty of letters) has administered our in-house placement test to incoming freshman students and new sophomore students since the spring of 2006 (twice a year). This placement test measures those students' reading ability in English and overall proficiency to provide streamed instruction appropriate to their proficiency levels to optimize their learning experience, and to provide multi-faceted English communicative skills. The test consists of four sub-sections made up of grammar, vocabulary, cloze and reading comprehension. It contains a total of 50 multiple-choice items and is administered to about 800 university students at a time. The purpose of the present study is three-fold: to examine the validity and reliability of the test using the Rasch-based statistical program, to report the consecutive three year investigation of students' reading ability, and to suggest the implementation of an item bank which will eventually have 400 linked items by means of equation. The validity can be examined whether the results fit the model or not in the Rasch measurement analysis. Also, the content validity and the face validity are discussed by using the interview survey and the questionnaire analysis. The reliability is investigated by the cronbach alpha plus the K-R 21 index. The benchmark for the acceptable boundary is over 0.7. The consecutive three-year investigation is examined through the comparison of students' test results (descriptive statistics and latent traits). The 50 items in each test are linked by 15 anchor items for the whole item bank. Once all the items are calibrated and the difficulty of each item is determined, each item can be put on the continuum of the scale according to their logit scores (difficulty level). These items along with a task can be stored as items in a bank. In the presentation we suggest that another model such as 2PL(2-parameter logistic model) is more suitable for illustrating item characteristics when analyzing linguistic test and also, show the result of applying 2PL.

第3室(D42) Part III

Developing Diagnostic Test for Teaching English in English

Kyung-Ae Jin (Korea Institute for Curriculum & Evaluation)

The purpose of this study is to provide primary and secondary teachers with criteria on TEE (Teaching English in English) and to develop their TEE abilities. This study developed ETKT (English Teaching Knowledge Test) and ETPT (English Teaching Practice Test) as TEE Testing tools to diagnose English teachers' knowledge of teaching methods, their communicative abilities in the class, and their English language proficiencies. ETKT is the first module which assesses English teachers' knowledge of teaching methods and major areas in teaching English language. Specifically, it consists T/F items asking English teachers' fundamental understanding of English language education such as second language acquisition, teaching methodology, lesson planning, classroom management, task-based activities, teachers' English language proficiency, development of teaching materials, and language testing as well as teachers' knowledge of English language such as phonetics, phonology, grammar/syntax, and pragmatics/socio-linguistics. ETPT is the second module that evaluates English teachers' performance in their classrooms. This instrument basically assesses teachers' teaching skills which encourage students to participate in their class and improve their spontaneous and creative talks in English. Researchers have enhanced the reliability and validity of these diagnostic instruments through item reviews and field tests. By piloting ETPT at 3 elementary schools and 3 middle schools, researchers verified items' reliability, validity, and rubric. The researchers also administrated ETKT with 127 elementary school teachers and 120 secondary school teachers, 247 in total. This pilot proved the items' quality such as item difficulty and discrimination as well as reliability and validity of this instrument. With the finalized diagnostic instruments, researchers conducted standard setting to decide cut points of examinees' levels. Specifically, this standard setting was a process to decide cut-scores that can categorize teachers' TEE abilities with TEE_Ace (Excellent) and TEE_Master (Extremely Excellent). Researchers used a modified Angoff method which rounded up and analyzed specialists' opinions about individual item's difficulties and levels to decide cut points.

第3室(D42) Part IV

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

Min-young Song (Korea Institute for Curriculum and Evaluation)

Since 1998, NAEA in Korea has been administered to small portion of elementary and secondary students to diagnose their educational achievements and provide fundamental reference data for improving the National Curriculum, and the result has not been released to the public. As accountability of educators, schools, and school districts has been raised as an important social issue, however, it has been decided that from 2009 on, all the 6th, 9th, and 10th graders are supposed to take the test and information about school performance on the test should be provided for the public. This has brought many significant changes in the procedures of developing and administering NAEA, and some of the major changes, especially focusing on the English subject, will be discussed in this paper.

第3室(D42) Part

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

Oryang Kwon, Ph. D (Professor, Seoul National University)

The present project was motivated by two factors: (1) the necessity for accurately assessing the students' English abilities and (2) the necessity for qualitative improvement in students' English proficiency in order to

meet the challenges of the rapidly changing world where international competitions are becoming tougher than ever before. In Korea, there have been many commercial tests of English proficiency that claimed to measure students' English abilities in several skill areas. These tests, however, varied greatly in their measurement scales and areas, making them inadequate for the students who have learned English as specified by the national curriculums. And yet, these tests have been taken by many Korean secondary and even elementary school students because some prestigious special purpose high schools required the scores of these tests for their admission decision making. Therefore, the necessity for developing instruments to accurately measure and assess the students' English abilities in the specific context of Korea has been acknowledged by many people both inside and outside the educational circles.

The second motivational factor is related with the reality that the students' practical English abilities are far below the desired levels, especially in the areas of speaking and writing, as a result of negative washback of the College Scholastic Ability Test (CSAT), a high-stakes test that does not have a speaking or writing section. A high-stakes test that includes speaking and writing has been suggested by testing specialists as a solution to induce a qualitative change of English classroom instruction.

From these two necessities, the present project was carried out for the following purposes:

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

To this end, the present project started with a discussion about the nature of the test and concluded that the test should be an achievement test, rather than a proficiency test, based on the school curriculum. Thus, the achievement criteria specified by the national curriculum became the main reference criteria for the test.

Next, the levels of the test were discussed. One quick solution would be to develop a test for each grade level of primary and secondary schools. However, this would be difficult in terms of the test practicality, which is an important component of a good test. Another possibility would be to develop one test for each school level, that is, one for elementary school (3rd to 6th grades), one for middle school (7th to 9th grades), and one for high school (10th to 12th grade). However, this solution, although practically simple, would be inadequate to accurately measure the students' achievements. A compromise of the above two was to tie up two grades for one test, thus developing five tests for the ten grades (3rd to 12th grades). Accordingly, two teams were formed: one team to develop elementary school tests, and one to develop three secondary school tests.

As for the skill areas to be assessed, the initial plan was to develop only the productive skill (i. e., speaking and writing) tests. However, after considering that most of the current tests assess all four skills, the present project also decided to include all four skills of listening, speaking, reading, and writing. Therefore, the present project was to develop three level tests for secondary schools.

The test development procedures started with a review of literature to set the theoretical background, and a review of currently available foreign and domestic English tests to draw up test specifications. After the reviews, operational definitions of the achievement criteria specified in the curriculum were made. Test time, item numbers, and item types were specified. The materials and types of linguistic input (i. e., spoken or written) for listening and speaking tests were also determined. The criteria for rating speaking and writing samples were prepared to rate the test performances.

Based on the test specifications, actual test items were written by professors, school teachers, and English native-speakers. The first draft was reviewed and revised by specialists, and the second draft also went through intensive reviews. The final draft contained test items as shown in Table 1.

TABLE 1
Numbers of Items for Each Part of the National English Proficiency Test Prototype

Levels	Listening	Reading	Speaking	Writing	Total
Secondary Level 1	30	30	5	3	68
Secondary Level 2	35	35	5	2	77
Secondary Level 3	40	40	5	2	87

The items were then pilot tested in sampled schools. The pilot tests, which included two elementary level tests, were administered to 970 students in 9 schools in the Seoul metropolitan area. An item analysis was conducted on listening and reading tests. Although some of the items showed weak discrimination power, the overall test items demonstrated that they served the test purposes well.

第4室 (D50) Part I

明示的知識・暗示的知識を測定するテストバッテリーの可能性と問題点

島田 勝正 (桃山学院大学)

本研究は、明示的知識(Explicit knowledge)と暗示的知識(Implicit knowledge)の区別が、時間制約によるものなのか、それともテスト項目の文法的適格性によるものなのかを調べた島田(2007,2008)の継続研究である。島田(2008)では、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)を配し、20文法範疇、80項目から構成される文法性判断テスト(Grammaticality Judgment Test; GJT)と、同一の80項目を対象とする誤文訂正テスト(Error Correction Test; ECT)を開発した。GJTには、時間制約の有無により Timed; TM と Untimed; UT の2種類がある。さらに、GR20項目とUG20項目から構成される口頭模倣テスト(Oral Imitation Test; OIT)を開発した。本研究では、さらに40項目の口頭英訳テスト(Oral Translation Test; OTT)を追加し、5つのテストによる8つの変数間の相関関係を分析した。探索的因子分析の結果は、時間制約の有無(TM:UT)による2因子モデルよりも口頭産出、文法的適格性判断、文法的不適格性判断の3因子モデルの方が適合度が高いことを示唆している。

第4室 (D50) Part II

言語テストにおける段階評価の実際：入試とプレイスメントテストのデータ処理

木村 哲夫 (新潟青陵大学)

従来、言語テストのデータは連続する間隔尺度として処理され、それに基づいて受験者の能力が評価されてきた。この点に関しては、古典的テスト理論においても、項目応答理論においても同様である。しかし、順序尺度を想定したテスト理論として発表されたニューラルテスト理論 (Shojima, 2008) は、テストのデータを非連続の順序尺度として処理し、評価において段階評価を行うものである。方法論的にも、教育社会学的にも、言語テストに段階評価を導入することの意義は大きい (荘島, 2007)。テストは選抜やクラス分けに用いるだけでなく、どの程度どのような能力を持っているかについて診断する機能を持っていることが望ましい。段階評価を導入することにより、段階評価により区別される各能力段階 (潜在ランク) の特徴を、Can-Do Chart との関連で示すことが、連続尺度のもとで検討するよりも容易に行えるであろう。このことは、言語テストの品質管理やアカウンタビリティの確保という観点からも、有効かつ重要な問題である。

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

第4室(D50) Part IV

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

楊 元 (筑波大学大学院人文社会科学研究科国際日本研究専攻博士後期課程)

(a) 研究目的

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

(b) 研究方法

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

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

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

(c) 結果とその考察

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

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

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

第4室(D50) Part V

日本語能力評価を考える：質的視点より

森谷 浩士(神田外語大学) 小林 美代子(神田外語大学)

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

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

Workshop (ワークショップ)

第5室 (D502)

TDAP を使用するテストデータ分析・ワークショップ

講 師 中村 洋一 (清泉女学院短期大学)

本ワークショップでは、大友・中村・秋山による、Test data Analysis Program: TDAP Ver. 2.02 を使用し、主に古典的テスト理論に沿ったテストデータ分析の実習をします。ワークショップ当日、使用する TDAP の CD-ROM を配布します。

古典的テスト理論は、従来から用いられている方法で、いくつかのテスト項目からなるテストを作成し、正答した項目の数を合計した得点、あるいは、正答数にそれぞれの配点を掛けて合計した得点を用いてテストデータ分析を行います。ワークショップでは、この「素点」、あるいは「正答数に基づく得点」と呼ばれる数値を基にして、基礎統計量・標準得点・テストの信頼性の検討、及びテスト項目の特性を検討する項目分析の実習を行います。特に、それぞれの分析結果の解釈に焦点を当てて検討し、よりよいテストを作成するために留意すべき点を考察することをねらいとします。

実習の流れは、次の通りです。

1. 解答データの作成・素点データの作成
2. 基礎統計量の算出と検討: 平均値・最頻値・中央値・標準偏差・分散・歪度・尖度
3. 標準得点の算出と検討: Z 得点・T 得点・5 段階尺度・9 段階尺度
4. 項目分析: 項目困難度・項目弁別力指数・実質選択肢数
5. 今後の課題

- * 本ワークショップでは、CD ドライブ付きのノート PC をご持参いただけると、より効果的な学びができます。
- * 申し込み: 当日参加でもかまいませんが、資料や会場準備の都合がありますので、できれば8月30日(日)までに申し込んでください。申し込みの方法については、別紙をご参照ください。

会場へのアクセス



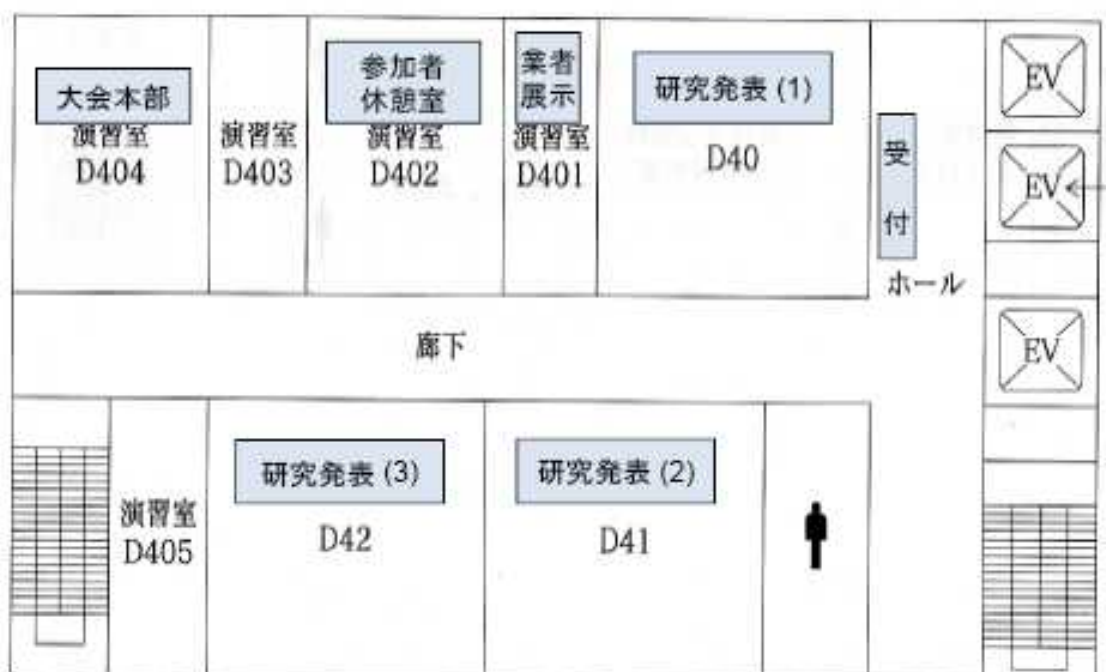
地下鉄東豊線「学園前」駅にて下車。3番出口直結。
 (「大通」駅より乗車5分、「さっぽろ」駅より乗車6分)

会場配置図

7号館 3階



7号館 4階



7号館 5階



展示協賛企業

株式会社 e ラーニングサービス

株式会社ピアソン・エデュケーション

国際教育交換協議会 (CIEE) TOEFL 事業部

Are you e-Testing?

あなたはeラーニングを使いこなしていますか？

沖縄で e ラーニング無料ワークショップを開催しています

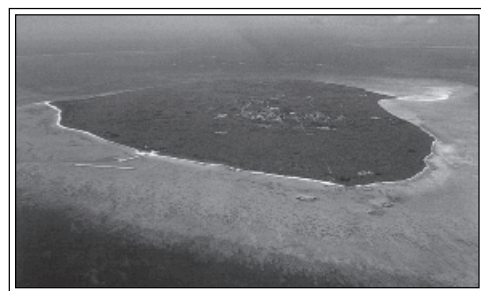


英語プレースメントテスト

PLATON-m

Moodle上で受験可能な実績のある英語
のオンラインプレースメントテストです。

1 回 840 円 / 人



e ラーニングのことなら。

上の写真は20年以上前の竹富島の
水牛車の写真です。竹富島では今で
も水牛車による島内観光が行われて
います。我々eラーニングサービスは
沖縄でeラーニングシステムの構築・
カスタマイズやコンサルティングを
主にしている会社です。特に最近では
eラーニングの中でもテスト機能に
注目しています。e テスティングとよ

ばれ、eラーニングで最も活用できる
機能です。弊社ではPROX法によるア
イテム分析やアダプティブテストの構
築などのツールを無償で提供したり、
構築のお手伝いを行っています。新し
い問題タイプの作成などもお引き受
けいたします。また沖縄ではeラーニ
ングの無料ワークショップを行ってい
ます。是非一度お越しください。

eLS Moodle Plugin配布サイト <https://e-learning.ac/moodle-resources/>

アダプティブテスト を作ってみませ んか？

ツールの提供とアイテムバンク構築
古典理論とPROX法に基づいたテストア
イテム分析機能とその分析結果を利用
可能なアダプティブテスト機能は無償で
提供いたします。構築全般のコンサルテ
ィング・実作業も承ります。

レンタル e ラーニ ングサーバを提供

用途に応じたプランをご用意

最大同時アクセス40名の小規模な e ラ
ーニングに最適なサービスです。必要に
応じて管理委託やヘルプデスクを準備
しております。

年間126,000円～

Contact us

株式会社 e ラーニングサービス

Web: <http://www.e-learning-service.co.jp>

Mail: askme@e-learning-service.co.jp

Tel : 098-886-2293