

日 本 言 語 テ ス ト 学 会 (JLTA)

第12回 全 国 研 究 大 会 (2008 年 度) プ ロ グ ラ ム

Handbook
of
the 12th Annual Conference
of
The Japan Language Testing Association

大会テーマ : Standards in Language Testing

日時 : 2008 年 9 月 14 日 (日) 8:30~17:35

会場 : 常磐大学 R棟
(〒310-8585 茨城県水戸市見和1-430-1, TEL(代表) 029-232-2511)

主催 : 日本言語テスト学会 (JLTA)

後援 : 茨城県教育委員会

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	(常磐大学)
小泉 利恵	(常磐大学)
伊藤 彰浩	(西南学院大学)
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

小泉 利恵	(常磐大学)
伊藤 彰浩	(西南学院大学)
塩川 春彦	(北海学園大学)

研 究 発 表 審 査 委 員

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

The 12th JLTA Annual Conference

Place Tokiwa University

Date September 14, 2008 (Sunday)

Timetable

8:30～	Reception (R Tower Lobby)	
9:00～9:15	Opening Ceremony (Room R002)	
9:20～10:00	Presentation Part I	Workshop (Room R103) 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 R103 Lounge : Room R002)	
12:30～13:10	Presentation Part IV	
13:15～13:55	Presentation Part V	
14:00～15:10	Plenary Speech “Defining, Setting and Validating Standards in Language Testing” Tony Green (University of Bedfordshire) (Room R002)	
15:20～17:00	Symposium Theme: “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) (Room R002)	
17:05～17:25	General Business Meeting (Room R002)	
17:25～17:35	Closing Ceremony (Room R002)	
18:00～19:30	Banquet (T Tower Student Hall)	

Exhibition R Tower Lobby

Headquarter & Lounge for Participants Room R002

大会日程表

8:30～	受付 (R 棟ロビー)	
9:00～9:15	開会行事 (R002 教室)	
9:20～10:00	研究発表 Part I	ワークショップ (R103 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：R103 教室 休憩室：R002 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 (R002 教室)	
15:20～17:00	シンポジウム (R002 教室)	
17:05～17:25	総会 (R002 教室)	
17:25～17:35	閉会行事 (R002 教室)	
18:00～19:30	懇親会 (T 棟学生ホール)	

業者展示 R 棟ロビー

大会本部・参加者休憩室 R002 教室

日本言語テスト学会第12回全国大会プログラム

9月13日(土)

16:30～18:00 **理事会** (常磐大学 R棟 1階 R103教室)

9月14日(日)

8:30～ **受付** (R棟ロビー)
(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** (R棟 地下1階 R002教室)
総合司会 小泉 利恵 (研究大会実行委員長・常磐大学)
挨拶 高木 勇夫 (常磐大学学長)
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** (R棟 1階 R103教室) 司会 中西 貴行 (常磐大学)
テーマ 「SPSS を使って分散分析をやってみよう」
講師 平井 明代 (筑波大学)・長橋 雅俊 (筑波大学大学院)

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：R棟 1階 R103教室 休憩室：R棟 地下1階 R002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** (R棟 地下1階 R002教室)
司会 藤田 智子 (東海大学)
紹介 浪田 克之介 (北海道大学名誉教授)
演題： Defining, Setting and Validating Standards in Language Testing
講師： Tony Green (University of Bedfordshire)

15:10～15:20 **休憩**

15:20～17:00 **シンポジウム** (R棟 地下1階 R002教室)
テーマ： Setting and Validating Standards on Language Tests
コーディネーター Tomoko Fujita (Tokai University)
パネリスト Michiko Nakano (Waseda University)
討論者 Tony Green (University of Bedfordshire)
Tomoko Fujita (Tokai University)

- 17:05～17:25 総 会 (R 棟 地下1階 R002 教室)
議長選出
報告 中村 洋一 (JLTA 事務局長・常磐大学)
- 17:25～17:35 閉会行事 (R 棟 地下1階 R002 教室)
- 18:00～19:30 懇親会 (T 棟学生ホール)
司会 印南 洋 (豊橋技術科学大学)・飯村 英樹 (常磐大学)

研究発表・ワークショップ一覧(概略)

Part	Time	第1室 (R001)	第2室 (R004)	第3室 (R003)	第4室 (R102)	第5室 (R101)	第6室 (R104)	第7室 (R103)
1	9:20 ～ 10:00	山下	村上	Newfields	Masaki, Otoshi, & Kuru	酒井	高波	Workshop 9:20～11:20
2	10:05 ～ 10:45	柳瀬	Fujita	Saito	宮崎	柳川	秋山(實)	
3	10:50 ～ 11:30	In'nami & Koizumi	跡部・ 境・ 野上・ 片岡	Zhou	奥田・ 大津	森谷	劉	
4	12:30 ～ 13:10	斉田	伊東・ 川口・ 太田	米野	深澤	Shite	中原	(空き)
5	13:15 ～ 13:55	小林・ 長谷川・ 町田	Thrasher	曲	藤森	島田	木村	(空き)

TOKIWA 



ANNIVERSARY

学校法人常磐大学は2009年に
開学100周年を迎えます

研究発表一覧(詳細)

第1室 (R棟 地下1階 R001 教室)

司会 Part I 小林 美代子 (神田外国語大学)
 Part II & III 渡部 良典 (上智大学)
 Part IV & V 片桐 一彦 (専修大学)

Part	発表者氏名 (所属)	発表題目
I	山下 仁司 ((株) ベネッセコーポレーション)	「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)
II	柳瀬 陽介 (広島大学)	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解明—
III	Yo In'nami (Toyoashi University of Technology) Rie Koizumi (Tokiwa University)	Language ability structure and Chapelle's (1999) construct definition and interpretation: Findings from the reanalysis of multitrait-multimethod studies
IV	斉田 智里 (茨城大学)	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—
V	小林 美代子 (神田外国語大学) 長谷川 信子 (神田外国語大学) 町田 なほみ (神田外国語大学)	子どもの英語力を測る: 語彙テスト開発の試み

第2室 (R棟 地下1階 R004 教室)

司会 Part I 斉田 智里 (茨城大学)
 Part II & III 小山 由紀江 (名古屋工業大学)
 Part IV & V 滝沢 雄一 (福島大学)

Part	発表者氏名 (所属)	発表題目
I	村上 京子 (名古屋大学)	就労外国人を対象とする日本語能力の測定
II	Tomoko Fujita (Tokai University)	Creating valid in-house can-do descriptors for a listening course
III	跡部 智 (慶應義塾大学) 境 一三 (慶應義塾大学) 野上 康子 ((株) 教育測定研究所) 片岡 夏子 ((株) 教育測定研究所)	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
IV	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)	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に影響を与えるのか?—初・中・上級者のタスク経験と自己評価値との関連の検証—
V	Randy Thrasher (Okinawa Christian University)	An attempt to improve our understanding of reading test difficulty

第3室 (R棟 地下1階 R003 教室)

司会 Part I 静 哲人 (関西大学)
 Part II & III J. K. Hubbell (法政大学)
 Part IV & V 秋山 朝康 (文教大学)

Part	発表者氏名 (所属)	発表題目
I	Tim Newfields (Toyo University)	The construct validity of "pronunciation" exercises in entrance examinations: Student perspectives
II	Hidetoshi Saito (Ibaraki University)	Comparison of Keyword-assisted, Practiced, and Impromptu Performances on an Oral Summary Classroom Test: A Pilot Study
III	Yujia Zhou (Graduate School of Tokyo University of Foreign Studies)	A comparison of speech samples of monologic tasks in speaking tests between computer-delivered and face-to-face modes
IV	米野 和徳 (山形県教育庁高校教育課)	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
V	曲 明 (東京外国語大学大学院)	グループ・オーラルテストにおける評価改善の研究

第4室 (R棟 1階 R102 教室)

司会 Part I Randy Thrasher (Okinawa Christian University)
 Part II & III 島谷 浩 (熊本大学)
 Part IV & V 中村 優治 (慶應義塾大学)

Part	発表者氏名 (所属)	発表題目
I	Masaki Michiko (Osaka International University) Otoshi Junko (Gunma Prefectural Women's University) Kuru Yukiko (Aichi Medical University)	Examining an EFL writing rubric for classroom use
II	宮崎 啓 (慶應義塾高等学校)	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
III	奥田 利栄子 (広島大学) 大津 理香 (茨城大学)	教員によるスピーチ評価と学生のピア評価との関係
IV	深澤 真 (茨城県立竹園高等学校)	スピーチ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
V	藤森 千尋 (東京大学大学院)	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関係について

第5室 (R棟 1階 R101 教室)

司会 Part I 印南 洋 (豊橋技術科学大学)
 Part II & III 塩川 春彦 (北海学園大学)
 Part IV & V 法月 健 (静岡産業大学)

Part	発表者氏名 (所属)	発表題目
I	酒井 英樹 (信州大学)	テキストの長さと再生言語が L2 リスニング方略に与える影響—ディクテーションと自由筆記再生法に焦点をあてて—
II	柳川 浩三 (University of Bedfordshire, Centre for Research in English Language Learning and Assessment (CRELLA), 神奈川県立大和西高校)	大学入試センター英語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
III	森谷 浩士 (神田外国語大学)	「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
IV	Kazuyuki Shite (Tokyo University of Social Welfare)	ESL vs. EFL in terms of learning difficulty in relation to implicit and explicit knowledge
V	島田 勝正 (桃山学院大学)	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

第6室 (R棟 1階 R104 教室)

司会 Part I 木下 正義 (福岡国際大学)
 Part II & III 野口 裕之 (名古屋大学)
 Part IV & V 大津 敦史 (福岡大学)

Part	発表者氏名 (所属)	発表題目
I	高波 幸代 (筑波大学大学院)	音韻認識とスペリングテストに関する一考察と実験結果
II	秋山 實 ((株) e-ラーニングサービス)	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
III	劉 東岳 (プロメトリック株式会社)	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
IV	中原 敬広 ((株) e-ラーニングサービス)	Moodle 版 TDAP3.0 の開発
V	木村 哲夫 (新潟青陵大学)	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

Plenary Speech (講演)
“Defining, Setting and Validating Standards in Language Testing”
Tony Green (University of Bedfordshire)

Standards, variously defined, are of central concern in language testing and assessment. In the first place, in defining what we are testing, we inevitably appeal to certain ‘standard’ language varieties. Increasingly, within Applied Linguistics, through the increasing recognition of so-called New Englishes, the values inherent in adopting a North American or British standard for English language teaching and testing are being called into question. The English for Specific Purposes movement raises similar issues relating to context-specific language use. Questions about the selection of language varieties employed in language tests both as a basis for judgements about learner production and as a source for comprehension tasks have yet to be adequately addressed by the profession. A second set of issues surround how we test language skills. Professional organisations such as ILTA, ALTE and EALTA have adopted quality standards for the conduct of language tests. Such standards, whether minimal or aspirational, serve as guidelines for professional practice. They lay down principles to help testing bodies to achieve best practice in developing instruments that are accurate, fair, transparent and fit for purpose. Such standards, however, also raise questions of enforcement, ownership and identity. The two sets of concerns come together in the practice of ‘standards setting’ in which standards represent defined levels of language ability that can be communicated across contexts. There is increasing pressure from policy makers to relate test performance to such external frameworks as the Common European Framework of Reference for Languages (CEFR). The draft manual for relating tests to the CEFR recognises that the process demands attention to the language skills being tested, to the technical qualities of the test and to the relationship between performance on the test and the CEFR descriptors. This paper will explore some of the issues involved in communicating outcomes in this way and will discuss some of the implications of frameworks for language testing practice.

Symposium (シンポジウム)

Theme: “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)

Michiko Nakano (Waseda University)

A Case Study at Waseda University: an application of CEFR in English Language Education

Open Education Center at Waseda University has been offering English Tutorial Programs which are based on CEFR. I will briefly introduce English Tutorial programs first and illustrate how we have been using CEFR in our textbooks and in our teachers’ manual. I will also illustrate an experiment to show that our teaching materials as well as teaching methods are based on the six levels of CEFR.

Paper Presentation (研 究 発 表)

第1室 (R001) Part I

「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

ベネッセと語学学校ベルリッツで共同開発した 4 技能の英語能力テスト、GTEC (Global Test of English Communication) のスコアおよび事前アンケートの分析から、「英語で話したり、書いたりする際に、日本語を先に考えず英語でそのまま考えて発信すること」の重要性和学習行動との関係を報告する。GTEC の被験者は、英語を仕事で必要とする大人のビジネスパーソンが大半を占める。英語で考えそのまま発信できる状態を最近では「英語脳」と称することがあるが、「無意識化/自動化 Automatisation」と被験者の英語能力の特徴や、その学習行動を分析し、英語学習・教育における示唆が得られた。

第1室 (R001) Part II

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解明—

柳瀬 陽介 広島大学

言語テストにおいて「コミュニケーション」という概念は鍵概念の一つであるが、その解明は十分ではない。本論では、これまでの言語コミュニケーション力論および関連性理論が十分に明らかにしてこなかった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する。この探究により、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拠 (Selbstreferenz; self-reference) と二重の偶発性(Doppelte Kontingenz; double contingency)などの概念、あるいは「システムと環境の差異」(Differenz von System und Umwelt; difference between system and environment)が明らかになる。これらの解明により、コミュニケーションとは従来言語学や言語テストの論で考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る。

第1室 (R001) Part III

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

Two central issues in the field of language testing are the structure of language ability (e.g., Alderson, 2000; Bachman & Palmer, 1996) and construct definition and interpretation (e.g., Bachman, 2007; Chapelle, 1999). The structure of language ability has been continuously investigated since Oller's (1979) unitary trait hypothesis was proposed (and later refuted). Construct definition and interpretation has been widely discussed in literature; however, the most comprehensive study is perhaps that by Chapelle (1999), which proposes three perspectives toward construct definition and interpretation depending on whether test performance is attributable to trait only, context only, or both trait and context. These two topics have been simultaneously examined by using a multitrait-multimethod (MTMM) design, where two or more traits are measured with two or more methods (Campbell & Fiske, 1959). In order to gain a better understanding of these two areas, the current study aims to quantitatively synthesize previous MTMM studies. An extensive literature search was conducted by searching databases, reviewing books and journals, and communicating with other researchers. Confirmatory factor analysis was used to estimate 10 rival models for each MTMM matrix. These models were classified into three types: trait models, method (context) models, and trait-and-method (context) models. Each of the 10 models was evaluated using the overall fit of the model with the observed data, and the best-fitting model was selected. The results and implications will be reported in the presentation.

第1室 (R001) Part IV

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—

斉田 智里 茨城大学

入学者全員を対象とした英語教育プログラムの構築が大学英語教育界で急速に進展している。TOEIC IP や TOEFL ITP あるいは GTEC といった外部の英語熟達度テストを全学に導入し、教育効果を検証する国公立大学が増加してきた。しかしこうした外部テストでないと教育効果が測定できないというわけではなく、より簡便に、より低コストで教育効果を測定する方法もある。本稿では、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実施することで、項目応答理論を用いて教育効果の検証を試みた事例について報告する。JACET8000 の語彙リストから、語彙サイズテストを2種類作成した。1つは頻度レベルの低い語をより多く抽出し、もう1つは頻度レベルの高い語をより多く抽出し、各100問とした。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、頻度レベルのより低い語からなる語彙サイズテストを成績上位者961名に、頻度レベルのより高い語からなる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するとともに、頻度順と困難度とを対応させた。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し、語彙サイズと対応させることで教育効果の検証を試みた。

第1室 (R001) Part V

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 神田外国語大学 長谷川 信子 神田外国語大学
町田 なほみ 神田外国語大学

本発表では、神田外語大学大学院で進めている早期英語研究プロジェクトの一環として開発した子どものための英語語彙テストについて、テストの概要及びパイロットテストの分析結果を報告した上で、信頼性と妥当性について考察する。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になり、小学生の英語活動に対する関心が一層高まってきている。そのような中で、評価に関する議論は、ほとんどなされておらず、教材や指導法などの議論と比較して、大きな立ち遅れが見られる。神田外語大学研究チームは、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)を認識した上で、語彙知識を中心に据え、Nation (2001)の枠組みを応用し、テスト項目を選定した。800余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について報告すると共に、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する。最後に、小学校段階における語彙を含めた英語力評価に関する諸問題と今後の研究課題についての示唆を考察する。

第2室 (R004) Part I

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

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 および、インタビューテスト、筆記試験を開発・実施した。Can-do-statements の「聞く」「話す」「やりとり」「読む」「書く」の各領域対面テストおよび読み書き判定シートの中の相関関係を検討した結果、口頭能力および書記能力ともかなり高い相関が見られた。日常生活上の頻度の高い日本語使用項目はできるが、使用する機会の少ない、あるいは母語で済ますことのできる項目は難易度の低いものでもできないため、教室習得環境の学習者に比べ、かなり異なる傾向を示し、Can-do-statements も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。全体に初級レベルの受験者が多いため、口頭能力の測定は自発発話が少なく、OPIのような従来の方法を使うことはできない。そのため独自のインタビュー等の方式を開発し実施した。この結果をもとに、就労外国人のための日本語能力レベル判定方法を確立し、学習者支援のガイドラインを策定していきたい。

第2室 (R004) Part II

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

Tomoko Fujita Tokai University

This study attempts to create valid and appropriate can-do descriptors for a listening course in an English language programme. Approximately three thousand Japanese college students in the course are divided into three different levels; advanced (A), intermediate (I), and basic (B) according to the results of the in-house placement test. Initially, the first draft of the can-do descriptors were written based on the descriptors for CEFR, the European Language Portfolio, the International English Qualifications, STEP, and TOEIC, and the ten teachers of the listening course examined them to see if they were appropriate for each level. This resulted in the completion of the can-do checklists along with 14 descriptors for each level. Four hundred fifty randomly chosen students from the three different levels responded to the statements on the checklists. The results were compared with their results of a listening proficiency test. The one parameter IRT analyses were conducted on the results using BILOG MG 3.0. The average ability levels (θ) of the can-do checklists for the students at each level are (A) 0.618, (I) -0.119, (B) -0.276 and the average θ of the listening proficiency test are (A) 1.094, (I) -0.054, (B) -0.780. The differences among the 3 levels are similar, and the correlation coefficient between these two measurements is 0.315 ($p > 0.01$). The presenter will also discuss how revisions were made to the can-do checklists in order to make them level appropriate by utilizing the item difficulty parameter for each descriptor.

第2室 (R004) Part III

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 境 一三 慶應義塾大学
野上 康子 (株) 教育測定研究所 片岡 夏子 (株) 教育測定研究所

慶應義塾は小学校から大学院までを備えた総合的な教育機関であり、文部科学省の助成を受けて2006年から行動中心複言語学習(AOP)プロジェクトに取り組んでいる。より良い一貫教育カリキュラムの開発を目指し、学習者の外国語能力実態調査の一環として、2007年度にヨーロッパ共通言語参照枠(CEFR)にもとづく European Language Portfolio (ELP)のチェックリスト(Schneider, Gunther & North, Brian, 2000) の日本語版を作成し、自己評価アンケートを実施した。本報告では得られた結果をもとに、CEFR チェックリストを日本人英語学習者に適用することの可能性を検討するとともに、CEFR にもとづく学習者の能力レベルの記述を試みる。項目応答理論に基づいて推定した各項目の困難度と想定されている CEFR レベルとの整合性はかなり高かった。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、発達的変化の仕方がスキルによって異なっていることがわかった。また、外部評価テストとしてコンピュータによる適応型テストである CASEC を用いて自己評価能力値との関連を調べたところ、比較的高い相関が観察された。

第2室 (R004) Part IV

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に影響を与えるのか？—初・中・上級者のタスク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

Can-do アンケートによる外国語能力の自己評価は一般的に信頼できるものとされている。一方、その評価に影響を及ぼす要因として、「不安」や「タスクの経験」などが指摘されている。本研究では、学習者の「経験」がどのように評価値に作用するのかを習熟度別、さらには習熟度間で比較検証し、自己評価アンケート結果を解釈する際の留意点を考察する。先行研究では、TOEIC-IP テスト受験者に業務に関連した言語タスク 65 項目の can-do アンケートを行い、その中でタスク経験の有無も尋ねた。タスク経験者の自己評価点の方が、非経験者の点より、程度の差はあるが、概ね高かったので、600 点取得者と 800 点取得者を抽出し、自己評価点の差がどのような条件で変わるかを調べた。タスク経験者と非経験者の自己評価点の差は、タスクの内容（難易度、スキル、シチュエーション）と評定者の習熟度に左右されることがわかった。この結果を受け、本研究では得点の範囲を広げて、習熟度の大きく異なる受験者（初級から上級）を対象に、さらに、経験・タスク内容・習熟度が自己評価にどのように影響を与えるかを追調査し、発表する。

第2室 (R004) Part V

An attempt to improve our understanding of reading test difficulty

Randy Thrasher Okinawa Christian University

Although Alderson separates text variables from reader variables and mentions more than 10 different elements that he claims contribute to determining the difficulty of a reading passage for a particular group of test takers, test writers rarely consciously consider any of these variables individually. We usually just use our experience to decide if a particular passage is going to be easy or difficult for the target test population. This paper discusses 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. Instead of looking only at the text itself and relying on the grammatical features and vocabulary of it as readability formulas do, the check sheet tries to determine the difficulty of the whole test task and is based on a theory of communication, Relevance Theory. Instead of focusing only on grammatical features and vocabulary, the proposed check sheet asks the evaluator to estimate what the target group of test takers brings to the reading task and what the text and questions demand of the reader in terms of communication and comprehension skills. The check sheet asks the test writer to estimate such things as the degree of background knowledge that the test takers would need to bring to that topic, the amount of cultural or encyclopedic knowledge the text requires, and the level of inferential ability required to answer the questions posed.

第3室 (R003) Part I

The construct validity of "pronunciation" exercises in entrance examinations: Student perspectives

Tim Newfields *Toyo University*

Many entrance exams include items requiring examinees to indicate the correct stress or intonation pattern of select words without any external audio input. Since 'pseudo-pronunciation' questions are an ongoing feature in many of high-stakes EFL entrance exams, it is worth exploring the validity of such test items. Does the ability to detect unspoken speech artifacts correlate highly with the ability to detect the same words if spoken in naturalistic speech? How well are students able to pronounce the same words in connected speech? Employing both qualitative and quantitative methodologies, this study contrasts how several dozen Japanese university students responded to three different sets of tasks: (1) a short series of paper-and-pencil word stress and hyphenation discrimination exercises, (2) a short listening transcription exercise with many words from the previous set, and (3) a short reading exercise containing many words from the same set. Comparing these three sources of data, two admittedly tentative conclusions will be offered: (1) the ability to discern word stress and hyphenation patterns does not correlate highly with the ability to pronounce most words in embedded speech, and (2) the ability to understand many words correlates only moderately with the ability to pronounce such words in embedded sentences. In light of these findings, the validity of pseudo-pronunciation items will be called into question. The presentation concludes by highlighting some of the limitations of the current study and suggesting directions for further research.

第3室 (R003) Part II

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 a two-way 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.

第3室 (R003) Part III

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.

第3室 (R003) Part IV

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 山形県教育庁高校教育課

学校現場において、「実践的コミュニケーション能力」育成にかかわる指導の一層の充実が求められる中、指導上、他の力とともに英語スピーキング力測定の必要性も一層高まった。こういった中、全国規模の英語教育改善実施状況調査(2006)によれば、定期試験等でスピーキングテストを実施している高等学校の割合はまだ十分とは言えない状況であり、評価が十分に行われていないか、行われていたとしても実際に発話をさせない間接テスト(indirect test)形式で実施されている現状がある。英語を話す動機付けや、妥当性の問題をふまえると、中村(2004)が主張するように、生徒への余波効果を意識し、実際に発話をさせるパフォーマンス・テストが好ましいと考えられる。本研究では、高等学校におけるコンピュータ普及率の高さに着目し、高等学校学習指導要領を拠り所に「適切さ1 (応答内容)」「適切さ2 (発話の速さ)」「正確さ (リズム・イントネーション)」という3つの評価項目をもつコンピュータを利用した半直接面接テスト(Computer-mediated Semi-direct Interview Test、以下 CSIT)を高等学校で実施し、その有用性を多相ラッシュ測定(Many-Faceted Rasch Measurement)を用いながら、妥当性、信頼性、実用性の3点からのアプローチにより検証した。発表ではその検証結果報告と学校現場への示唆を示したい。

第3室 (R003) Part V

グループ・オーラルテストにおける評価改善の研究

曲 明 東京外国語大学大学院

グループ・オーラルテストは、試験官が同時に何人も測れるので、実用的、経済的なテストであると言われている。しかし、試験官が同時に何人も測らなければならない、更に分析的評価尺度にした場合、同時に数人の被験者について、数項目の評価項目も評価しなければならないため、採点作業は困難で、評価の信頼性と妥当性に疑問がもたれている。そこで、本研究では、日本人中国語学習者（大学生）を対象にしたグループ・オーラルテストの評価改善に焦点を当て、信頼性、妥当性、実用性の観点から見た評価改善方法について提案を行った。スピーキングの評価結果には、「評価者」という要因だけではなく、「生徒」や「評価尺度」といった要因も複雑に絡みあっており、評価の改善を検討するためには、これらの要因ごとの影響を見なければいけない。従って、本研究では、まず、今回のテストで使われた分析的評価尺度と総合的評価尺度間の相関係数を算出することによって、分析的評価尺度の妥当性を検討した。また、評価の信頼性と実用性を高める方法として、一般化可能性理論（Generalizability Theory）を用いて、評価の信頼性と実用性のバランスを検討する方法を提案した。今回の中国語グループ・オーラルテストのデータを使用して、具体的に、評価項目数や評価者数という要因を変化させることによって、信頼性係数がどのように変動するかを検討し、教育現場にとって信頼性と実用性を両立できるような評価方法を探った。今回のデータでは、一般化可能性係数は $G=.60$ を保つには、評価項目は最低3つ、評価者数（採点トレーニングを行う前提で）最低2人であれば、可能であることが明らかになった。本研究はグループ・オーラルテストの評価の信頼性と妥当性の改善に関して、教育現場での使用という実用性を考慮に入れた2つの方法で検討した。明らかになったことを活用すれば、グループ・オーラルテストの評価において飛躍的な効率化が期待できるであろう。

第4室 (R102) Part I

Examining an EFL writing rubric for classroom use

Masaki Michiko *Osaka International University* Otsoshi 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.

第4室 (R102) Part II

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

エッセイライティング分析的評価の各構成要素はそれぞれ特徴をもっているものの、互いの特徴に関連性がある場合、下位項目能力を独立して測定したことにはならない。本研究は、分析的評価の主な側面である修辞面：Rhetorical Features (RF)と言語面：Linguistic Features のうち、RFの部分にのみ焦点を当て、下位項目スケール(Content, Organization, Cohesion, Voice)の構成概念妥当性を考察した。特に、Content は RF 部分の全体的な要素を含んでいると思われるため、(1)他の構成要素の得点で Content の能力をある程度予測できるのではないか？ということに着目し、さらに、予測できるとしたら(2)他の RF の構成要素のうちどれが Content と大きく関連を持つか？について調査を行った。EFL 高校生 74 名(有効データ 70 名)の説得文エッセイを 2 名の採点者で評価し、重回帰分析を行った結果、Content は Voice との関連性が高く、Organization は Cohesion と高い相関を示した。広く使用されている ESL Composition Profile (Jacobs et al., 1981) のように、Content と Organization という 2 点のみを RF の評価スケールに設定することが妥当であることを示唆し、さらに、本研究の調査結果を参考に、いくつかの評価スタンダードの記述子への考察も加える。

第4室 (R102) Part III

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

本研究は教員によるスピーチ評価(TA)と学生のピア評価(PA)との一致の度合いを探ったものである。大学1年生6クラスの参加を得、3クラスずつを2人の教員が担当した。全てのクラスで同じ評価基準を用いたものの、一方の教員は評価基準について説明する際に各項目の評価点(1～5点)の大きさの違いについて実演しながら説明し、もう一方の教員は口頭説明のみを与えた。各クラスとも、4人から成る小グループでスピーチ及び評価を4回繰り返し、5回目に一人ずつのスピーチを実施し、それを教員と学生が同時に評価した。スピーチはビデオ撮影され、教員はビデオを見て互いのクラスのスピーチについても評価した。2人の教員による評価の平均点と学生による評価のクラス平均点との間で相関を調べた結果、6クラス全体で見ると TA と PA の間に有意な相関があることが分かった。基準項目別に見ると「正確さ」については相関が低いことが分かった。また、基準についての説明方法の違いが TA と PA の相関の強弱に大きく係わることも分かり、「評価点(1～5点)の違いについて実演しながら説明した」クラスの学生はそうでなかったクラスの学生より教員評価と似た評価ができたのが分かった。

第4室 (R102) Part IV

スピーチ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

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を発表する。昨年の発表(深澤, 2008)では, 自己評価の総合評価における妥当性を検証したが, 本研究の目的は, (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点である。研究は, 日本人高校生 79 名(有効データ数 62)を対象に調査を行い, 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った。研究の結果, 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。

第4室 (R102) Part V

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

本研究は, 高校1年生, オーラルコミュニケーション I 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である。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調べた。結果としては, 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。

第5室 (R101) Part I

テキストの長さと言語再生がL2リスニング方略に与える影響—ディクテーションと自由筆記再生法に焦点をあてて—

酒井 英樹 信州大学

本研究は、L2 リスニングテストのうち、ディクテーションと自由筆記再生法を取りあげて、受験者がテスト中に用いる L2 リスニング方略を質問紙票によって調べようとするものである。本研究の目的は、(a) まとまった英文をまとめて聞かせる課題と、まとまった英文ではあるが一文ずつポーズを入れて再生させる課題におけるリスニング方略を比較することによって、英文の長さの影響を調べることと、(b) 英語を聞いてできるだけ正確に英語を書くディクテーションと、英語を聞いてできるだけ正確に日本語でその内容を書く自由筆記再生法におけるリスニング方略を比較することによって、再生言語 (L1 か L2) の影響を調べることである。参加者は、日本語を L1 とする英語学習者である大学生 38 名であった。方略に関する質問紙票として、自由記述及び先行研究に基づき、26 項目が作成された。再生言語とテキストの長さの影響を、効果量 (Cohen's d) に基づいて分析した。その結果、テキストの長さが再生言語の違いよりも使用された方略の選択に大きな影響を及ぼしていたことがわかった。さらに、長いテキストを聞かせるディクテーションは、さまざまなリスニング処理を要求している課題であることが示唆された。

第5室 (R101) Part II

大学入試センター英語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 *University of Bedfordshire, Centre for Research in English Language Learning and Assessment (CRELLA)*, 神奈川県立大和西高校

研究の目的は、大学入試センターリスニング試験 (以下 J-NCT[L]) の妥当性検証を行い、試験の改善点を指摘し、高校生の英語リスニング力の向上に寄与することである。妥当性検証を行うにあたっては、社会・認知妥当性の枠組みを(Weir 2005, 2008)用いる。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(situational authenticity)と言語使用者の認知処理(interactional authenticity)とがテストタスクにどの程度反映されているかを問うものである。大学入試センター試験に英語リスニング試験が導入されて3年が経った。J-NCT[L]の妥当性検証は、しかし、ほとんどなされていない。また、大学入試センターも実証的証拠(empirical evidence)を提出していない。そこで、今回の発表ではまず、妥当性検証に使う L2 リスニングモデルを構築・提示し、リスニングテストタスク要因を包括的に提示する。そして、この二つの枠組みを使って、J-NCT[L]が社会・認知妥当性をどの程度実現しているかを検証した分析結果をあきらかにする。具体的な研究課題と研究方法は次の三つであった。

1. J-NCT[L]が言語・社会的コンテキストをどの程度実現しているか、量的・質的分析を行った。
2. J-NCT[L]が学習指導要領をどの程度反映しているか、量的・質的分析を行った。
3. J-NCT[L]は、実際の言語使用時の認知処理をどの程度反映しているか。専門家三人が質的分析を行った。

第5室 (R101) Part III

「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 神田外国語大学

本研究は、談話処理方略の転移 (Harrington, 1987; Koda, 2005) という観点から、日本語の談話の特徴である「省略」に着目し、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。特に、異なる母語背景を持つ学習者（中国語母語話者と英語母語話者）間で、理解の方略に違いが見られるかを考察した。英語は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、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探った。先行研究では、省略に関しても、産出データを扱うものを中心であるが (Jin, 1994; Jung, 2004)、本研究では、省略部分が産出データに現れないことが習得の証明にはならないことを論じ、「解釈」に焦点を当てる。省略のある自然な文章と、省略要素を復元した文章を読んで、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を実施した。その結果、読みやすさの判断及び省略要素の解釈共にグループ間に違いが見られた。本研究の結果は、談話処理方略に母語の影響がある可能性を示唆するものである。

第5室 (R101) Part IV

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

Kazuyuki Shite Tokyo University of Social Welfare

This study attempted to replicate Ellis (2006), which found that some linguistic features were difficult in terms of implicit knowledge, but easy in explicit knowledge, or vice versa after making sure a Principal Component Factor Analysis produced a two-factor solution for a battery of tests designed to assess either implicit or explicit English grammatical knowledge. While the participants of his study were L2 learners who had learned English in diverse learning contexts, mostly ESL contexts, those of the present study were all Japanese EFL learners so that the effects of learning contexts on SLA would be observed. Thus, learning difficulty in relation to these two types of knowledge in the EFL context was investigated. Fifty-seven Japanese university students participated in a battery of tests: Oral Imitation Test (OIT) and Timed Grammaticality Judgment Test (GJT) to measure implicit knowledge, and Untimed GJT and Metalinguistic knowledge test to measure explicit knowledge. After a factor analysis was performed, the mean accuracy percentage score of the OIT and Timed GJT was subtracted from that of the Untimed GJT and Metalinguistic knowledge test so that difference in difficulty between implicit and explicit knowledge could be examined. It was found that whereas Ellis (2006) detected 6 structures in which the difference between the mean scores was over 20 %, the present study had 11 structures. Therefore, the difference in difficulty between the EFL learners' two types of knowledge seemed bigger than the ESL learners.' Moreover, the extent to which ELF learners tend to develop explicit knowledge compared to implicit knowledge is discussed.

第5室 (R101) Part V

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

本研究は、文法性判断テスト(Grammaticality Judgment Test; GJT)において、問題文の提示時間の制限の有無により、明示的/暗示的知識(Explicit/Implicit Knowledge)の2つの異なる文法能力を測定することができるか否かを調べた島田(2007)の継続研究である。1つの文法範疇毎に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。OITではGRとUGの相関が高く、2つのGJTでは両者の相関は低かった。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、文法的適格性の有無(GR:UG)による2因子モデル、さらにOIT, GJTGR, GJTUG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6室 (R104) Part I

音韻認識とスペリングテストに関する一考察と実験結果

高波 幸代 筑波大学大学院

音韻認識とは、話し言葉における音声構造を理解し、語の音素構造に敏感である、感性のことで、特に読み書きの能力と、非常に深く関係している(門田,2004; Harley, 2001)。一般的に、L1では音韻認識を自然に身に着けるが、インプット量の異なるL2では獲得が困難、とされている(Treiman, 2003)。本研究では、特につづり字学習と音韻認識の関連に注目し、中学校1年生に計15語の未知語を、異なる3つの方法によって5語ずつ学習させた。音と文字の2要素をそれぞれ①音(発音)単独学習、②文字単独学習、そして③音・文字同時学習、の3方法に分けた。調査前に、学習者の日常的なつづり字学習について調査し(RQ1)、その後、3つの方法で学習をした後に行った直後テストの結果比較(RQ2)、最後に調査の前後で行った質問紙調査(選択肢あり・自由記述)を比較し、学習者がつづり字学習にどの程度意識を向けるようになったか、を質的に分析(RQ3)した。つづり字学習の直後テストにおいては、③音・文字同時学習の得点に有意差があった。学習者の意識調査も並行して行った。3つの方法のうち最も有用と思えるものを事前選択させたが、調査後に同じ質問をし、有用なもの、有用でなかったものを選択させ、その理由を細かく書かせた。自由記述の内容分析の結果、学習方法の提示により、学習者に音と文字の対応を意識づけることができた、ということが明らかになった。アルファベットの単なる羅列として単語を認識するのではなく、音を持つ単語として意識させることができた。つづり字学習法を提示し、事後テスト、遅延テストを行うことで、学習者に、つづり字学習を意識して行うことの重要性を示すことができ、この点に教育的な示唆がある、と考えられる。

第6室 (R104) Part II

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。

第6室 (R104) Part III

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。

第6室 (R104) Part IV

Moodle 版 TDAP3.0 の開発

中原 敬広 (株) e ラーニングサービス

大友、中村および秋山が開発したテストデータ分析プログラム TDAP2.02 は、単体のツールとしての完成度は高い。しかし、e ラーニングが普及し始めた現在、LMS の持つオンラインテスト機能を利用する場合には、いくつか改善すべき点が残っていた。秋山は、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 tlap 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た。本発表では、これを最新版の Moodle 用に作り直し、さらに、1) 構成したテストのテスト情報関数をそのテストを構成するテストアイテムの項目情報関数とともに重ね合わせてグラフ表示することでテスト設計を補助する機能、2) 受験者シミュレータを使って受験結果を確認できるシミュレーション機能、3) テストアイテムに使用されている語彙の頻度レベル判定機能を持つアイテムエディタ : E-Edit、などを付加した。これにより、Windows 版 TDAP2.02 の持つ機能をすべて備え、かつ、新たに強力な補助機能を持たせることができ、客観的な測定ツールとしてのテストを適切に作成することが容易になった。

第6室 (R104) Part V

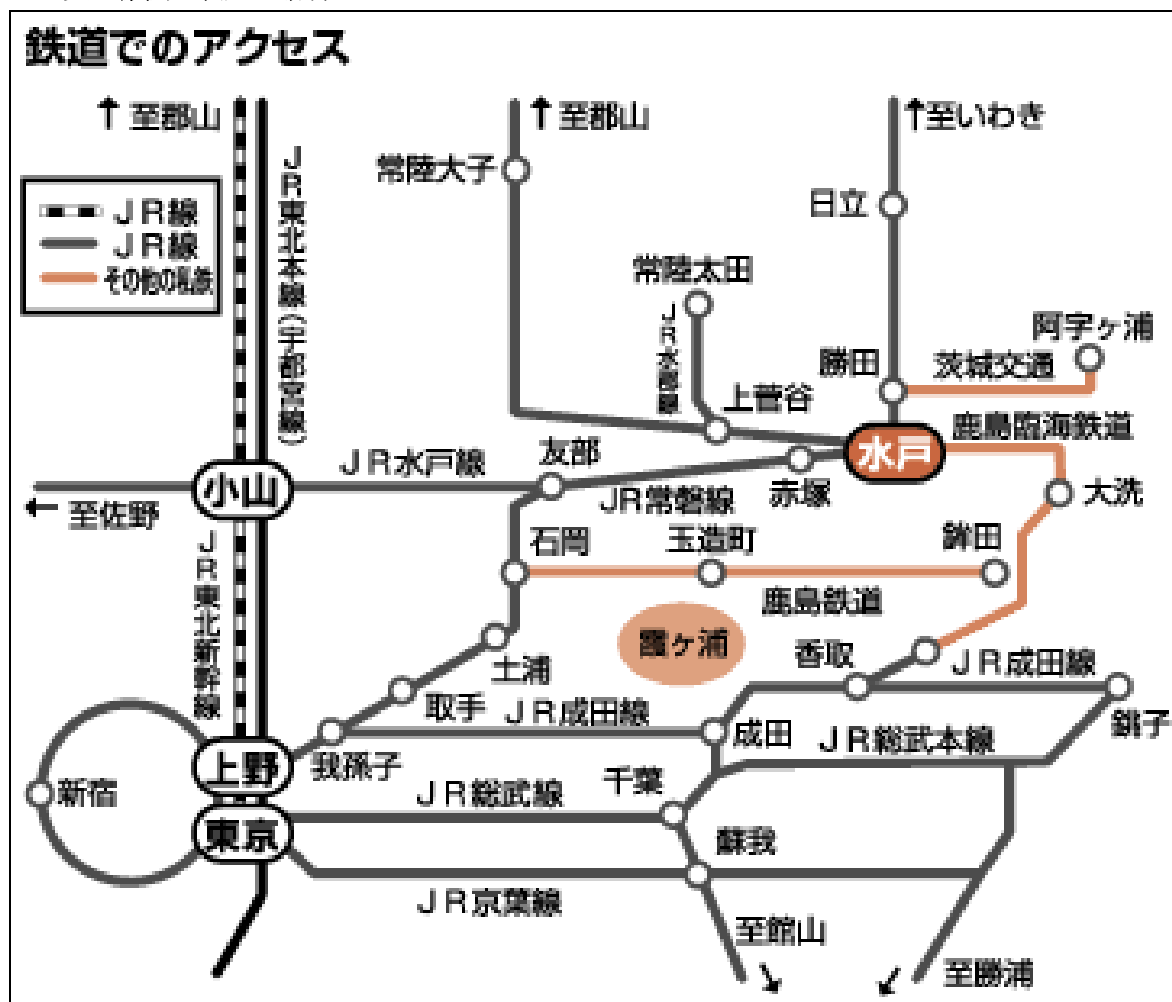
習熟度別クラス編成のための英語基礎力判定標準化テスト作成の試み

木村 哲夫 新潟青陵大学

大学に入学してくる学生の英語力が多様化していることから、英語の授業を習熟度別クラス編成で行う（あるいはその必要性を感じている）大学が増えてきている。学生の英語基礎力を判定に、英検や TOEIC などいわゆる資格試験や、英語力判定のために開発された様々なテストが利用されることも多いが、近年の IT 技術の進歩により、かつては個人レベルで行うことが技術的にも費用的にも考えられなかったテストデータを含む学習データの管理と分析が、安価にかつ容易に行えるようになってきた。本研究は、オープンソースウェアでありながらも充実した機能を持つ LMS として定評のある Moodle を使い、英語基礎力を判定するテストを作成・実施し、データを管理・分析し、標準化されたテストとして、習熟度別クラス編成のための基礎資料を提供する試みを紹介する。テスト項目としては独自のものと、英検の過去問題を使用許諾を得て利用した。データ分析手法としては、1 パラメータと 2 パラメータのロジスティック・モデルの両方を試した。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行い、本研究のテストの妥当性の検証も行った。

■水戸まで (To Mito)

公共交通機関ご利用の場合



都心からご利用の場合は下記が便利です。

■「上野駅」から「水戸駅」までJR常磐線の特急「スーパーひたち」「フレッシュひたち」。(所要時間約65分)。片道4,220円

■「東京駅」と「水戸駅」北口までを結ぶ高速バスが運行。「自由ヶ丘」下車。(所要時間下り約100分、上り約140分)。片道切符は2,080円。ツインチケットを買いと往復3,500円(片道1,750円。バス内でも買えます)。

※時間帯により発着場所が新宿または上野の場合があります。

[From Downtown Tokyo to Mito]

1. Take the Joban line (Japan Railways; JR) from Ueno Station to Mito Station. It takes about 65 minutes by the JR special express train.
2. Take the JR Highway Bus from Tokyo Station to Mito Station and get off at Jiyugaoka Bus Stop. It takes about 100 minutes from Tokyo to Mito.

できるだけ公共機関でお越しくださいますよう、お願い申し上げます。

■水戸駅・赤塚駅から常磐大学まで (From Mito Station or Aakatsuka Station to Tokiwa University)



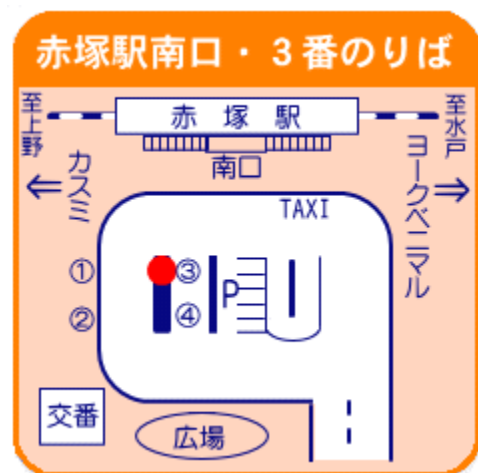
■水戸駅北口 5 番バス乗り場から自由ヶ丘下車 (所要時間約 20 分)

※直接「常磐大学」経由のバスもあります。その時刻表は下のバス時刻表をご覧ください。

■赤塚駅駅前ロータリー停留所(水戸駅行き)から常磐大学前下車。

[From Mito Station to Tokiwa University]

Take a bus from Stop #5 at Mito Station (North exit). It takes about 20 minutes from Mito Station to Jiyugaoka Bus Stop. Some buses go directly to Tokiwa University, but the service is limited (see the Timetable below).



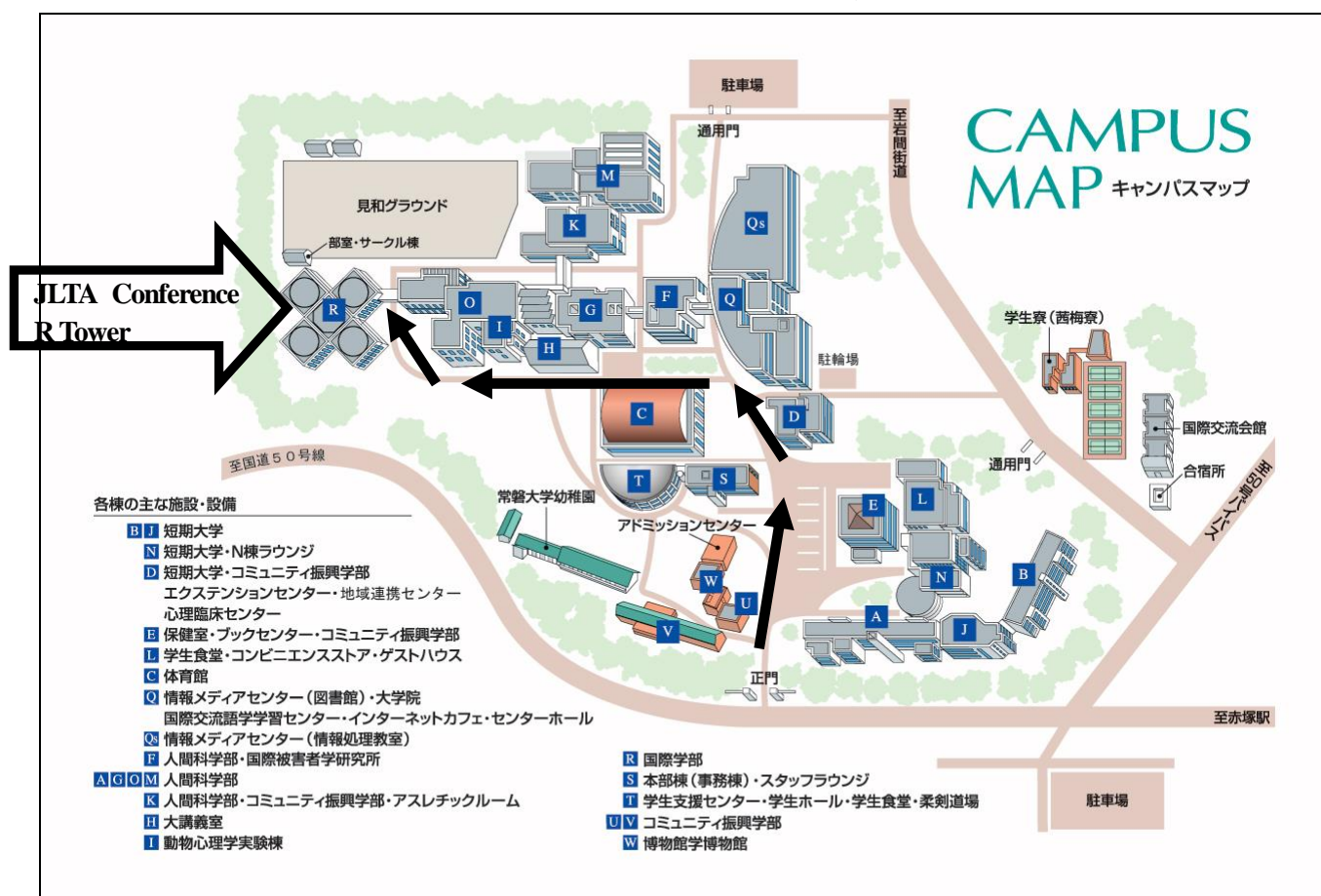
■バス時刻表 (Bus Timetable)

	水戸駅北口→常磐大学 Mito Station North Exit→Tokiwa University		赤塚駅北口→常磐大学 Akatsuka Station North Exit→Tokiwa University		常磐大学→水戸駅北口 Tokiwa University→Mito Station North Exit		常磐大学→赤塚駅北口 Tokiwa University→Akatsuka Station North Exit	
	JR Bus	茨城交通	JR Bus	茨城交通	JR Bus	茨城交通	JR Bus	茨城交通
7	59	59		44		51		
8	18 31	18 59	26 41	14 42	32 47	21 49	12 31 44	12 31
9	4 39	23 42	12 38	15 48	18 44	22 55	17 52	12 36 55
10	18	17 33	6 54	10 54	12	17	31	30 46
11	0 19 59	00 48	27 54	27	0 33	01 34	13 32	13
12	23	23	54	18 55	0	25	12 36	01 36
13	19 59	02 53	17	05 38	0 23	02 12 45	32	15
14	32	23	18 30	23 33	24 36	30 40	12 45	06 36
15	46	02 33 58	53	38	59	45	59	16 46
16	7 47	31	16	00 43	22	07 50	20	11 44
17	34	01 30 38	0 44	05 37 47	6 50	12 44 54	0 47	14 43 51
18	35	13 38	0	23	6	30	48	26 51
19		18	0	10 43	6	16 49		29
20		02		28		34		13

注：水戸駅北口⇄常磐大学は、270 円。赤塚駅⇄常磐大学は、150 円。ただし、JR バスで、Suica か pasmo を使うと、4 割引（それぞれ 160 円、90 円）。水戸駅北口⇄自由が丘は、230 円で、通常 10 分間隔で運行。タクシーだと、1,300 円程度。

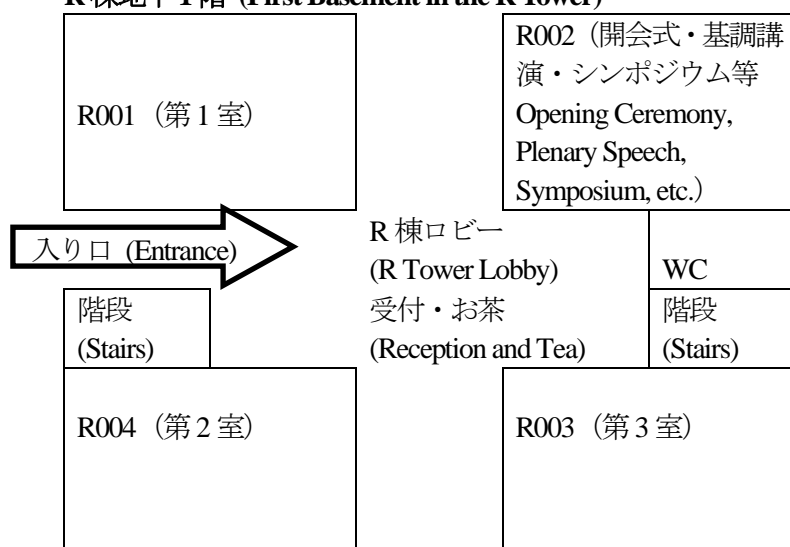
■学会が開かれる R 棟まで (To R Tower)

バスから降りると、常磐大学正門前です。

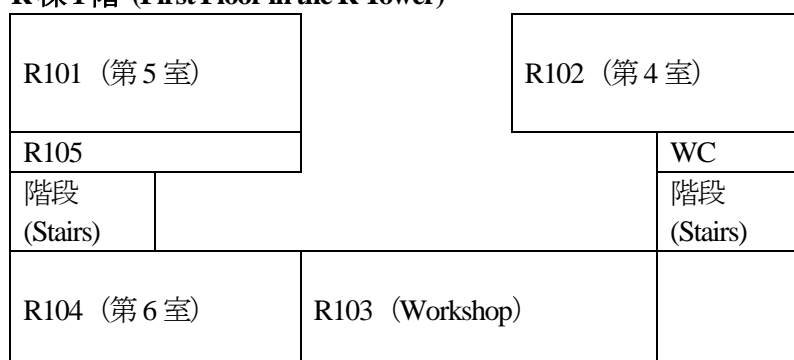


■R 棟の配置図 (Layout of the R Tower) 実際は円形の棟です。

R 棟地下 1 階 (First Basement in the R Tower)



R 棟 1 階 (First Floor in the R Tower)



■タクシー会社(一部)の電話番号 (Telephone number of taxi companies)

- (1) 茨城第一交通有限会社 TEL : 029-241-6611
茨城県水戸市見川 1-1233-1
 - (2) 観光第一交通株式会社 TEL : 029-253-1123
茨城県水戸市石川 1-4002-38
 - (3) 安全第一交通有限会社 TEL : 029-253-1123
茨城県水戸市石川 1-4002-38
 - (4) 新星タクシー TEL : 0120-21-9191
茨城県水戸市宮町 2 丁目 1-6
 - (5) 関東水戸タクシー (株) 本社 TEL : 0120-21-1996
茨城県水戸市城南 2 丁目 7-17
- (常磐大学近辺 1, 2, 3 ; 水戸駅北口近辺 4 ; 水戸駅南口近辺 5)