

The 10th JLTA Annual Conference

Place Building No. 21,
Ryukoku University (Fukakusa Campus)
Dates October 28 (Saturday) & 29 (Sunday), 2006

Day One (October 28, Sat)

Reception (8:50-) [Exhibition Lobby, 6th floor]

Opening Ceremony (9:20-9:50) [Place 603]

Break (9:50-10:00)

Presentation Part I & Part II (10:00-11:25) ***Presentation titles are given on pages 3-8.**

- Part I (10:00-10:40)

- Part II (10:45-11:25)

Room 403 (A) Papers 1 & 2

Room 404 (B) Papers 3 & 4

Room 407 (C) Papers 5 & 6

Presentation Part III (11:35-12:15)

Room 403 (A) Paper 7

Room 404 (B) Paper 8

Room 407 (C) Paper 9

Room 503 (D) Paper 10

Announcements (12:15-12:25)

Lunch Break & Committee Meetings (12:25-13:45)

Presentation Part IV & Part V (13:45-15:10)

- Presentation IV (13:45-14:25)

- Presentation V (14:30-15:10)

Room 403 (A) Papers 11 & 12

Room 404 (B) Papers 13 & 14

Room 407 (C) Papers 15 & 16

Room 503 (D) Papers 17 & 18

Break (15:10-15:20)

Presentation Part VI & Part VII (15:20-16:45)

- Part VI (15:20-16:00)

- Part VII (16:05-16:45)

Room 403 (A) Papers 19 & 20

Room 404 (B) Papers 21 & 22

Room 407 (C) Papers 23 & 24

Room 503 (D) Papers 25 & 26

10th Anniversary Celebration & Party (17:10-19:50)

[Dining Hall, Basement, Building No. 4]

Day Two (October 29, Sun)

Reception (8:40-) [Exhibition Lobby, 6th floor]

Presentation Part VIII (9:00-9:40)

Room 601 (E) Paper 27

Room 602 (F) Paper 28

Room 603 (G) Paper 29

Break (9:40-9:55)

Plenary Speech (9:55-11:25) “Asian Language Testing and the Threat of Scales”

Bernard Spolsky

[Room 603]

Announcements (11:25-11:30)

Break (11:30-11:40) *Non-members will have lunch after this.

General Business Meeting (11:40-12:05) [Room 603]

Lunch (12:05-13:15)

Symposium (13:15-15:15)

Theme: Language Testing in an Asian Framework

Coordinator: Steven Ross (Kwansai Gakuin University, Japan)

Panelists: Lianzhen He (Zhejiang University, P.R. China)

Oryang Kwon (Seoul National University, R. Korea)

Yuji Nakamura (Keio University, Japan)

[Room 603]

Break (15:15-15:25)

Closing Ceremony (15:25-15:35)

* The group photo will be taken after the ceremony.

[Room 603]

第 10 回 大 会 プ ロ グ ラ ム

10 月 27 日 (金)

16:30 ～ 18:00 理事会 (ホテル法華クラブ京都)

10 月 28 日 (土) (大会 1 日目)

8:50 ～ 受付 (6 階展示ロビー)
(PC 利用発表者：発表教室で機器接続確認)

9:20 ～9:50 開会行事 (603 教室)

総合司会 法月 健 (大会運営委員長・静岡産業大学)
Soo im Lee (大会実行委員長・龍谷大学)

挨拶 Randy Thrasher (JLTA 会長・沖縄キリスト教学院大学・
国際基督教大学名誉教授)

会場校挨拶

10:00～11:25 研究発表 (発表 30 分, 質疑 10 分) Part I 10:00～10:40
Part II 10:45～11:25

A 室 (403 教室) 司会 清水 裕子 (立命館大学)

[1] Part I A Study of Face Validity in a High School English Grammar Test: Inter-personal
Authenticity and Affective Responses Made by Test-takers
山下 次郎 (東京都立国分寺高等学校)

[2] Part II 清泉アカデミック Can-Do 尺度の開発とその評価
長沼 君主 (清泉女子大学)
宮嶋 万里子 (清泉女子大学)

B 室 (404 教室) 司会 飯村 英樹 (常磐大学)

[3] Part I スピーキングテストの種類が持つ波及効果研究
正頭 英和 (関西大学大学院博士課程前期課程)

[4] Part II パソコンによるスピーキング・テストと対面式スピーキング・テストに
おけるパフォーマンスの比較
周 育佳 (東京外国語大学大学院博士課程)

C室（407 教室） 司会 藤田 智子（東海大学・東京工業大学大学院）

[5] Part I Predicting Reading Comprehension Item Difficulty with Respect to Content Characteristics

Kiyomi Yoshizawa (Kansai University)

Soo im Lee (Ryukoku University)

[6] Part II Are we talking about the same test? How do teachers and learners view the impact of the TOEIC on language learning?

Mark Chapman (Hokkaido University)

11:25～11:35 休 憩

11:35～12:15 研究発表（発表 30 分，質疑 10 分） Part III

A室（403 教室） 司会 **J. K. Hubbell**（法政大学）

[7] Part III Meta-Validation of Institutional TOEIC Research in Japan

Steven J. Ross (Kwansei Gakuin University)

Hiromu Yamada (TOEIC Steering Committee)

B室（404 教室） 司会 小山 由紀江（名古屋工業大学）

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

秋山 實（合資会社eラーニングサービス）

井川 悦朗（NPO英語運用能力評価協会）

白戸 治久（NPO英語運用能力評価協会）

C室（407 教室） 司会 島田 勝正（桃山学院大学）

[9] Part III リスニングテストにおいて情報の検索および統合が項目困難度を与える影響

土方 裕子（筑波大学大学院博士課程、日本学術振興会）

D室（503 教室） 司会 大坪 一夫（麗澤大学）

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

庄司 恵雄（お茶の水女子大学）

野口 裕之（名古屋大学）

春原 憲一郎（海外技術者研修協会）

12:15～12:25 連絡

12:25～13:45 昼 食 (役員会 506 教室)

13:45～15:10 研究発表 (発表 30 分, 質疑 10 分) Part IV 13:45～14:25
Part V 14:30～15:10

A 室 (403 教室) 司会 前田 啓朗 (広島大学)

[11] Part IV 海外短期英語研修の口頭英語面における効果と測定

横川 博一 (神戸大学)

鈴木 正紀 (Ordinate Corporation)

[12] Part V テスト改訂における妥当性検証—新 TOEIC フィールドスタディー
結果

三橋 峰夫 (財団法人国際ビジネスコミュニケーション協会)

B 室 (404 教室) 司会 塩川 春彦 (北海学園大学)

[13] Part IV リスニング・リコールの採点について—アイデア・ユニットを分
析単位にした場合—

飯村 英樹 (常磐大学)

[14] Part V 並べ替え問題の部分採点法

安間 一雄 (玉川大学)

C 室 (407 教室) 司会 土方 裕子 (筑波大学大学院博士課程)

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

笹尾 洋介 (京都大学大学院修士課程)

[16] Part V ライティングにおける産出語彙の利用可能性: より良い評価へ貢
献するために

占部 昌蔵 (兵庫教育大学大学院修士課程)

D 室 (503 教室) 司会 Mark Chapman (Hokkaido University)

[17] Part IV IRT, testing, and online technology: the future of autonomous
vocabulary learning?

Brent Culligan (Aoyama Women's Junior College)

Charles Browne (Aoyama Gakuin University)

[18] Part V New Ways of Testing and Teaching Vocabulary Online

John Paul Loucky (Seinan Women's University)

15:10～15:20 休 憩

15:20～16:45 研究発表 (発表 30 分, 質疑 10 分) 発表Ⅵ 15:20 ～16:00
発表Ⅶ 16:05～16:45

A室 (403 教室) 司会 長沼 君主 (清泉女子大学)

[19] Part Ⅵ 高校生英語リスニング力に関する Can-do statements の妥当性検証
吉池 陽子 (ARCLE 研究員)
根岸 雅史 (東京外国語大学大学院)
沓澤 糸 ((株)ベネッセコーポレーション)

[20] Part Ⅶ 英検 Can-do 調査報告
柳瀬 和明 ((財)日本英語検定協会)
松平 知樹 ((財)日本英語検定協会)

B室 (404 教室) 司会 小泉 利恵 (常磐大学)

[21] Part Ⅵ 日韓大学生の英語「方略能力」に関する一考察
達 川 奎 三 (広島大学)

[22] Part Ⅶ 次世代英語教育におけるコミュニケーションとその評価ーコミュニケーションに対する理論的考察とそれを応用した評価モデルの提言ー
山中 司 (北陸大学非常勤 慶応義塾大学大学院博士課程)
鈴木 佑治 (慶應義塾大学)

C室 (407 教室) 司会 Steven Ross (Kwansei Gakuin University)

[23] Part Ⅵ The Effectiveness of a 15-minute Writing Sample for Assessing Writing Ability
Melvin R. Andrade (Sophia Junior College)

[24] Part Ⅶ Critical language testing: The General English Proficiency Test
Chih-Min Shih (Chin-Yi Institute of Technology, Taiwan)

D室 (503 教室) 司会 吉澤 清美 (関西大学)

[25] Part Ⅵ I R Tを利用したテストの改訂
藤田 智子 (東海大学・東京工業大学大学院)
前川 眞一 (東京工業大学大学院)

[26] Part Ⅶ テスト自動管理・作成プログラム (Item Banking System) の開発
平井 明代 (筑波大学)

16:35～16:40 連絡

17:10～19:50 **10 周年記念行事・懇親会** (4 号館地下グリル)
挨拶 **大友 賢二** (JLTA 名誉会長、筑波大学名誉教授)
「わが愛しの JLTA」

司会 **小山 由紀江** (名古屋工業大学)
J. K. Hubbell (法政大学)

10 月 29 日 (日) (大会 2 日目)

8:40 ～ 受付 (6 階展示ロビー)
(PC 利用発表者：発表教室で機器接続確認)

9:00～9:40 **研究発表** (発表 30 分, 質疑 10 分) **Part VIII**

E 室 (601 教室) 司会 **島谷 浩** (熊本大学)

[27] **Part VIII** 英語関係節の習得研究における可変的研究結果と文法性判断テスト
伊藤 彰浩 (愛知学院大学)

F 室 (602 教室) 司会 **安間 一雄** (玉川大学)

[28] **Part VIII** **ELPA 診断テスト (文法) の妥当性検証：項目応答理論による分析を
中心に**
小泉 利恵 (常磐大学)
酒井 英樹 (信州大学)

G 室 (603 教室) 司会 **静 哲人** (関西大学)

[29] **Part VIII** **Creating pedagogic tasks based on analyses of English university
entrance examinations**
Keita Kikuchi (Waseda University)
Brian Wistner (Tokyo Junshin Women's College)

9:40～9:55 休憩

9:55～11:25 **講演** (603 教室)

司会 **Soo im Lee** (龍谷大学)
紹介 **Randy Thrasher** (JLTA 会長)

演題： Asian Language Testing and the Threat of Scales
講師： **Bernard Spolsky** (Bar Ilan University, Israel)

11:25～11:35 連絡

11:35～11:40 休憩

11:40～12:05 総 会（603 教室）

議長選出

報告 中村 洋一（JLTA 事務局長・常磐大学）

12:05～13:15 昼 食

13:15～15:15

シンポジウム（603 教室）

Language Testing in an Asian Framework
（Asian Framework における Language Testing）

コーディネーター: **Steven Ross** (Kwansei Gakuin University, Japan)

パネリスト: **Lianzhen He** (Zhejiang University, P.R. China)

Oryang Kwon (Seoul National University, R. Korea)

Yuji Nakamura (Keio University, Japan)

15:15～15:25 休 憩

15:25～15:35 閉会行事（603 教室）

* 閉会の後、集合写真

展示協賛企業

合資会社 e ラーニングサービス

英語運用能力評価協会

株式会社 Lexxica

日本英語検定協会

株式会社 教育測定研究所

(財) 国際ビジネスコミュニケーション協会

Plenary Speech (講演)

Asian language testing and the threat of scales

Bernard Spolsky (Bar Ilan University, Israel)

When I consulted Google (that tempting new servant of scholarship and knowledge) for leads on my topic, it sent me first to a beautiful site that presented the difference between Chinese, Japanese and Korean musical scales, and next to a frightening account of the invasion of Florida in 1992 by the Asian Cyclad Scale, an insect which rapidly killed queen sago trees in Miami and subsequently spread across the US as infected plants were shipped across the country. I am tempted by those two colorful examples to ponder two questions: will the attempt to develop a Common Asian Framework lead to an effort to build a single scale for language proficiency for all Asian countries (as seems to be happening in Europe), and will the infection of a single scale based on current Western models force conformity rather than permit intelligent localized diversity.

Of course the underlying language testing question about scales is not new. Thorndike (1910) assumed for many years that a testing scale was a database of graded and judge-rated examples to be used by an examiner to assign a score to a new production, but ultimately came to believe (Monroe 1939) that it was possible to produce a scaled language test, on which each item was one unit more difficult than the last. Most of our psychometric techniques assume that language proficiency forms a single unidimensional scale, although Henning (1992) showed us that even tests designed to be multidimensional fulfill the requirement of psychometric unidimensionality. But even a few minutes' thought will produce countless examples of multidimensionality in various aspects of language proficiency, showing the need for profiles rather than single scales. However, we also know that tests users demand the simplicity of ranking and cut-off scores. What are our social responsibilities as language testers faced with this paradox?

Henning, Grant. (1992). Dimensionality and construct validity of language tests. *Language Testing*, 9(1), 1-11.

Monroe, Paul (Ed.). (1939). *Conference on examinations under the auspices of the Carnegie corporation, the Carnegie foundation, the International institute of Teachers college, Columbia university, at the Hotel Royal, Dinard, France, September 16th to 19th, 1938*. New York City: Teachers College, Columbia University.

Thorndike, Edward L. (1910). Handwriting, *Teachers College Record* (Vol. 11).

Symposium (シンポジウム)

Language Testing in an Asian Framework

Steven J. Ross (Coordinator)

Kwansei Gakuin University

School of Policy Studies

The concept of a framework for a pan-Asian system of language assessment has become increasingly attractive as diverse and complex language testing practices are now employed for social, educational, meritocratic, economic, and strategic purposes. This diversity presents formidable challenges to those who seek a cohesive and useful framework for language assessment. A framework provides the basis for interpreting how evidence can be understood to instantiate or refute a theory (Notturmo, 1994). Multiple frameworks may be antithetical to a coherent grounded theory because they easily license contradictory cultural conventions. They also can undermine the basis for a code of practice.

This symposium seeks to explore the possible common ground for an Asian framework linking the ways Chinese, Korean, and Japanese language assessment scholars' accounts of 'what is meant to know a language' (Spolsky, 1985) can be shared and understood. In this sense, an immediate question before us is whether a 'framework' is to be developed as a singular pan-Asian concept. Is to be understood in the plural, or as the subset of commonalities shared by different Asian communities of practice?

The three symposium participants address the issue of a pan-Asian framework from the diversity of their own national contexts. He Lianzhen focuses on social issues related to the assessment of English as a foreign language in China. Prof. He examines recent moves to develop national college entrance examinations, and considers the influence of coaching and cramming schools on efforts to reform language assessment. She also addresses the relevance of ILTA's Code of Practice on language assessment practices in China.

Also addressing centralized entrance examination reform is Oryang Kwon. Prof. Kwon considers the long term effect of introducing listening on the national college entrance examination in Korea. He considers also the current policy of using many different kinds of commercial tests in Korea for certification, admissions, exemption, and as indicators of school achievement.

Prof. Nakamura Yuji outlines the antecedents of international benchmarks and standards, and takes on the issue of 'international English', which is currently defined in relation to US and UK standards of usage. Prof. Nakamura will consider three different approaches to creating a common framework: The Common European Framework, test development theory, and the quest motivating the symposium – finding the common ground linking testing practices in China, Korea, and Japan.

The role and extent of language testing in China

He Lianzhen (Panelist)

Zhejiang University, P.R. China

Testing in China has a long history and the word *kaoshi* (test or testing) has become a household word. As early as the Sui and Tang dynasties, the imperial examinations were used for the selection of government officials, and the imperial examination system played a positive role in ancient China for a time. In modern China, testing has been a very important part of the education system, drawing much attention from the public, and the social influence of tests has been widely discussed.

The present paper will discuss the role and extent of language testing in China, addressing issues like the social uses of tests in general, the influence of standardized testing, the washback of some high-stakes tests like the National College Entrance Exam, Graduate Entrance Examination, College English Test (CET) etc, the role of coaching schools, the relevance of ILTA's Code of Practice to testing practices in Chinese culture and so on. The aim of the discussion is to share the experiences of language testing practice in China with our Japanese and Korean colleagues and to find out what the similarities and differences are in the three nations.

English Language Testing in the Korean Framework

Oryang Kwon (Panelist)

Seoul National University

English language testing is now drawing attention in Korea, as Koreans are becoming interested in learning English for communication purposes more than ever. Traditionally, paired with the age-old grammar-translation method of teaching, English language testing had been mainly a matter of measuring the testees' grammatical knowledge, vocabulary, and reading comprehension.

In 1993, with the innovative College Scholastic Ability Test (CSAT) that introduced a listening section for the first time in the history of college entrance examination in Korea, test writers and teachers began to embrace the idea that listening is an important part of English language ability.

The testing industry has produced or imported English language tests for all levels of testees, from elementary school students to adults. Most of these tests offer measurements for only listening and reading comprehension, mainly for practicality reasons. Some recent tests, however, do offer assessment in speaking and writing, although they are still small in number. Because of the increasing discontentment with the insufficient assessment of English skills through the listening and reading tests, speaking and writing tests, especially the speaking tests, are beginning to draw the attention of more serious employers and school administrators.

I will review the current situation of the English language testing in Korea in terms of the available tests, whether indigenous or foreign, and the general characteristics of these tests and the trend of the testing profession. It is hoped that the common issues and differences in language testing in Asian countries will be identified and ideas shared for the improvement of language testing.

The Necessity for Co-operation in Language Testing in Asia

Yuji Nakamura (Panelist)

Keio University

Is there a rationale for considering the necessity for co-operating in language testing in Asia? Let us take a look at the following three current events or statements.

First, some examination boards are already discussing the issue of how to make English language testing more relevant to the international needs of many test-takers, rather than continuing to consider relevance purely in relation to what is real and grammatical for the relatively small number of EIL (English as an International language) and ENL (English as a Native language) speakers (Jenkins, 2006). Davies et al. (2003) argue that existing international tests of English are 'biased' because they 'represent the old colonial Standard English of the UK, USA, etc.' In Japan such existing international tests as TOEFL, TOEIC and IELTS have been quite popular.

Second, Spolsky (2005) claimed that each Asian nation must develop its own set of goals and guidelines for teaching English and not rely only on models supplied by British and American agencies or businesses. He believes that developing an Asian framework will lead to common understanding of local differences (Proshina, 2006). Japan is one of the Asian nations, and it is worthwhile to consider one individual country in a larger Asian setting in terms of language testing.

Third, there was a proposal of an Asian Common Framework of Reference for Languages in the 2nd Asia TEFL international conference in Seoul. Although it does not exclusively mention teaching, TEFL naturally will include testing as well. The point of this framework is for us to have a broader perspective beyond our national boundaries.

But this third one also has urged language testers to think about one of the most important aspects of testing--washback/impact. Teaching, learning, and testing are two sides of one coin. When the teaching part of TEFL in Asia is being discussed, how can language testers keep silent? When you can make a good test, you can teach more effectively.

The three events or statements mentioned above have convinced us of considering the co-operation between three countries in language testing in the framework of Asia.

There are some approaches to explore the possibility of an Asian Framework of Language Testing. One is to refer to the Common European Framework of Reference as the Asia TEFL did. Or another is to follow a testing theory such as Bachman and Palmer's framework. Still another approach is to present our current situations on language testing from three countries and share the similarities and differences, uniqueness and commonness. It is usually said that comparing and contrasting the differences can lead to the findings of a common core. Probably, it can be our framework.

In my presentation, I will mainly talk about the following testing issues in Japan: 1) admissions tests and placement tests, 2) the social uses of tests in general, 3) the influence of standardized testing, and 4) testing in the domain of commerce. Also, I will consider the constraints and drawbacks as well as the benefits of an Asian Framework. I will conclude my talk by having the audience consider what specific roles an Asian language testing framework has, and how our co-operation can contribute to creating a better structure of language testing in Asia.

References

- Bachman, L.F. & Palmer, A.S. (1996). *Language Testing in Practice*. Oxford University Press.
- Davies, A., Hamp-Lyons, L. & Kemp, C. (2003). 'Whose norms? International proficiency tests in English?' *World Englishes* 22,4,571-84.
- Jenkins, J. (2006). The spread of EIL: a testing time for testers. *ELT Journal*, 60,1, 43.
- Notturmo, M. A. (ed.) (1994) *The Myth of the Framework*. London: Routledge.
- Proshina, Z. G. (2006). Conference Reviews—3rd Asia TEFL International Conference: TEFL for Asia: Unity within Diversity. *Asian Englishes*, 8,2, 78-81.
- Spolsky, B. (1985) What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing* 2: 180-192.
- Spolsky, B. (2005). Towards a common Asian framework of reference of reference for English learning, teaching and assessment. Special invited speech at the 3rd Asia TEFL International Conference in Beijing, November 2005.

Research Proposals (研究発表)

[1] A Study of Face Validity in a High School English Grammar

Test: Inter-personal Authenticity and Affective Responses Made by Test-takers

山下 次郎 (東京都立国分寺高等学校)

本研究は、高等学校で行われる定期考査の問題に関する Face Validity (受験者の情意的反応及び言語使用実感) を検証することで、どんな問題が高い Face Validity を持つのかを考察し、日々の教育活動の要の一つであるテスト開発の一助とすることを目的とする。2005 年 5 月に「英語グラマー定期考査問題」と質問紙 (各問題に対する受験者の「好み」と「言語使用実感」を調査) を実施し、以下の観点から、量的及び質的分析・検証を行った (n=324 人)。

- (1) 好まれる問題、「好み」と難易度との相関
- (2) 「言語使用実感」が強い問題、言語使用実感と難易度との相関
- (3) 面白さを感じる問題
- (4) 質問紙で最も面白さを感じるとされた問題の特徴
- (5) 2 番目に面白さを感じるとされた問題の特徴
- (6) 受験者の「好み」に関する因子分析
- (7) 「言語使用実感」に関する因子分析

分析の結果、難易度と「好み」や「言語使用実感」との相関は低い、「言語使用実感」、「好み」、「面白さ」を感じる問題はほぼ共通し、受験者が得意とする問題形式の Face Validity が必ずしも高くないことが判明した。今後は、日々行われているコミュニケーション重視の英語教育を忠実に反映したテストを作りが望まれる。それによって、コミュニケーション能力を測定できる態勢が整い、授業とテストとの両輪で生徒へプラスの波及効果を及ぼしていくことが期待できる。

[2] 清泉アカデミック Can-Do 尺度の開発とその評価

長沼 君主 (清泉女子大学)

宮嶋 万里子 (清泉女子大学)

Can-Do Statements (CDS) には Common European Framework (CEFR) に見られるような日常的な言語行動を記述した BICS 的なものと、TOEFL (iBT) に見られるようなアカデミックな場面での言語行動を記述した CALP 的なものとがある。最近では TOEFL (iBT) に続いて、新 TOEIC テストでもテストスコア解釈のための Score Descriptor としての CDS の利用が行われるようになってきているが、CEFR におけるような到達段階の参照基準としての利用も有効であろう。

そこで本研究では清泉女子大学で開発されたアカデミック Can-Do 尺度とその評価について、実際の調査データおよびプレイスメントテストとの関連をもとに考察を行う。清泉 CDS は 4 技能、各 5 項目からなり、それぞれの項目で 4 段階の到達段階により評価を行っている。調査は 1 年入学時、1 年次末、2 年次末の学生を対象として行われ、能力発達段階ごとの変化が分析された。

【3】スピーキングテストの種類が持つ波及効果研究

正頭 英和（関西大学大学院博士課程前期課程）

実践的なコミュニケーション力が注目されている英語教育において、スピーキングの評価のあり方は非常に注目を集めており、スピーキングテストはその一つの方法として注目されている。しかし、スピーキングテストは実施コストが高いため、学校という環境下では回数を多く行うことができないという現状がある。そのためスピーキングテストの持つ波及効果を活用することは非常に意義のあることだと考えられる。本研究の目的は、スピーキングテストが学習者の学習に与える影響を調査することである。そのような影響の中でも、特にスピーキングテストの形式に注目した。本研究においては、スピーキングテストの形式をダイアログとモノログの2種類とし、ダイアログテストが学習者の学習に与える影響と、モノログテストが学習者の学習に与える影響を調査した。分析方法には質問紙とインタビュー、ノートテイキングの分析を用い、量と質の両方の側面から分析を行い、それぞれのテストが学習者に与える波及効果について検討する。発表時にはこれらの分析の結果を報告し、スピーキングテストをより効率よく行うための1つの示唆を提示する。

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

周 育佳（東京外国語大学大学院博士課程）

スピーキング能力を測る方法として、インタビューのような直接テストと、媒介（テープ、電話など）を必要とする半直接テストがある。近年パソコンの急速普及につれ、実施の容易さという面から、パソコンによるスピーキング・テストの使用が増加している。対面式からパソコンによる提示モードの変化は、受験者のパフォーマンスにどのような影響を与えるかを明らかにすることは、パソコンによるスピーキング・テストのスコア解釈に大きい意味を持っているといえよう。

本研究は、パソコンによるスピーキング・テストとインタビューのスコアの関係とパフォーマンス指標における違いを解明することを目的とする。参加者は日本人高校生 20 名である。パソコンテストは、GTEC for students のスピーキング部分を使用し、インタビュー・テストには、GTEC のタスクと難易度と内容面で対等のタスクを実施する。それぞれのテストは、四つのタスクタイプ（oral reading, sentence repetition, telling story と expression opinion）を含む。本発表は、両テストスコアの比較、また発話の正確さ、複雑さ、流暢さ、自己修正などの指標の比較の結果について報告する。

[5] Predicting Reading Comprehension Item Difficulty with Respect to Content Characteristics

Kiyomi Yoshizawa (Kansai University)

Soo im Lee (Ryukoku University)

In some educational settings, item analyses are not plausible prior to the administration of a particular test due to test security. In those settings, it is up to each item writer or test designer to estimate how difficult each item is to their relevant test takers. The present study aims to provide information to those people involved in test making processes concerning the relationship between reading comprehension item difficulty and content characteristics. First, a coding system was developed based on the findings of previous studies (e.g., Feedle & Kostin, 1993; Buck, Tatsuoaka, and Kostin, 1997; Alderson, 2000; MacMillan, 2006) to examine the relationship between item and text characteristics. Next, the coding system was used to analyze 40 reading comprehension items in two official TOEIC practice tests with respect to item and text characteristics. Then, the relationship between the content characteristics and the level of item difficulty was examined. To this end, the two practice tests were administered at two institutions and the total number of test-takers was 381. Their responses were analyzed using Rasch Measurement Model and items and persons with outfit indices exceeding the acceptable range were deleted from the subsequent analyses. One of the findings is that test-takers found items difficult when the items can be identified by multiple characteristics. Another example is when test-takers need to make inferences about the people in the prompt because it is not explicitly written.

The results of the analyses with examples and practical implications will be presented in the presentation.

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

Mark Chapman (Hokkaido University)

Recent assessment literature from the general education field (Macellan, 2001; Lane, Parke, & Stone, 2002) suggests that learners and teachers can have very different opinions of how testing can impact learning. While teachers commonly believe that testing drives learning, students have presented a different picture with the washback effect of tests being less direct than claimed by teachers. This paper aims to explore the differing views of teachers and learners with regard to the impact of testing on learning. The main focus is how the TOEIC impacts language learners and learning, especially learning strategies, motivation for learning and test anxiety.

The author employed a qualitative approach as recommended by Ushioda (1994, 2001) and interviewed 11 teachers and 15 learners. All of the learners were employees of a major Japanese company enrolled in intensive language courses. The teachers all had a minimum of 18 months experience of teaching business students who regularly took the TOEIC. The interviews were conducted in the first language of the interviewees; English in the case of the teachers and Japanese for the learners with the learner interviews being conducted with the assistance of an interpreter. The interviews were transcribed (and translated into English for the learner interviews) before being analyzed quantitatively and qualitatively.

The presentation will outline similarities and differences in how teachers and learners believe the TOEIC impacts language learning. The main findings indicate that teachers need to exercise caution in their sometimes misguided beliefs about the degree of impact of testing on learning.

[7] Meta-Validation of Institutional TOEIC Research in Japan

Steven J. Ross (Kwansei Gakuin University)

Hiromu Yamada (TOEIC Steering Committee)

TOEIC validity research within Japanese institutions has increased in recent years. Many analyses of TOEIC validity involve empirical research designed to test explicit hypotheses about phenomena such as covariance between test scores and criterion outcomes, score changes over time, and convergent evidence between objective and subjective assessments of proficiency. As most data analysis methods entail distributional and internal consistency assumptions, evidence of validity needs to be examined in light of these assumptions. Otherwise the results may lead to serious inferential errors about test validity.

The present research will critically examine specific examples of TOEIC validity research in three broad areas: correlation evidence between predictors and criterion measures; Conditions under which change scores can be interpreted; Criteria for triangulation of objective and subjective evidence of validity. Emphasis will be placed on examining the statistical assumptions employed for each of the three types of validity research. For the analysis of criterion-related validity, an examination of the effects of restriction of range and attenuation of both predictor and criterion measures will be presented; In change score analysis, the assumption of continuous time on task will be examined in light of apparent discontinuity of gains; For the triangulation issue, data exhibiting a regression model used to compare the relative predictive power of objective and subjective assessments of achievement will be demonstrated.

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

秋山 實（合資会社eラーニングサービス）

井川 悦朗（NPO英語運用能力評価協会）

白戸 治久（NPO英語運用能力評価協会）

近年普及してきたeラーニングプラットフォームには、オンラインテスト機能が備わっており、手軽にテストを作成・実施し、結果を分析することができる。しかし、プレースメントテスト等のテストでは、1) 受験者数は千人規模になる、2) 一斉に実施する必要がある、3) 実施日時が限られており、確実に実施する必要がある、4) 全員同じ状態で表示・再生できなければならない、5) 音声、映像、アニメーションの制御（音量・再生・一時停止）や回答方法のバリエーションが要求される、等の条件を満たす必要がある。筆者らは、上記の条件のうち2) から5) の条件を完全にクリアする新しい方式を考案・実装し、試用した。その結果、a) テスト実施中の機器のトラブルに対しても、再開が容易であること、b) 受験者の誤操作に対してロバストなシステムであること、c) 今まで実現できなかった語順整列問題などが使用できる、など十分に満足のいく結果が得られたので報告する。また、条件の1) については、テストセンターでの利用を考慮して、10,000 人同時受験可能な設計としたバージョンを開発し、シミュレーションした結果を報告する。

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

土方 裕子（筑波大学大学院博士課程、日本学術振興会）

リスニングテストの困難度は、多くの要因に影響を受けるとされている (e.g., Brindley & Slatley, 2002)。Buck and Tatsuoka (1998) は「解答に必要な情報 (necessary information; NI)」を “the information in the text which the listener must understand to be certain of the correct answer” と定義し、NI の位置と特性が項目難易度を決定する主要な要因であると主張した。本研究はこれらの先行研究や、聴解中の情報統合プロセスに関するモデルを踏まえ、解答に必要な情報の範囲が英検の項目困難度を与える影響を明らかにすることを目的とする。具体的には (1) 英検問題における NI の特徴はどのようなものか、(2) NI と得点率にはどのような関係があるか、(3) NI と項目弁別度にはどのような関係があるか、の3点をリサーチ・クエスチョンとする。これらのリサーチ・クエスチョンをカイ2乗検定、クラスター分析、項目弁別力の算出などを用いて検証する。

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

庄司 恵雄 (お茶の水女子大学)

野口 裕之 (名古屋大学)

春原 憲一郎 (海外技術者研修協会)

財団法人技術研修協会AOTSでは、年間約五千名以上の技術研修生を主としてアジア各地から受け入れ、150時間の本研修前日本語教育を国内4か所のセンターにおいて実施している。この予備教育の成果を測定する方法として、従来、語彙・文法・聴解に関する全国統一基準による研修後試験が実施されてきたが、口頭能力については統一テストがなく、センターごとに独自にテストが実施されてきたため、全国統一基準による口頭能力試験を開発することになった。そのプリテストを2004年度から2005年度にかけて148名（7か国）に対して実施したところ、2006年度本実施への見通しがつく評価者間信頼性係数が得られた（庄司・野口・春原・石沢：2006）。しかしながら、本実施にあたっては、さらに評価表の構成について改善を図ることによって、いっそう高い信頼性を確保する必要がある。本発表では、プリテストにおいて相対的に評価者間信頼性係数の低かった課題について、その対応する評価表を再検討した結果を報告する。

パイロット研究に使用した発話テストは、録画済みのビデオテープから提示される課題に反応させる方式で、庄司ほか(2003、2004)を参照し、質問に応答するなど6類の課題で構成した。各課題は相互に独立しており、先行課題の結果が後続課題の遂行に影響を及ぼさないよう設計されている。

評価は、解答の量的側面を測定するチェック・リスト法と、質的側面を測定する査定基準参照による程度評価法を併用し、被験者の上位群・下位群それぞれ2名ずつを抽出して同協会所属教師がランダムに2名ずつのペアに分かれて評定を行った結果、0.8台の評定者間相関係数が得られた。しかしながら、評価表を個別に検討すると、チェック・リストによる評定に比較し、査定基準参照による程度評定では判定に「ゆれ」を生じる部分があることが判明した。この原因として、課題の質・評価表の構成、両面の問題点が予測されるが、本発表では、評価表を取り上げ、評価者×被験者の評定データを分散分析モデルを適用して分析、評価者間信頼性を再度精査するとともに、本実施における評定者の数についても言及したい。

[11] 海外短期英語研修の口頭英語面における効果と測定

横川 博一（神戸大学）

鈴木 正紀（Ordinate Corporation）

海外短期語学研修（以下「研修」）に参加する学生の多くは、「英語をたくさん聞いて話せるようになりたい」という漠然とした目的・期待を持っている。しかし、その効果は TOEIC や TOEFL などによって測定されることが多く、口頭による英語能力がどの程度向上したか、直接測定されることは少ない。

神戸大学とオーディネート社では、オーディネート社の「口頭英語能力自動試験 SET-10」を用い、オーラル面における研修の効果の測定を試みた。SET-10 は、口頭による英語能力を直接測定するもので、電話で受験し、自動採点される。また、総合点と文章構文、語彙、流暢さ、発音のサブスコアが 20～80 点の間で報告される。

この試験を、豪州・メルボルン大学において約 3 週間の研修に参加した 35 名の学生を対象に、研修の前後に実施した。その結果、総合点 (36.3→38.9)、文章構文(35.1→38.1)、流暢さ(37.9→41.1)において統計的に有意な伸長が見られ、短期間でもオーラル面の能力向上が期待できることが明らかになった。

本発表では、研修前後の得点の推移などを示すとともに、通常の教室環境での授業実践における口頭英語能力との比較などを行いながら、海外短期英語研修の口頭英語面における効果をどう測定するかについて議論する。

[12] テスト改訂における妥当性検証—新 TOEIC フィールドスタディー結果

三橋 峰夫（財団法人国際ビジネスコミュニケーション協会）

様々なテストにおいて、内容の改訂はしばしば行われている。理由としては、最新のテスト理論やテクノロジーの導入であったり、測られるべき内容が時間の経過と共にテスト開発時とは変化したためそれらを是正するといったことが考えられる。しかしながら理由の如何に係わらず、重要なことは改訂後のテストが改訂の目的や意図に合ったものになっているかを検証することである。これが行われないのであれば、何のためのテスト改訂であったのかが不透明なものとなる。

2006 年 5 月 TOEIC テストが、”More Authentic”というコンセプトの下リニューアルされた。それに先立ち、開発元である Educational Testing Service (ETS) 主導による Pilot Study が実施され、新・旧バージョン間における比較検証調査が行われた。

本発表ではその調査結果を中心に、新たに取り入れられたテスト作成における手法 ”Evidence-centered design”や、各設問ごとの Item analysis の実例について紹介する。

またテスト作成者側の意図とは別に、受験者が今回の改訂についてどのように評価しているかを調査するために行ったアンケートの結果も新 TOEIC テストの受験者データ（平均点、標準偏差等）等と併せて紹介する。

[13] リスニング・リコールの採点について—アイデア・ユニットを 分析単位にした場合—

飯村 英樹（常磐大学）

リコール（筆記再生法）は、学習者の読解力や聴解力を測定するための手法である。リスニング・リコールの場合、学習者に聞き取り終了後、テキストの内容を筆記再生させ、その内容を何らかの分析単位を用いて分割したテキストと照らし合わせて、それぞれの箇所が正しく再生されているかを採点の基準とする。

採点の際に使用される分析単位は様々であるが、アイデア・ユニットを用いた研究が最も多い。しかしアイデア・ユニットを用いた研究であっても、その分析単位の定義及び採点の手続きは定まっていない。

本研究では、4つのアイデア・ユニット — 1. Spencer' s system (reported in Meyer, 1975), 2. Meyer (1975), 3. Carrell (1985), 4. Sakai (2005) — に従って日本人大学生を対象に実施したリコール・テストを採点し、その結果を多肢選択式英語能力判定テスト（CASEC）との相関分析を通して、それぞれのアイデア・ユニットの妥当性について検証する。

[14] 並べ替え問題の部分採点法

安間 一雄（玉川大学）

言語テストにおける形式の 1 つである並べ替え問題（語彙・文・パラグラフ等）においては全問正解でない限り合理的な部分採点法は知られていなかった。広範に行われている全問一致法では正解の連鎖のみに得点を与え、それ以外のバリエーションはすべて不正解とするため、例えば正解が A-B-C-D-E の連鎖で解答が (i) E-A-B-C-D であるとき、最後の項目の位置が間違っているだけにもかかわらず、すべてを逆順にした解答 (ii) E-D-C-B-A と同様に 0 点とされてしまう。合理的な部分採点法を開発することにより、相対的に正解に近い (i) を、より遠い (ii) よりも高く評価し、評価の妥当性を高めることができる。今回、さまざまな形で提案されている部分採点法を吟味した結果、正解からの項目の移動回数分のペナルティーを与える方法が認知上合理的でかつ直観と合致することがわかった。この方法は人工知能分野における「最小項目移動」に相当する。しかしながら移動が重なるにつれ、どのように移動したら正解に辿り着くのかを計算することは困難になる。そこで、与えられた解答において正解の連鎖と同じベクトルを有する項目連鎖を最大にする項目の組み合わせを見つける「項目連鎖最大化法」を考案し、これが最小項目移動と等価であることを示した。この方法を用いれば目算で部分点を一意に計算できる。

[15] 学習者の語彙サイズはテストによってかわるのか

— 京都大学での事例研究結果からの考察 —

笹尾 洋介 (京都大学大学院修士課程)

語彙はコミュニケーション能力の基盤であり、語彙の大きさが学業成績、リーディング、ライティング、その他一般言語能力を予測できると考えられている。これまで語彙サイズを測るための各種のテストが考案され実施されてきたが、テストの違いによりテスト間で語彙サイズの推定が異なるという主張がある。しかしながら、この点に関しての実証研究はいまだ乏しい。

そこで本研究は、テストによってどの程度語彙サイズが異なるのかについて、実証データで示すことを目的とする。具体的には、京都大学1回生126名を対象として、『JACET 8000』を2000語ごとに四段階のレベルに区分し、各レベルから無作為に30語ずつ、合計120語を選出し、異なる四種類のテストにより被験者の語彙サイズを測定した。研究方法はLaufer and Goldstein (2004)を参考にし、発表語彙、受容語彙のそれぞれについて、選択式、記述式の二種類により被験者の語彙サイズを測定した。推定語彙サイズは基準とした『JACET 8000』の8000語に、テストした120語の正答率をかけて算出する。この調査の成果を報告する。

Laufer, B. and Z. Goldstein. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning* 54, pp.399-436.

大学英語教育学会基本語改訂委員会 編. (2003). 『大学英語教育学会基本語リスト JACET List of 8000 Basic Words』 (通称 JACET8000).

[16] ライティングにおける産出語彙の利用可能性: より良い評価へ 貢献できるために

占部昌蔵（兵庫教育大学大学院修士課程）

語彙力はライティング能力の一部と考えられ、ライティングにおいて大きな役割を果たすと考えられている。また、ESL 学習者のライティング評価は、ライティングのある種の語彙分析と相関が高いとも言われている。しかし、語彙が、ライティングの評価にどの程度影響しているのかについては、あまり多くは調べられてきていない。本研究では、①「評価者の下した判断は、産出語彙から見てどの程度妥当なのか」という点と、②「産出語彙からライティング能力を予測することがどの程度可能なのか」という点を明らかにすることを目的とする。

参加者は約 20 名の日本人英語教師（中学・高等学校）で、100 名の日本人英語学習者（高校 2 年生）が受けた自由英作文のテストを評価した。評価規準は、ESL Composition Profile を使用した。実際に書かれた語彙と評価の語彙パートの点数とにどの程度の関係があるのかを分析する予定で、その際に、項目応答理論を使用して、評価者の中で信頼性の高いグループとそうでないグループに分け、それぞれのグループと客観的語彙指標とのあいだにどの程度の相関があるのかを分析する予定である。

[17] IRT, testing, and online technology: the future of autonomous vocabulary learning?

Brent Culligan (Aoyama Women's Junior College)

Charles Browne (Aoyama Gakuin University)

How many words do our students know? Is it possible to accurately identify the words that each learner knows and which ones should be studied next? How can online technology be used to test and teach these words? We will discuss both the theoretical underpinnings and practical applications of the online testing and teaching of vocabulary for EFL students.

To date, most vocabulary research has utilized word frequency both as a measure of a learner's lexical size as well as the sequencing of vocabulary learning. However, this practice assumes a relationship between word knowledge and frequency.

Working from the findings of a large-scale testing project of EFL learners in Japan, which applied Item Response Theory to lexical decision tasks, we obtained difficulty estimates for over 5000 of the most frequently used words in the English language.

Three observations emerged. First, we found that vocabulary frequency was not a good predictor of vocabulary difficulty. Even students with large vocabulary size (3000+ words) had large gaps in their knowledge of high frequency words. This suggests that efficient language teaching should help to fill in these gaps.

Second, vocabulary frequency does allow us to accurately calculate a learner's vocabulary size.

Finally, it allows us to design individualized learning programs at the learner's level, based on the most frequent words which they do not know. The widespread popularity of electronic technologies such as cell phones and PCs makes it possible to efficiently test and teach vocabulary in ways that until now were not possible.

[18] New Ways of Testing and Teaching Vocabulary Online

John Paul Loucky (Seinan Women's University)

http://www.call4all.us//home/_all.php?fi=0

call4all@mx7.tiki.ne.jp

Various new types of vocabulary development online programs and tests are compared in this study to consider how useful they may be for guiding individual or classroom L2 vocabulary instruction. It explored how these programs seek to test and teach new L2 vocabulary more innovatively and accurately. Several vocabulary programs are reviewed, compared and combined within a comprehensive, recycling model of second language vocabulary acquisition (SLVA), in an attempt to help maximize effectiveness in teaching and testing.

This presentation will explore how some new online programs seek to test and teach new L2 vocabulary development more innovatively and accurately. Several vocabulary programs will be reviewed, compared and combined within a comprehensive model of SLVA, in an attempt to help maximize effectiveness in teaching and testing second language vocabulary acquisition. These include:

Key Vocabulary Constructs Tested:

- 1) V-Check-Vocabulary Size assessed by Accuracy of Recognition
- 2) DAVIE Vocabulary Knowledge Scale-Variety Aspects, but mainly assessed by Accuracy of Production including
 - a) L1 Translation Ability
 - b) L2 Defining Ability
 - c) Collocational Recall
 - d) Productive Use in Whole Sentences, showing
 1. Semantic Accuracy of Meaning in Context
 2. Syntactical Accuracy of Grammatical Structure
- 3) Semantic Field Keyword Quizzes-Degree of Structure
- 4) Word Champ-Monitoring Online Vocabulary Learning Using
 - a) Student-made Flashcards
 - b) Teacher/Class-made Flashcards
 - c) Computer-Generated Tests of these
 - d) Using CMS to Monitor Learning in Detail
 - e) Adding Links to Related Texts for Contextualization
- 5) Lexxi-Study's Online Vocabulary and Language Learning System
 - a) Basic Word Meanings
 - b) Generate Word Lists and Tests from any Text
 - c) Combining Common Collocations, Corpora and Concordancers
 - d) Tracking Student/Teacher's Writing and Reading Texts

Various vocabulary development programs and tests will be presented, and their different stages and pedagogical use demonstrated. How useful they can be for guiding individual or classroom vocabulary instruction will also be compared.

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

吉池 陽子 (ARCLE 研究員)

根岸 雅史 (東京外国語大学大学院)

沓澤 糸 ((株)ベネッセコーポレーション)

近年、学習者の言語能力を記述する方法として、実際に英語を使用する場面で「できること」を具体的に記述しようとする Can-do statements が注目されている。高校生の英語力を測るテストの一つである、ベネッセコーポレーションの GTEC for STUDENTS では、テストスコアとあわせて、Can-do statements を学習者にフィードバックしている。これらの Can-do statements 開発は、学習者の自己申告による「できること」とテストのスコアの関連を見ることが多い。しかし、このような Can-do statements はあくまでも自己申告であるため、その妥当性を疑う声も多い。そこで本研究は、実際の学習者のパフォーマンスに着目して、既存の Can-do statements の改良を試みた。今回は学習者のリスニング能力に着目し、英語によるラジオ、その中でも、日本の学習者が実際に体験可能なインターネットラジオに着目して調査を行った。調査では、日本の高校生を対象に、インターネットラジオを利用したタスク形式の調査を行い、高校生の英語能力を示す指標としてベネッセコーポレーションによる GTEC for STUDENTS の結果を使用した。調査の結果、実際のパフォーマンスに着目したことで、日本の高校生の実態に即した Can-do statements へと改良する手がかりが得られたと考える。

[20] 英検 Can-do 調査報告

柳瀬 和明 ((財)日本英語検定協会)

松平 知樹 ((財)日本英語検定協会)

(財)日本英語検定協会は実用英語技能検定(以下、英検)を7つの級で実施しており、毎年非常に多くの英語学習者が受験している。各級における「典型的な合格者像」を探るため、合格者を対象とした「英語を使ってどのようなことができる自信があるか」ということについての調査を実施し、それに基づいて英検 Can-do リストを作成した。

調査はアンケート形式で、2003 年から 2005 年まで 3 回行い、有効回答者数は延べ 20,000 人を超えた。各級の合格者(合格直後)に 4 技能別に作成した能力記述文を掲載した調査冊子を送付し、各項目に対して自信の度合いを 5 段階で回答してもらった。

英検の 7 つの級は学習初期段階から上級まで幅広くカバーしているため、調査冊子を段階別に 5 セットに分けた。隣接する調査冊子セットにアンカー項目を配置し、段階反応モデルによってすべての能力記述文を等化した。これらの結果を踏まえ、各項目の 5 段階の回答状況等を考慮して各級のカットラインを判断した。

課題として残された点もあるが、今回の結果が英語学習者、英語教育関係者に貢献できる点も少なくない考える。

[21] 日韓大学生の英語「方略能力」に関する一考察

達 川 奎 三 (広島大学)

国の内外を問わず英語教育界は、コミュニケーション能力育成のためにあらゆる領域・分野で指導と評価の工夫・改善を試みている。「コミュニケーション能力」の「下位構成素」の中で、学校教育においてあまり力点を置いて指導されていなかったものに「方略能力」がある。コミュニケーションが挫折・頓挫した場合に必要となるのはそれらを「修復」する能力であり、また、そのような事態に陥らぬよう「予防」する能力であろう。本研究では、Celce-Murcia, et al. (1995) を視座にした英語学習者の「方略能力」に関する診断テストの開発を提案する。さらに、大学生(日本人 245 人、韓国人 200 人)にこの診断テストに取り組んでもらい、その結果を分析・考察することによって学習者の英語方略能力を概観し、教育的示唆を得ようとするものである。調査結果から日・韓両国の大学生は英語での「話題や談話を途中で回避したり、取り下げたりする」場面や、「対話者に対して理解の可否を自ら確かめる」ことに慣れていないことが分かった。そして、様々なコミュニケーション上の障害を乗り越える知識は備えているが、自ら言い換えたり、確認のための応答をすることによってコミュニケーションの挫折を前もって積極的に回避する方略を身につける必要性を確認できた。また、言語切り換えの方略においては母語の違いによる顕著な影響が見られた。

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

山中 司 (北陸大学非常勤講師、慶應義塾大学大学院博士課程)

鈴木 佑治 (慶應義塾大学)

本研究は日本の言語・外国語教育、特に英語教育において、次世代のコミュニケーションの変容を展望し、新たな言語コミュニケーションの理論研究と、その具現としての「アウトカム」を評価するモデル構築、そしてそのグローバルな提示の必要性を広く訴えるものである。本研究を叩き台とし、「コミュニケーション」というものを正面から研究する、新たな研究が試みられることを強く提言したい。

上記の目的を実現するために、本研究は大きく2つの視点に分けて議論を行う。第1の論点は、変貌する次世代のコミュニケーション論の重要性を議論し、その意義を明らかにした上で、言語教育という文脈の中での「コミュニケーションとは何か」という根本的な問題を議論する。第2の論点は、その応用・実践事例として、英語教育における「評価の問題」を取り上げる。「広義な」コミュニケーション論を包含した評価モデルを構築するため、筆者による新たな英語評価モデルの提示を行う。これによって発信の側面や、言語以外における多メディア・多感覚的に訴える総合的な表現についての評価を提言する。不完全ながらも新たなモデルを提示することで、今後のワーキング・グループ等の結成を促す。

[23] The Effectiveness of a 15-minute Writing Sample for Assessing Writing Ability

Melvin R. Andrade (Sophia Junior College)

This paper describes action research involving the preparation of a test of English writing ability to be included in a placement examination. The examination, which included tests of reading, grammar, vocabulary, and error recognition in addition to writing, needed to be limited to one hour to fit the examination day schedule. Based on the observation of test takers in previous years--primarily variation in reading speed--it was concluded that at least 45 minutes would be needed for the non-writing portions of the test. Thus, 15 minutes was allocated for writing. Another factor was that the writing test had to be short enough to be scored quickly by a limited number of scorers. With these points in mind, the research questions became the following: (1) Can a 15-minute test of writing ability tell us enough about the students' ability to distinguish between high, medium, and low ability writers? (2) Can such a short test provide information about writing ability beyond sentence level grammar and syntax? That is, can it indicate, for example, coherence, unity, and organizational competence? To answer, these and related questions, a pilot study was conducted with four groups of Japanese learners of English in Japan: two groups of second-year university students (English majors and non-majors), and one group each of first- and second year junior college students (English majors). It was predicted that there would be significant differences between first- and second-year English major students, between junior college and university students, and between English major and non-English major students. The paper discusses the complexity of the findings and the factors that contributed to the decision to modify the time limit for the writing test.

[24] Critical language testing: The General English Proficiency Test

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

Shohamy (1998, 2001) proposed her theory of critical language testing, asserting that test takers should be entitled to question the test itself, its value and its methods. She believed that testers could benefit from the valuable insights of test takers. Although Shohamy has proposed her theory of critical language testing for almost ten years, very little research has focused on this subject. Therefore, I investigated test takers' views of the General English Proficiency Test¹ (GEPT) and its uses for gate-keeping purposes by interviewing 29 test takers at two Taiwanese institutes of technology. Surprisingly, a majority of participants agreed that the GEPT should be used as one of their degree requirements. In addition, with their firsthand experience, participants suggested that the Language Training and Testing Center, which develops the GEPT, should administer listening comprehension questions in a different manner, develop a more authentic speaking test, eliminate the noises uttered by other test takers on their surrounding examinees in the speaking test, roll back the registration fees, do a facelift of the answer sheets, provide test centers at well-known locations, endeavor to acquire international recognition for its certificates, and administer the GEPT more frequently. In summation, research findings showed that test takers could view the test critically and provide critical insights into the administration of the test. Therefore, their views of the test and its uses should be valued by testers.

¹ The GEPT, a Taiwanese English proficiency test which debuted in 2000, has been taken for 1.7 million times in Taiwan and has been adopted by some Taiwanese schools as one of their degree requirements.

[25] I R Tを利用したテストの改訂

藤田 智子 (東海大学・東京工業大学大学院)

前川 眞一 (東京工業大学大学院)

ある大学英語プログラムの約6000人の新入生を、上中下の英語習熟度別レベルに分ける為のプレースメントテストを、I R Tを利用してより妥当なテストに改訂するための試みを報告する。

I R Tを利用した第一回目テスト改訂は、項目分析を行い、項目困難度の高すぎる問題(3.00)を削除した。また、そこに不足していた下のレベル切り点に近い困難度(-1.0)の項目と、弁別力の高い項目をパイロットテストの項目から優先的に補った。しかしこれは暫定的な改訂であり、テスト情報曲線が上下の切り点で高くなる(精度が高くなる)という目標とする形に近づけることはできなかった。

妥当な自前のプレースメントテストを作るには、被験者がそのプログラムでどのようなアカデミックパフォーマンスを発揮できるかを予想するための項目を非常に多く作成する必要がある。そこで今回はこの大学英語プログラムで学期毎に作成し、期末成績判定のために実施していた英語統一テストを、テストの等化によって、大幅な作問作業をすることなく、プレースメントテスト改訂に利用する計画である。それぞれのテストの5～6項目ずつを共通被験者にパイロットテストとして受けてもらい、共通項目デザインを利用した項目パラメタの等化を行う予定である。

[26] テスト自動管理・作成プログラム（Item Banking System）の 開発

平井 明代 （筑波大学）

小学・中学・高校・大学のどのレベルの教員であっても、生徒や学生に試験問題を作成し、実施する。過去に使った問題の中で、再度使用したい良い問題、別のところから見つけた将来使えそうな問題など、整理して保管しておきたい。そして、そこから問題を組み合わせて、簡単にクラスルームベースの小テストや学年全体の模擬試験や期末試験などを簡単に作り、実施することができれば時間的効率もよくなる。また、どのような問題を選択し組み合わせてテストにしたかを一目で確認することができれば、バランスよく出題することもできる。

筑波大学では、少しでもテスト作成の実用性を上げるために、毎年2回実施している筑波英語検定試験用に、項目応答理論を使って、すべての問題を同一尺度で測定した問題を簡単に保管し、そこから複数のテストフォームを自動的に作成するコンピュータプログラムを MS-ACCESS データベースソフトを使って開発し、使用している。

本研究では、この筑波英語検定試験用に開発したプログラムを基に、一般教員向けに、誰でも自分の問題を簡単に管理し、そこから自動的に問題を組み合わせてペーパーテストを作成するプログラムを開発している。発表時点で完成に近づきつつあるプログラムの紹介および、筑波英語検定試験での使用実践を報告する。

[27] 英語関係節の習得研究における可變的研究結果と文法性判断テスト

伊藤 彰浩 (愛知学院大学)

1980年代から行われている英語関係節の習得研究では、文法性判断テストが誘出法のひとつとして用いられている。文法性判断テストは他のテスト方法とは異なる研究結果を導くことが多い。これを根拠に異質的言語能力観を提唱する研究者がいる。その一方で、研究結果が可變性を示す傾向は言語データを提供した被験者群の言語能力の様相を反映しているのではなく、単にテスト方法による影響だと指摘する言語テスト研究者も存在する。本研究では、初めに英語関係節の習得研究に利用されてきた典型的な文法性判断テストを例示し、その問題点を指摘する。次に、英語関係節に関する知識・能力を測定する文法性判断テストを作成する際に考慮しなくてはならない要因を列挙し、諸要因を統制した上で新しい文法性判断テストを構築する。そして、その文法性判断テストを大学生に実施し、含意的および類型論的普遍文法理論に基づく仮説「名詞句の接近度階層」の妥当性を検証する。新しい文法性判断テストは他のテスト方法と異なる研究結果を導くか、それとも同一の研究結果を導くだろうか。調査結果に考察を加えながら、第二言語習得研究と言語テスト研究の相互作用はいかにあるべきかについても検討したい。

[28] ELPA 診断テスト (文法) の妥当性検証：項目応答理論による 分析を中心に

小泉 利恵 (常磐大学)

酒井 英樹 (信州大学)

診断テストとは、受験者の長所・弱点を見つけるために使われるテストであり、日本人英語学習者に向けた診断テストは知る限りない。このような状況の中で、特定非営利活動法人 (NPO)・英語運用能力評価協会 (ELPA) は、日本人英語学習者のための診断テスト (「ELPA 診断テスト」) の開発プロジェクトを立ち上げた。ELPA 診断テストの主な受験対象者は日本人中学生・高校生であり、文法・語彙・音声セクションからなる。文法セクションは、あまり体系的には教えられていないが、読解・聴解の基礎として重要である「名詞句の把握」、特に「基本的な名詞句の境界把握と内部構造の把握能力」を測定するために作成された。本研究の目的は、文法セクションの試行版の妥当性を検証することである。様々な妥当性の検証方法 (e.g., Messick, 1996) の中で、本発表は、項目応答理論を中心に用いる。参加者は中学 2・3 年生、高校 1・2 年生の 740 人で、72 項目の多肢選択式問題を受験した。項目応答理論の分析の結果、項目と受験者は高い割合でモデルに適合しており、文法セクションについての肯定的な妥当性の証拠が得られた。

**[29] Creating pedagogic tasks based on analyses of English
university entrance examinations**

Keita Kikuchi (Waseda University)

Brian Wistner (Tokyo Junshin Women's College)

University entrance examinations are often criticized for focusing on grammar-translation tasks which create washback effects on pedagogical methods in junior high schools, high schools, and test preparation courses. Considering current task-based approaches to ESL/EFL pedagogy (Ellis, 2003; Kumaravadivelu, 2006), the validity and appropriateness of traditional grammar-translation approaches to language education in Japan, in addition to primarily testing grammatical knowledge, come into question. In this paper, in order to investigate the constraints of washback on the pedagogical options available to English teachers in Japan, we first review the findings from two research articles that focused on prestigious university entrance examinations in Japan (Kikuchi, 2006; Brown & Yamashita, 1995). Then, the content and test methods of three private university entrance exams, three public university entrance examinations, and the center test which Kikuchi (2006) examined in his study are analyzed in order to create sound pedagogical tasks designed to prepare test takers. Communicative tasks are drawn from the research literature (Norris, Brown, Hudson, & Yoshioka, 1998), and the validity and appropriateness of implementing these tasks to prepare Japanese high school students for the entrance examinations are considered. Additionally, the findings are situated and analyzed within Bachman's (1990) framework of communicative language ability. Finally, the implications for English language education at various levels in Japan are discussed.

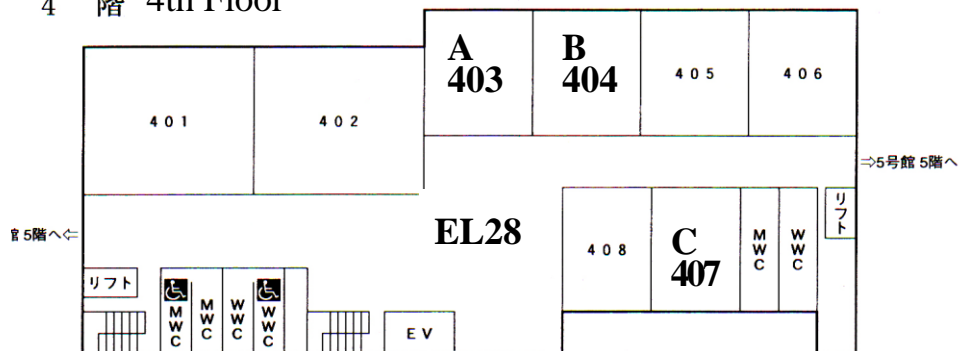
深草キャンパス 21 号館 Fukakusa Campus, Building No. 21



4 階 4th Floor

Research Proposals (研究発表)

- A: Room A 403
- B: Room B 404
- C: Room C 407
- D: Room D 503
- E: Room E 601
- F: Room F 602
- G: Room G 603



REC: Reception (受付)

CM: Committee Meetings (役員会) 506

OC: Opening Ceremony (開会行事) 603

S: Symposium (シンポジウム) 603

PL: Plenary Speech (講演) 603

GBM: General Business Meeting 603

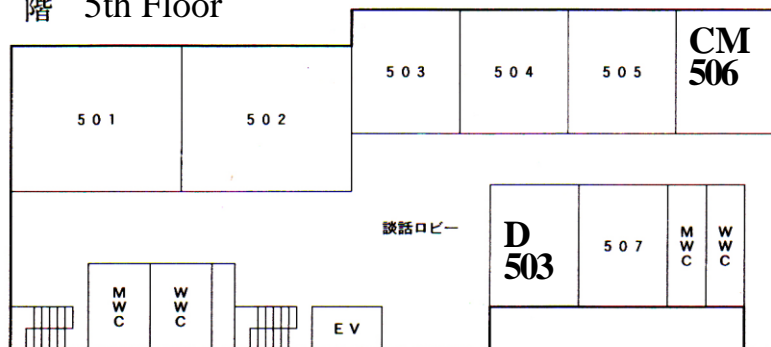
(総会) 603

CC: Closing Ceremony (閉会行事) 603

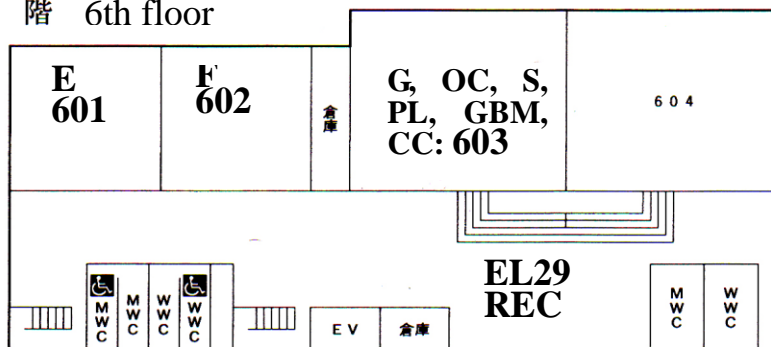
EL28: Exhibition Lobby for October 28 (Sat) (10月28日(土)展示ロビー)

EL29: Exhibition Lobby for October 29 (Sun)

5 階 5th Floor



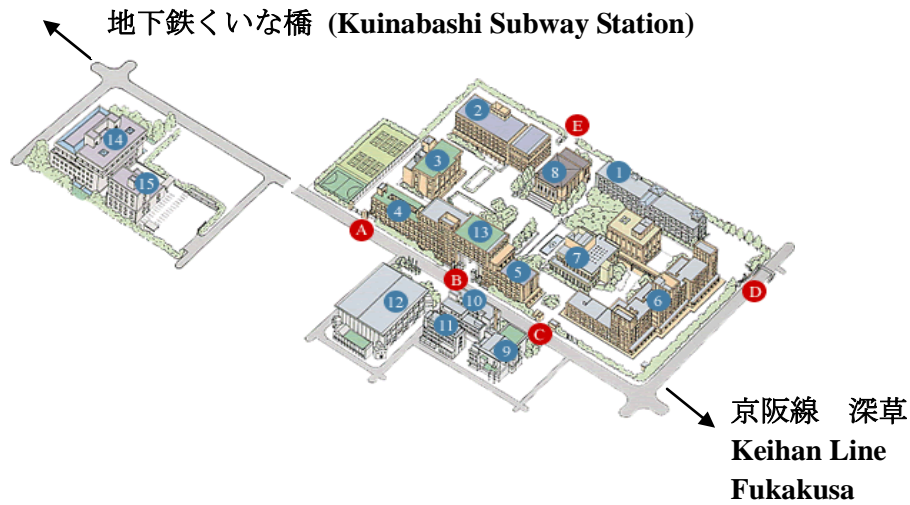
6 階 6th floor



* 飲み物は 4 号館地下の自動販売機をご利用下さい。

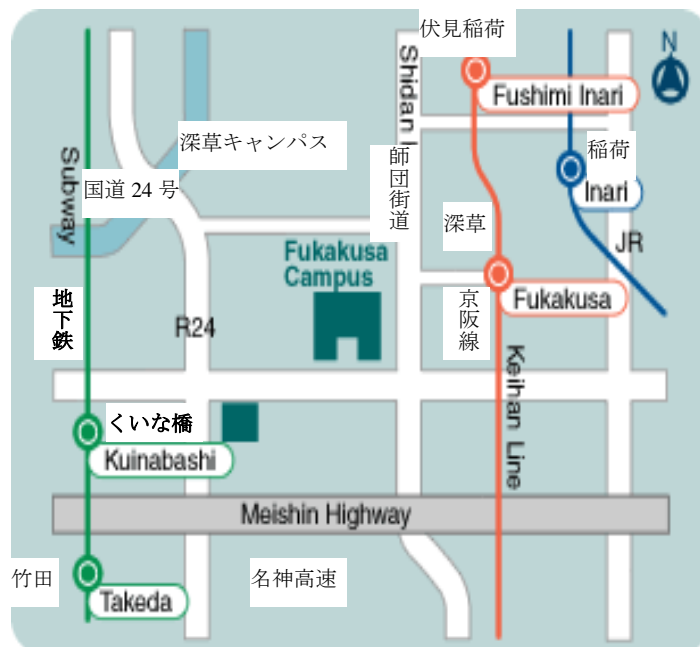
Drinks are available from the vending machines in the basement of Building No. 4

深草キャンパス (Fukakusa Campus)



- ① 1号館 Building No.1 ② 2号館 Building No.2 ③ 3号館 Building No.3
 ④ 4号館 Building No.4 (グリル Dining Hall) ⑤ 5号館 Building No.5 ⑥ 紫英館 Shieikan
 ⑦ 図書館 Library ⑧ 顕真館 Kenshinkan ⑨ 学友会館 GakuyuKaikan (Student Hall)
 ⑩ 紫朋館 Shihokan ⑪ 紫陽館 Shiyokan ⑫ 体育館 Gymnasium
 ⑬ 21号館 Building No.21 (学会会場 Conference Venue)
 ⑭ 紫光館 Shikokan ⑮ 至心館 Shishinkan
 (A) 西門 West Gate (B) 正門 Main Gate (C) 通用門 Side Gate (D) 東門 East Gate
 (E) 北門 North Gate

アクセスマップ (Access Map)



- ・地下鉄「京都」駅から竹田方面へ「くいな橋」駅下車(所要時間 5 分)、地下鉄くいな橋駅から北東徒歩 10 分
 Kyoto (subway) → Kuinabashi (5 min) About a 10-minute walk northeast from Kuinabashi Subway Station