

日 本 言 語 テ ス ト 学 会 (JLTA)

第 7 回 全 国 研 究 大 会 (2003 年 度) プ ロ グ ラ ム

The Seventh Annual Conference

of

The Japan Language Testing Association

大会テーマ:「言語学習とテスト:テストが言語学習に果たす役割」

Language Learning and Testing: The Role of Testing for Language Learning

日時: 2003 年 10 月 25 日 (土) 8:50 ~ 17:45

会場: 熊本大学大学教育センター棟 (黒髪キャンパス北地区)
(〒860-8555 熊本市黒髪 2-40-1)

日本語テスト学会 (JLTA)
The Japan Language Testing Association (JLTA)

事務局
〒389-0813 長野県埴科郡戸倉町芝原 758
TEL 026-275-1964 FAX 026-275-1970
e-mail: youichi@avis.ne.jp
URL: <http://www.avis.ne.jp/~youichi/JLTA.html>

全 国 研 究 大 会 本 部 委 員

大友 賢二 (常磐大学・筑波大学名誉教授)

Randolph Thrasher (沖縄キリスト教短期大学・国際基督教大学名誉教授)

中村 優治 (東京経済大学)

中村 洋一 (常磐大学)

全 国 研 究 大 会 運 営 委 員

島谷 浩 (熊本大学)

法月 健 (静岡産業大学)

中村 優治 (東京経済大学)

大坪 一夫 (麗澤大学)

小山 由紀江 (名古屋工業大学)

塩川 春彦 (北海学園大学)

智原 哲郎 (大阪女学院短期大学)

Elizabeth Heiser (International Pacific College)

伊藤 彰浩 (愛知学院大学)

藤田 智子 (東海大学)

Soo-im Lee (龍谷大学)

林 孝憲 (東京経済大学非常勤)

第 7 回 大 会 プ ロ グ ラ ム

10月24日(金)

17:00 ~ 18:30 理事会 (熊本ホテルキャッスル)

10月25日(土)

8:50 ~ 受付

9:20 ~ 9:35 開会の挨拶 (C301 教室) 総合司会 島谷 浩 (熊本大学)

会長 大友 賢二 (常磐大学・筑波大学名誉教授)

9:40~12:20 研究発表 (発表 30 分, 質疑 10 分) I 9:40 ~ 10:20

II 10:20~11:00

第1室 (B201 視聴覚室) 司会 大津 敦史 (福岡大学)

① 発表 I Can-do 項目作成のための英語能力評価共通枠組みの検討

長沼 君主 (清泉女子大学)

和田 朋子 (東京外国語大学大学院)

田中 敦英 (桐朋中・高等学校)

根岸 雅史 (東京外国語大学)

② 発表 II 日本人英語学習者の英語メンタルレキシコン—感情を表す形容詞の定着度—

折田 充 (八代工業高等専門学校)

第2室 (B202 視聴覚室) 司会 Terry Laskowski (熊本大学)

③ 発表 I A Possibility of Norm-Referenced Testing for Educational Purposes in Japan – A TOEIC Case Study –

MIZUMOTO Atsushi (Graduate School of Language Education and Information Science, Ritsumeikan University)

④ 発表 II Enhancing the Learning, Teaching and Testing of New Vocabulary by Systematically Using a Depth of Lexical Processing Scale and a Vocabulary Learning Strategies Taxonomy

John Paul Loucky (Seinan Women's University)

第3室 (D201 教室)

司会 石井 和仁 (福岡大学)

- ⑤ 発表I プレイスメントテスト記述課題の量的・質的分析
西村よしみ (筑波大学留学生センター)
- ⑥ 発表II TOEIC で見る新カリキュラムの成果
吉川 勝正 (熊本学園大学)

第4室 (D202 教室)

司会 山本 廣基 (西南女学院大学)

- ⑦ 発表I 語彙知識によるスピーキング能力の予測可能性
小泉 利恵 (筑波大学大学院博士課程)
- ⑧ 発表II 日本語 Can-do-statements の IRT 尺度化
野口 裕之 (名古屋大学)
和田 晃子 (北海道大学)
熊谷 龍一 (名古屋大学)
脇田 貴文 (名古屋大学)

第5室 (D203 教室)

司会 浪田 克之介 (北海道情報大学)

- ⑨ 発表I A Vocabulary Level Test at the 1000-Word Level
Paul Westrick (Kyushu University)
J. Lake (Fukuoka Jo Gakuin and Kyushu University)
- ⑩ 発表II Examining the Rationale of Vocabulary Size Tests Based on Word-Frequency Levels
KATAGIRI Kazuhiko (Reitaku University)

11:00~11:10 休 憩

11:10~12:30 基 調 講 演 (C301 教室)

司会 木下 正義 (福岡国際大学)
紹介 会長 大友 賢二 (常磐大学)

Topic: **Interactive Impacts of English Testing, Teaching, and Learning in Korea**

Speaker: **Oryang Kwon** (Seoul National University)

12:30~13:30 昼 食 (役員会：1階第一会議室、休憩室：1階多目的会議室)

13:30～14:50 研究発表 (発表 30 分, 質疑 10 分) I 13:30～14:10
II 14:10～14:50

第 1 室 (B201 視聴覚室) 司会 **Jeff Hubbell** (法政大学)

⑪ 発表 I The Role of the TOEIC in Measuring Language Learning
Sandra V. Williams (Hiroshima Kokusai Gakuin University)

⑫ 発表 II Multiple Imputation in Latent Growth Curve Analysis
Steven Ross (Kwansei Gakuin University)

第 2 室 (B202 視聴覚室) 司会 小山 由紀江 (名古屋工業大学)

⑬ 発表 I An Investigation of Method Effects in Reading Comprehension Tests: Text Structure and Response Format
Miyoko Kobayashi (Graduate School of Language Sciences, Kanda University of International Studies)

⑭ 発表 II 実践報告—理科系の大学生に求められる英語力とテスト、およびその対策
河野 円 (星薬科大学)

第 3 室 (D201 教室) 司会 塩川 春彦 (北海学園大学)

⑮ 発表 I Introducing Exit Competencies into a University English Course
Brendan Rodda (Kanda University of International Studies)

⑯ 発表 II Standardized Testing as a Graduation Requirement: The Case of the CET (College English Test) in China
Robert J. Fouser (Kyoto University)

第 4 室 (D202 教室) 司会 大坪 一夫 (麗澤大学)

⑰ 発表 I 非母語話者の発話に対する日本語教師の評価～教育経験の多少はどのような影響を及ぼすか～
鈴木 秀明 (神田外語大学)

⑱ 発表 II WEB上の英語語彙力テストの開発
風斗 博之 (東北学院大学)

第5室 (D203 教室)

司会 法月 健 (静岡産業大学)

⑱ 発表 I Self-Assessment, Peer-Assessment, and Teacher-Assessment on Japanese University Students' writing an English Composition

Sumie Matsuno (Aichi Prefectural University)

14:50~15:00 休 憩

15:00~16:50 シンポジウム (C301 教室)

Topic: **The Role of Testing in Language Learning**

Coordinator & Panelist: **Yuji Nakamura (Tokyo Keizai University)**

Panelists: **Ken Norizuki (Shizuoka Sangyo University)**

Garold Murray (Tokai University)

Yoshinori Watanabe (Akita University)

17:00~17:30 総 会 (C301 教室)

司会 島谷 浩 (熊本大学)

議長選出

報告 事務局長 中村 洋一 (常磐大学)

17:30~17:45 閉会の挨拶 (C301 教室)

副会長 **Randolph Thrasher** (沖縄キリスト教短期大学・
国際基督教大学名誉教授)

18:00~19:30 懇 親 会 (くすの木会館レセプションホール)

司会 **Terry Laskowski** (熊本大学)

發表要旨 (ABSTRACTS)

基調講演

Interactive Impacts of English Testing, Teaching, and Learning in Korea

Oryang Kwon (Seoul National University)

An educational reform is multi-faceted, and is not obtainable through a single approach. It needs to include changes in all the major facets of education, such as curriculum, textbook, teacher training, testing, teaching methodology, and administrative supports. Since the early 1990s, the government and academic circle of Korea have innovated and renovated several components of English education; they changed the curricula, textbooks, teacher training and retraining system and curricula, English testing, and classroom teaching methodology.

The 6th National Curriculum was promulgated in 1992 to become effective from 1995 for middle-school English and 1996 for high-school English. The 6th Curriculum was innovative in that it discarded the old grammatical syllabus and adopted a modified version of functional syllabus. Other characteristics of the 6th Curriculum included comprehension before production, emphasis on communicative competence, and fluency over accuracy. All these characteristics meant drastic changes from the traditional grammatical syllabus. The new curriculum made it necessary to adopt a new type of test, which put an emphasis on communicative competence.

English language testing was the second to be changed in the series of English education reform. From 1992 to 1993, with the help of a group of ELT professors who had earned their doctoral degrees in ESL/EFL, the MOE administered a series of seven pilot tests for the new College Scholastic Ability Test (CSAT), which was first administered in 1993 as the nation-wide official college entrance examination for the school year of 1994. The CSAT English test writers, who were a group of ELT professors, proposed a new model of proficiency test emphasizing communicative ability, which included a listening comprehension test for the first time in its history and a reading comprehension test that emphasized the ability to understand the topic, main idea, and other discourse knowledge. The washback effect of the new CSAT was immediate and far-reaching. High-school and even middle-school English tests imitated the format and focus of the CSAT. The CSAT English test also made teachers and students pay attention to listening and communicative reading.

The textbooks based on the changed curricula spared more pages for communicative activities than previous ones did. Listening activities were included for the changed teaching and learning principles. When obligatory English instruction was introduced in 1997 into elementary schools, the textbooks for the first two years (i.e., 3rd and 4th grades) contained only pictures, as the curricula allowed only spoken English for the first two years of English instruction.

The 7th National Curriculum, which was to be effective in 2000 for middle school and 2001 for high school, continued to emphasize communicative competence, and yet included grammar as an important part for language learning. The 7th Curriculum also stipulated that classroom assessment be done through performance assessment, especially elementary school English.

Performance assessment posed difficulties for English teachers, for the teachers were not capable of implementing such assessment. The MOE provided the teachers with some practical guidelines and assistance through in-service training and supervision. The performance assessment thus became an important buzzword in English education in Korea.

In 2000, the MOE began to urge teachers to teach English through English (TETE). This reform, coupled with the performance assessment, began to have a tremendous impact on classroom teaching and learning. Although not all teachers are doing TETE, the step certainly increased the amount of English use by the teachers.

English testing, teaching, and learning in Korea have undergone reforms for a decade, and their interrelationships were observed in several ways.

シンポジウム

The Role of Testing in Language Learning

Coordinator & Panelist: Yuji Nakamura (Tokyo Keizai University)

Panelists: Ken Norizuki (Shizuoka Sangyo University)

Garold Murray (Tokai University)

Yoshinori Watanabe (Akita University)

The purpose of this symposium is 1) to bring together a range of papers in the field of language testing, learning, and teaching; 2) to allow participants discuss the issues concerned; and, 3) to answer the question "How language teachers can make effective use of the studies of language testing for theoretical and pedagogical purposes?"

Yuji Nakamura, Coordinator & Panelist

Impacts of External English Tests on Learners

One of the aims of using internal or external tests is to create a positive impact on the classroom. This paper presents some of the results of an external English test conducted on a large scale with university students. Although the focus is the impact of an external test on student motivation and attitudes, other findings related to teacher reactions and classroom language teaching are also reported. Implications are suggested for the use of external tests.

Ken Norizuki

Growth- and Learning-Oriented Testing and Assessment

Tests are often considered as tools of selection, grouping and grading, or even punishment. All of them (except for the last) are indispensable functions of testing, but they are not what testing is all about. This presentation will focus on in-class growth- and learning-oriented testing and assessment, which is designed to help students learn the language and enhance their ability and motivation to use it in a meaningful context.

Garold Murray

Language Testing and Learner Autonomy: Strange Bedfellows

The role of testing in language programs that promote learner autonomy and self-directed learning is problematic. This paper briefly looks at the theoretical issues, describes how teachers are meeting the challenges of putting theory into practice, and concludes by exploring needs and implications.

Yoshinori Watanabe

Does Evaluation Motivate Learners? - Not Necessarily!

It is normal to claim that if there is a 'good' test, students work harder for it than otherwise. However, a body of recent research into washback effects of language testing indicates that washback is an extraordinarily complex PHENOMENON rather than monolithic. This presentation attempts to explicate the relationship between motivation and evaluation practices in EFL, by referring to the research findings that have recently been made in the field of motivation and washback. Several suggestions will be offered as to how to motivate learners BY EVALUATION and how to investigate the issue empirically.

研 究 発 表

① Can-do 項目作成のための英語能力評価共通枠組みの検討

長沼 君主 (清泉女子大学)

和田 朋子 (東京外国語大学大学院)

田中 敦英 (桐朋中・高等学校)

根岸 雅史 (東京外国語大学)

近年、項目応答理論(IRT)を用いた絶対尺度による評価が注目を集めている。例えば、TOEFL や TOEIC をはじめとして、CASEC や英語コミュニケーションテストといったテストでもその背後では項目応答理論が用いられ、一貫したスコアを保証している。ただし、ここで注意しなくてはならないのは、これらのスコアはあくまでも「ものさし」であって、英語を使って何ができて、何ができないかを保障しているわけではないことである。

そこで、本研究では、能力をテストによるスコアではなく、何ができるかといった視点から客観的に自己評価をさせる can-do 項目に着目し、言語能力評価における既存のフレームワークを概観しながら、能力評価の共通の枠組みを作成し、言語能力の発達段階を記述することを試みた。

能力評価の観点はそれぞれの評価基準の作られた背景によるところが大きく、共通枠組みを設ける中で、逆にその点が浮き彫りになった。TOEIC や英語コミュニケーション能力テストでも、can-do 項目を用いた分析をもとにコメントの作成が行われているが、このような共通枠組みを用いて、目標とする対象者に求められる能力を評価するのに最適な観点を選択していく必要があるだろう。

② 日本人英語学習者の英語メンタルレキシコン—感情を表す形容詞の定着度—

折田 充 (八代工業高等専門学校)

本研究は、日本人英語学習者のメンタルレキシコン形成における、英単語の定着度の違いを探る目的から、3種類のタスクを用いて感情を表す4種類30個の形容詞についてその定着度を調べた。実験では、Haastруп and Henriksen (2000)で用いられた3つのタスク(Primary Sorting Task; Card-Sorting Task; Situation Task)を日本人大学生に課し、次のような研究課題に基づき結果を分析した。

- ① Primary Sorting Taskの方が、Card-Sorting Taskよりやさしいか。また、HAPPYカテゴリ(*cheerful, pleased, glad*など)が、AFRAIDカテゴリ(*anxious, frightened, scared*など)よりやさしいか。
- ② 3つのタスクは次の順番で難度が上がるか： Primary Sorting Task < Situation Task < Card-Sorting Task。
- ③ Card-Sorting Taskについて、HAPPYカテゴリやANGRYカテゴリ(*furious, mad, annoyed*など)がAFRAIDカテゴリやSADカテゴリ(*miserable, sorrowful, depressed*など)よりもやさしいか。

発表では上記の研究課題について実験の結果を報告した後、次の諸点を検討したい。

- (a) 実験結果に関連する要因としてどういったものがあるか。
- (b) タスクの難易度を左右する要因にはどういったものがあるか。
- (c) タスクによる出来不出来の差が、メンタルレキシコン内における英単語のネットワーク形成度に関連すると言えるか。
- (d) よりよいメンタルレキシコンの構造を探るタスク作成のためには、どういった配慮が必要か。

(参考文献)

Haastруп, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10, 221-239.

③ A Possibility of Norm-Referenced Testing for Educational Purposes in Japan – A TOEIC Case Study –

**MIZUMOTO Atsushi (Graduate School of
Language Education and Information Science,
Ritsumeikan University)**

Tests play an integral role in language teaching and learning. Especially, in a country like Japan, where English is learned as a foreign language and the amount of learners' exposure to English in daily lives is limited, tests often tend to be regarded as the center and the yardstick of one's English proficiency. Hence, the whole idea of studying and teaching English for most people could boil down to "study for tests." In spite of this widespread use of tests, however, the consensus in the testing community is that there is no such thing as a perfect test although most of the lay people do not know this undeniable fact. In this study, a possibility of regarding a norm-referenced test, the TOEIC, as a criterion-referenced test was explored. A questionnaire and a question analysis were carried out. Test takers revealed their unawareness toward the differences among tests administered to them. Furthermore, the TOEIC questions analyzed showed the distinct characteristics of the TOEIC, which has common questions test takers should learn in every administration. Based on these findings, a 12-hour TOEIC preparation course mainly focusing on teaching questions that are most likely to be asked was conducted. The scores of the students, with the idea that the TOEIC can be studied for, after the course showed clear score gains compared with those before the treatment. The result suggests a possibility of altering the notion of conventional test practices employed in the TOEIC into one for educational purposes.

**④ Enhancing the Learning, Teaching and Testing of New Vocabulary
by Systematically Using a Depth of Lexical Processing Scale and a
Vocabulary Learning Strategies Taxonomy**

John Paul Loucky (Seinan Women's University)

Although developing good guessing skills is very important, especially when one is without access to a dictionary, training students to use both bilingual book and computerized dictionaries by maximizing their retrieving, recording, rehearsal and recycling functions for most rapid TL access and acquisition now seem to be far more important skills for low proficiency language learners to develop, especially as better software, portable electronic bilingual dictionaries and Internet-mediated machine translation equipment becomes more widely available (See Author, 2002a, c, d, & 2003b, c, d).

This study aimed to investigate how more Systematically Use of a Depth of Lexical Processing Scale and a Vocabulary Learning Strategies Taxonomy could help improve both the learning, teaching and testing of various aspects of both L1 and L2 vocabulary learning, while looking at these research questions:

- 1) What vocabulary learning strategies (VLS) do Japanese college students use when learning more advanced English academic vocabulary (EAP)?
- 2) How may VLS identified in previous studies be more clearly classified and extended by research, and more effectively practiced and learned by foreign language students?
- 3) Can a set of the most essential phases of vocabulary learning be taught logically, simply and systematically so as to affect positive change in learners' VLS awareness and habits?
- 4) Does a particular approach to or combination of strategies in academic vocabulary learning (EAP/ESP) help to predict participants' degree of success on English proficiency or reading level exams, indicating a greater possibility of future success in study or work abroad in English-speaking countries?

⑤ プレイメントテスト記述課題の量的・質的分析

西村よしみ (筑波大学留学生センター)

筑波大学留学生センターでは、日本語授業に参加する新入留学生に対し、各学期プレイメントテストを実施している。テストの内容は、多肢選択式テスト(文法・聴解・語彙)・SPOT・記述課題で、所要時間は2時間強である。記述課題は、制限時間15分の内に与えられたテーマについて記述する。本研究の目的は、記述課題を量的・質的観点から分析し、特に多肢選択テストと相関を検討することにより、記述式テストを評価するための有効な指標を探る。研究方法は、記述課題の電子化されたもの(N=110)を総語数、T-unit数、T-unit/総語数、正解T-unit/総T-unit数、Idea unitなどの指標と多肢選択式テストの下位テストとの相関を検討する。更に、質的分析の指標としてIdea unit群の質的分類を行う。Idea unitの分類カテゴリーは以下に示す。1) 同一表現 2) 類似表現 3) 言換え表現 4) 脱特定化 5) 抽象化表現 6) 具体化表現 7) 合成表現 8) 統一表現 9) 推論表現 10) 個性的表現 11) コメント表現 12) 誤り表現

分析結果は、総語数(漢字+かな)と文法の相関は($r=.698$ $p<0.01$)、正解T-unit/総T-unitと文法テストとの相関は($r=.668$ $p<0.01$)、Idea unit総数と文法は($r=.679$ $p<0.01$)というかなり高い相関が見られた。

本研究により、記述式テストと多肢選択テストの相関、及び多肢選択テストでは測定できないIdea unitの質的分析を通してプレイメントテストで測定する日本語能力像を推測することができるのではないかという指針を得た。

⑥ TOEIC で見える新カリキュラムの成果

吉川 勝正 (熊本学園大学)

熊本学園大学経済学部国際経済学科では 2000 年 4 月より英語のカリキュラムを刷新し、新カリキュラムを導入。この新カリキュラムの最大の特徴は、TOEIC (IP テスト) がカリキュラムの中に組み込まれており、学生は 1 年次より年 2 回 (5 月下旬頃と 11 月下旬頃) 受験し、その点数が各英語の科目の成績評価の重要な資料となる点である。その他の特徴を挙げると、(1) 英語の履修希望者は 1 年次の春に TOEIC を受験し、300 点未満の学生は後期週 2 コマの基礎英語の履修が義務づけられる。300 点以上の学生も履修してよい。1 年前期は英語の授業は全くない (2) 英語の正規の授業 (英語 1~4) は 2 年前期より始まる (2) 英語 1~4 はすべて半期制で、それぞれ 2 単位で週 2 コマ。4 単位コースの学生は 2 年後期で終了し、8 単位コースの学生は 3 年後期で終了 (3) 8 単位コースの終了目標として TOEIC400 点を掲げている (4) 基礎英語と英語 1 と 3 は日本人教師が担当し、英語 2 と 4 は母語話者が担当。

以上の新カリキュラムを受けた 8 単位コースの学生 (2000 年入学の約 80 名と 2001 年入学の約 50 名) の入学時から終了時までの半年毎の TOEIC の点数の推移を中心にカリキュラムの成果を考え、さらに今後の課題を挙げておきたい。

⑦ 語彙知識によるスピーキング能力の予測可能性

小泉 利恵 (筑波大学大学院博士課程)

語彙知識はスピーキング能力の一部と考えられ、スピーキング能力で大きな役割を果たすと考えられている。しかし、「語彙知識が、スピーキング能力や発話のどの部分にどの程度影響するか」については、あまり調べられてきていない。本研究では、(1) その点を検証し、(2) 「語彙知識からスピーキング能力をある程度予測し、適切なスピーキングテスト (ST) のタスクを選択することが可能か」を探索することを目的とする。本研究は、ST の実用性向上や、スピーキング能力の実証的なモデル構築に向けての第一歩として位置付けられる。

参加者は約 170 名の日本人英語学習者 (中学 3 年生 150 名・高校 1 年生 20 名) で、2 種類のテスト (ST と筆記テスト) を受験した。ST は、面接官と受験者が対面して行われた。本研究では、5 つのタスクの中で長めの発話を引き出した 3 つ (自己紹介・好きなものについて・2 枚の絵の比較) を分析対象とした。書き起こした発話は、語彙・流暢さ・正確さ・複雑さの観点で分析し、評価を行った。筆記テストには、発表語彙・受容語彙・文法問題が含まれていた。分析では、項目応答理論・共分散構造分析などが用いられ、発表語彙知識がスピーキング能力に強く影響することなどが示された。

⑧ 日本語 Can-do-statements の IRT 尺度化

野口 裕之 (名古屋大学)

和田 晃子 (北海道大学)

熊谷 龍一 (名古屋大学)

脇田 貴文 (名古屋大学)

日本語 Can-do-statements は、日本語能力試験の妥当性を検証するための外的基準として、開発された自己評価方式による、日本語能力の評価尺度である。「読む」「書く」「話す」「聞く」の4技能につき各15項目、全部で60項目の質問から構成される。例えば、「郵便局・銀行での説明がわかりますか」などの質問に、「1 全くできない」から「7 問題なくできる」までの7段階で回答させる。これまでに項目分析、因子分析、信頼性分析などが実施され、1) すべての項目が、一定程度以上の識別力を持ち、2) 質問項目群は、全体として一因子構造を示し、3) 技能別・全体ともに高い信頼性を示し、4) 大学のプレイスメント・テスト(大部分の項目が日本語能力試験から抜粋)とは高い相関を持つ、などの結果が報告されている。本報告では、この日本語 Can-do-statements に対して、項目応答理論による分析を実施した結果について示す。具体的には、1) 複数のIRTのモデルの中から最適なモデルを示す、2) 多値型のIRTモデルを用いて、評定尺度の等間隔性について確認する、3) 尺度に含まれる項目の難易度の分布の適切性を検討する、4) テスト情報量を用いて測定精度をきめ細かく評価する、5) IRT特性尺度値と従来の加算得点との関係を検討する、6) 回答者の属性によって回答傾向の異なる(DIF)項目を同定する、などである。IRT尺度化によって、日本語 Can-do-statements をより洗練度の高い測定尺度にすることができる。

⑨ Vocabulary Level Test at the 1000-Word Level

Paul Westrick (Kyushu University)

J. Lake (Fukuoka Jo Gakuin and Kyushu University)

The Vocabulary Levels Test (VLT) (Nation, 1983; 1990) is widely used in language education research for estimating vocabulary sizes of ESL/EFL students (Nation, 2001; Beglar, 1999; Schmitt, N, Schmitt, D. & Clapham, C, 2001; Read, 2000). The VLT is used to correlate vocabulary size with other linguistic research measures, as part of a test battery in placement tests, to gauge the appropriate range of graded readers for extensive reading, to gauge vocabulary size for materials development and other pedagogical purposes. Traditionally the levels have been at the 2000, 3000, 5000, 10,000 and “University” word levels. The 2,000-word level tests include items from the 1-1,000 and the 1,001-2,000 levels; however, in an EFL context or with lower ability students there may be a need to assess more precisely below the 2,000-word level. The Westrick (2001) 1,000-word level test was designed to do this. In this paper, we explain the development, analysis and validity of this 1,000-word level test and suggest future research directions and uses.

Beglar, D. and Hunt, A. (1999) Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing* 16, 2: 131-162.

Nation, I.S.P. (1983) *Testing and teaching vocabulary. Guidelines* 5, 1: 12-25.

Nation, I.S.P. (1990) *Teaching and Learning Vocabulary*. Mass: Newbury House.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 18, 1.

Westrick, P. (2001). Making and using vocabulary tests for vocabulary size and placement decisions. Presentation for Fukuoka JALT. December 1, 2001.

⑩ Examining the Rationale of Vocabulary Size Tests Based on Word-Frequency Levels

KATAGIRI Kazuhiko (Reitaku University)

Mochizuki (1998) developed one form of the Vocabulary Size Test (VST), which consists of seven sections (1,000, 2,000, 3,000, 4,000, 5,000, 6,000, and 7,000 word-frequency levels). Since then, Mochizuki has developed the two more forms of the VST so far. Katagiri and Mochizuki (2002) confirmed the parallelism of the three forms. In this study, the author will examine whether the rationale of estimating vocabulary size of Japanese learners of English through each section (each word-frequency level) is appropriate from the psychometrical view point.

University students whose major is English took six sections (2,000 - 7,000 word-frequency levels) in each of the three forms of Mochizuki's VST. Since six variables (subtotals of correct answers in each section) were obtained in each form, Pearson's product-moment correlation coefficients among the six variables were calculated in each form, and explorative factor analyses were conducted.

The results showed that six sections (2,000 - 7,000 word-frequency levels) had high correlation with one another (especially the stronger correlations with the closer word-frequency-level sections) and that only one factor was extracted in each of all the three forms. This may indicate that no section (no word-frequency level) is heterogeneous and that each word-frequency-level section is sequentially connected. Theoretically no section (no word-frequency level) should be heterogeneous and that each word-frequency level section should successively be connected; empirically, in this study, it may partially have been confirmed from the psychometrical results obtained through the interaction between the examinees and the VSTs.

⑪ The Role of the TOEIC in Measuring Language Learning

Sandra V. Williams (Hiroshima Kokusai Gakuin University)

This study observes the results of the TOEIC administered to the same subjects on two different occasions. There are 3 aims in conducting this study. The first is to describe the test results. The second is to interpret the test scores. The third aim of this study is to analyze the correlation between different sets of scores. At the time of the administration of the test, the subjects were students at a two-year college for women. They were English majors and their native language was Japanese. The subjects were required to take the test at the beginning of the first semester after they enrolled in the college. They were also required to take the TOEIC at the beginning of their second year. The study begins with a description of how the subjects performed on the test. The central tendency, as determined by such statistical indicators as the mean, the mode, the median and the midpoint generated on each administration of the test, is described in this study. A discussion of how the subjects' performance varies from the central tendency is also conducted with reference to the standard deviation and the test variance. The correlation between the reading and listening scores on each test is considered. Finally, the results of this study are summarized and the implications of these results on English language education are considered.

⑫ Multiple Imputation in Latent Growth Curve Analysis

Steven Ross (Kwansei Gakuin University)

That last decade has witnessed growing interest in language program impact analysis among language assessment specialists. Among the tools devised for impact analysis are longitudinal analyses such as repeated measures and latent growth curve models. A key issue in longitudinal analysis is the phenomenon of participant attrition. Early measurement waves usually have more complete data than latter waves in most repeated measures designs. The present study addresses methods for dealing with missing data in a sample longitudinal design using Institutional TOEFL scores. Three missing data methods are used on the same data set: list-wise deletion, replacement of missing data through linear regression, and multiple imputation through expectation maximization. The three versions of the reconstituted data set are then tested in a covariance structure analysis devised to model a latent growth curve. The three versions of the same data set are then subjected to a multiple group covariance structure analysis designed to test the equivalence of their respective latent growth curves as well as the equivalence of constraints. Implications for the use of imputation will be the focus of discussion.

⑬ An Investigation of Method Effects in Reading Comprehension

Tests: Text Structure and Response Format

**Miyoko Kobayashi (Graduate School of Language
Sciences, Kanda University of International Studies)**

Test developers must minimise the influence of intervening factors relating to test conditions if they are to produce an accurate measure of learners' language abilities. Following Bachman's model of method facets (1990) (later revised in Bachman and Palmer, 1996), this study focuses on the nature of input and of expected response by manipulating text structure and response format, both of which play a significant role in reading comprehension tests. The types of text structure investigated are: problem-solution, causation, description and association (Meyer 1985), and the types of response format examined are: cloze test, open-ended questions, and summary writing (Kintsch and Yarbrough 1982).

The study analyses the results of reading comprehension tests which were delivered to 754 Japanese university students. The students were randomly divided into twelve groups, with each receiving one of the twelve test sets. The main finding is that the two variables in question had a significant impact on the students' test performance. When texts were clearly structured, the more proficient students achieved better results in summary writing and open-ended questions. This suggests that well-structured texts help differentiate between students with different levels of proficiency. Test developers have hitherto taken little notice of the impact of text structure and response format on test results, and they will enhance the validity of their tests by paying more attention to these factors.

This paper further reports the results of an investigation into reading passages for university entrance examinations in Japan to illustrate the current practice of text selection for these high-stakes tests.

⑭ 実践報告—理科系の大学生に求められる英語力とテスト、およびその対策

河野 円 (星薬科大学)

理科系の大学生の英語学習において、テストは大きなインパクトを持っている。テストは学習の目的ではないのだが、彼らが受ける英語のテストが増え、大学のカリキュラムもそれに対応せざるをえない。本発表では、学生をとりまく「テスト環境」の多様化を論じ、それを考慮しながら教育を行っている実践を報告する。

彼らが受けるテストは、1) 学内の定期試験、2) 就職に必要な TOEIC などの検定試験、3) 留学希望者にとっては TOEFL、4) 国内の大学院進学希望者にとっては、大学院の入試、の4種類で、留学や大学院進学希望者は年々増加している。

これらのテストが要求する英語力は、速読のスキル、大意を把握し英語・日本語で要約するスキル、語彙の知識、リスニング能力、専門分野の基礎的用語の知識、英文エッセイの力、訳読のスキルなどである。授業での具体的なメソッドや教材を示し、図書館や情報教室とのタイアップにも触れながら報告を行う。

学生の反応として、たとえば、TOEIC 対策では、英文の量やリスニングのスピードに圧倒されてしまうこと、また、科学論文の訳読では、段落全体の意味を考慮しながら一文一に正確な日本語訳をつけることができない、などの、傾向を指摘する。

そこから浮かび上がる指導上の留意点を論じ、さらには大学の英語教育のあり方について、フロアとのディスカッションを行う時間もとりたい。

⑮ Introducing Exit Competencies into a University English Course

Brendan Rodda (Kanda University of International Studies)

The presentation will provide information about the introduction of exit competencies into BEPP, a general skills course for first-year students at KUIS. It will also look at how this process is influencing the course curriculum in general.

We are currently working on specifying clear and detailed exit competencies and plan to trial assessment tasks based on these later this year. Full implementation of the exit competencies assessment system is scheduled for the 2005-06 academic year.

This system will allow students who satisfy the exit competencies to move out of BEPP early and will require students who do not satisfy them to remain in the course beyond their first year. Students will be assessed on their performance on a series of tasks that require the integration of a receptive and a productive skill (e.g. listening-speaking), reflecting the course's focus on integrative skills. Tasks will be similar to communicative tasks used in the classroom and will take 20-30 minutes to complete. Students will be aware of the exit competencies they need to satisfy but will not have specific knowledge of the tasks, so will not be able to prepare or practice their actual performance.

⑩ Standardized Testing as a Graduation Requirement: The Case of the CET (College English Test) in China

Robert J. Fouser (Kyoto University)

This paper serves as a comprehensive introduction to the CET (College English Test), which has been required for all students of four-year universities to receive a bachelor's degree since 1987. The test is developed by the National College English Testing Committee of China and is taken by 2.5 million students annually. The test covers listening, reading, and writing and has two versions: Band 4 and Band 6. Band 4 covers material taught in the first four semesters of university and is a graduation for all students. Band 6 covers material taught in three semesters of university and is required for English majors and, graduate students, depending on the university. Students who have not passed Band 4 cannot take Band 6 and students who fail Band 6 cannot retake it. The CET Spoken English, which was developed in 1999, is used as a supplemental test to Band 4 and Band 6.

Through a careful examination of examples from Band 4 and Band 6 tests from previous years, I will compare the two tests with each other and with other commonly used standardized tests in terms of organization, item characteristics, and possible difficulty. Using the JACET 8000 list of basic words, I will analyze the vocabulary and readability of reading passages to provide a quantitative basis for comparison. Finally, by referring to the literature and translations of official information from Chinese, I will discuss on the pedagogical and social implications of the CET on English teaching in higher education in China.

⑰ 非母語話者の発話に対する日本語教師の評価～教育経験の多少 はどのような影響を及ぼすか～

鈴木 秀明 (神田外語大学)

本研究は、日本語教師が非母語話者の発話の評価する際に教育経験はどのような影響を及ぼすかを調査するものである。先行研究では、発話能力の評価において、教師と一般人では評価の仕方に違いがあることは報告されているが、教師の教育経験の多少による評価の違いについての研究はまだ見られない。

調査協力者は、日本語教師 40 名で、平均教育経験年数は 5.6 年であった。調査では ACTFL-OPI のサンプルテープを使用し、発話テープを調査協力者に聞いて評価してもらった。調査では、①全体的評価、②分析的評価、③発話評価を行う際の教師の自己基準を記述してもらおうというタスクを実施した。全体的評価では発話を全体的印象で 5 段階評価してもらい、分析的評価では、発話をテキストタイプ、文法、音声、語彙、流暢さ、機能、内容の 7 項目について ACTFL のガイドラインに基づいて作成された基準で 5 段階評価してもらった。さらに、発話能力を評価する際の着目点や評価基準について、記述形式で回答してもらった。調査で得られた結果を教育経験の多いグループと少ないグループに分け、グループごとに分析した。本発表ではグループごとの分析結果について述べ、教育経験が発話能力の評価に及ぼす影響について考察を行う予定である。

⑱ WEB上の英語語彙力テストの開発

風斗 博之 (東北学院大学)

筆者は一般英語の授業履修者を中心に、ネット上に作成した英語単語テストを受験させ、各受験者の語彙力を測定する試みを行っている。受験者からは、同時に TOEIC や TOEFL のスコアを提供してもらい、それらのスコアをもとに語彙力を表示する。テストは IRT (項目応答理論) に基づいて作成し、問題項目や受験者が増えるほどその測定の正確さも増えるように作られている。受験者の能力・項目の難易度や識別度の計算には最尤推定法を使用している。問題の形式は現在、英和および和英の記述式であるが、選択式、絵や例文、音声を用いたものなど様々な形式を開発中である。

テスト問題は再編集し、別のテストセットとして再利用することができる。語彙力の正確な評価が受験者の学習意欲をかきたてる。また、テストのデータから得られた、問題項目の難易度から英単語の難易度を導き出すことができる。たとえば、よく使われる頻度順のリストとは異なる、英日方向の英単語の難易度順 (知名度順) のリストなどを作成し、語彙学習教材で利用することなどが可能となる。

発表では、個々の問題の難易度の計算法を含め、CGIプログラムでの実際の処理方法なども解説する。

⑱ Self-Assessment, Peer-Assessment, and Teacher-Assessment on Japanese University Students' Writing an English Composition

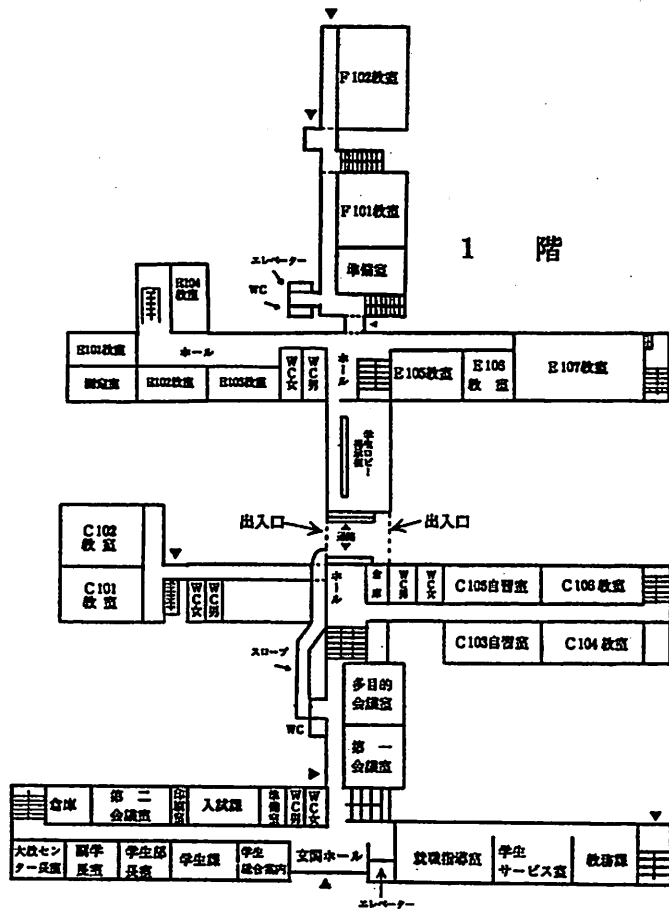
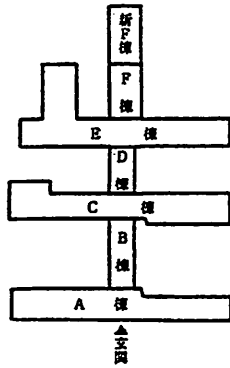
Sumie Matsuno (Aichi Prefectural University)

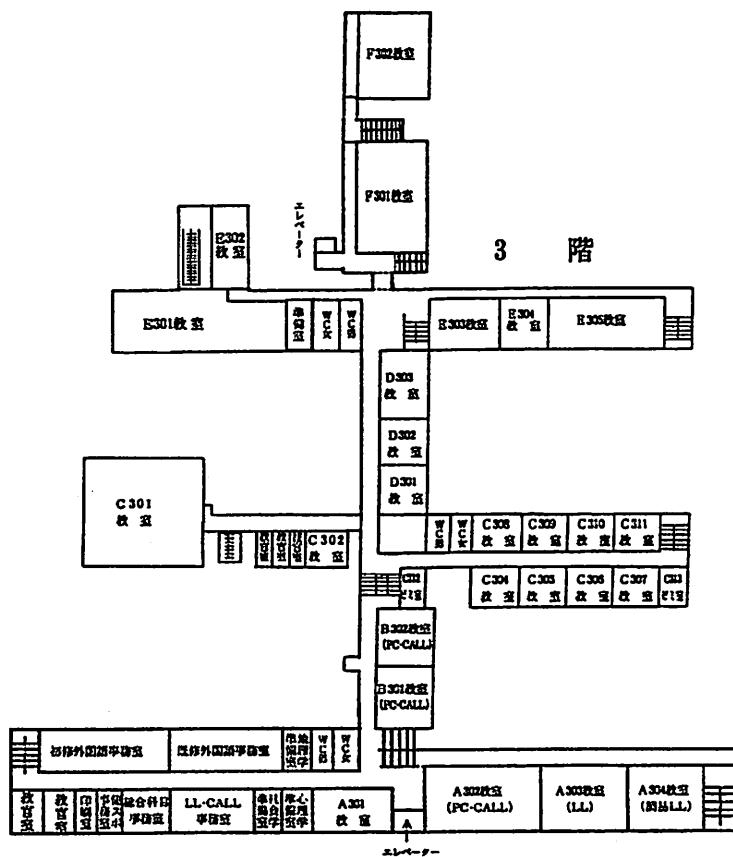
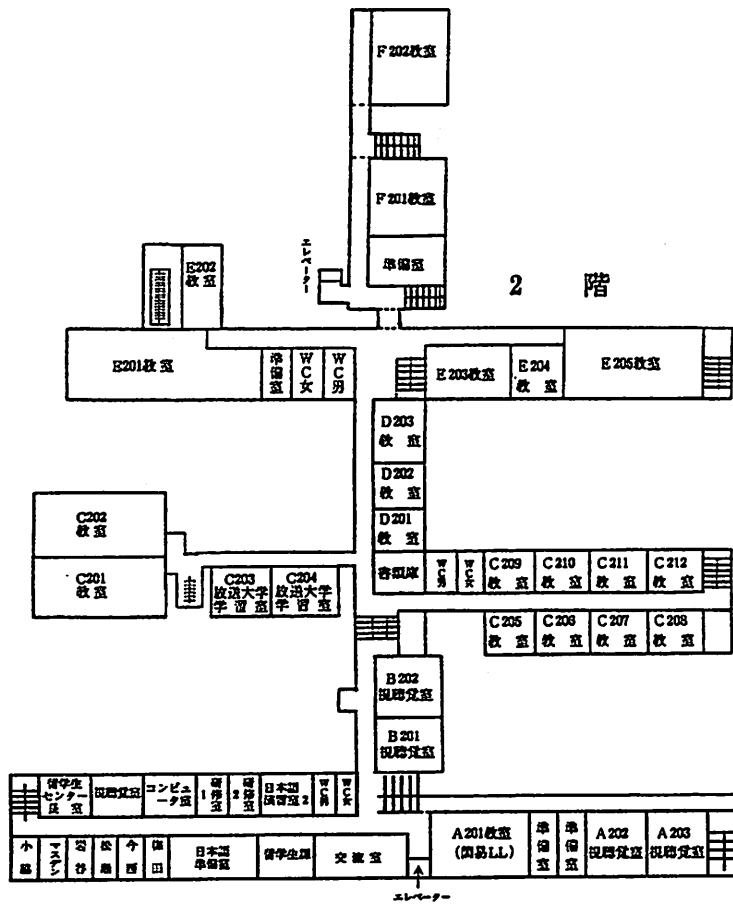
This study intended as an investigation of self-assessment, peer-assessment, and teacher assessment. Using the Multi-faceted Rasch analysis, the following questions were answered; To what degree are raters' severities, writers' abilities, and items' difficulties various?; To which items were student raters harsh or gentle?; To which items was the teacher rater harsh or gentle?; Which writers, raters, and items could not fit the model?; To what degree did self-assessment, peer-assessment, and teacher-assessment have bias?

After learning how to write an English essay for eight sessions, twenty-eight Japanese students (in the first study), and ninety-seven Japanese students (in the second study), enrolled in prefectural or national universities in Nagoya Japan, wrote a one-page (about 250 words) essay. The topic of the essay was: *Please discuss the advantages and disadvantages of university students having a cellular phone.* After writing the essay, in class, all participants evaluated the same three essays and I gave my own comments and grading in front of them. Then they evaluated their own essay and their peers' essays at home, which was worth 10 % of their grade in total. In addition to the student evaluations, I evaluated all essays.

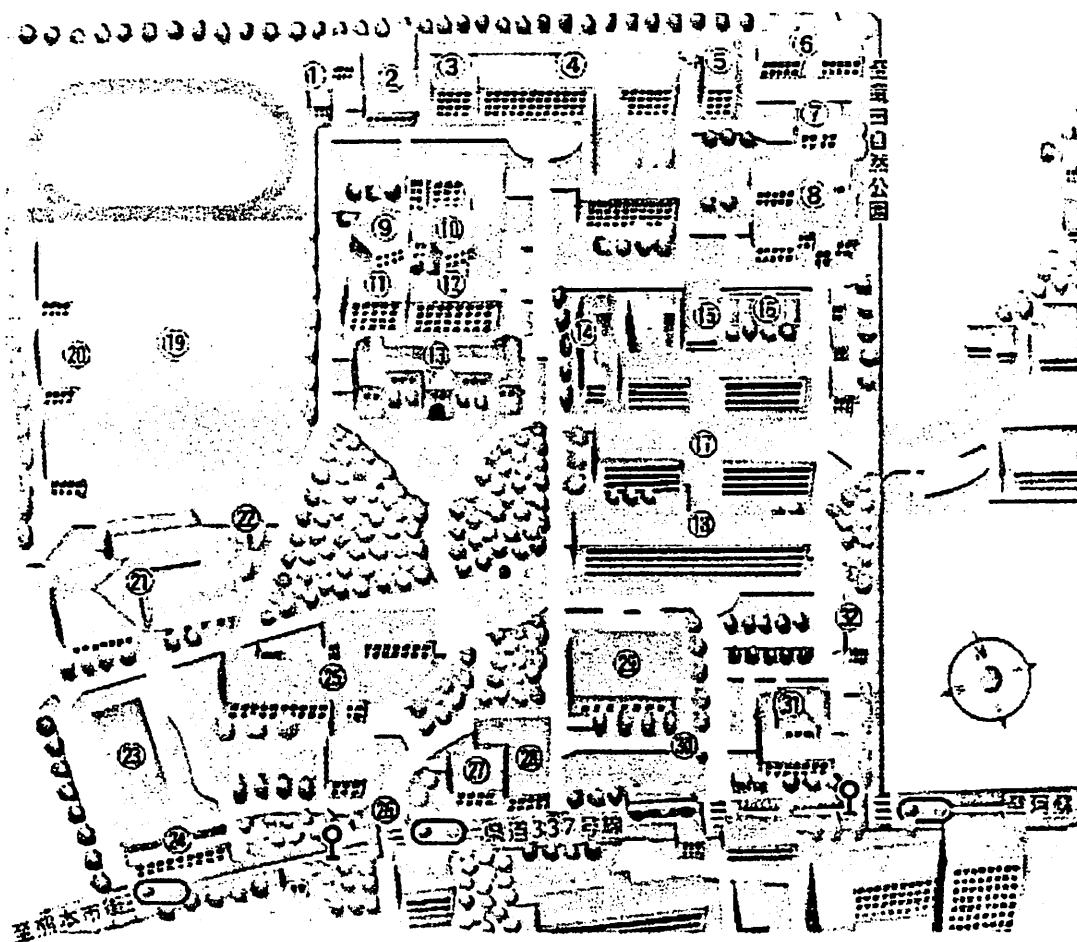
The results indicated the writers did pretty well on writing an essay. On the other hand, there was likely to be considerable variation in the rater severity. Regarding the items in the peer assessment, overall grammar and use of variety of sentence structures were the most seriously scored items; spelling was the least severe item to obtain high scores. Similarly, the teacher was harsh to grammar and lenient to spelling. Besides, many participants evaluated their own essays without any bias; however, five students rated their own scores with bias. All of them rated themselves lower than their expected scores.

熊本大学 大学教育センター棟 配置図





熊本大学 黒髪キャンパス 北地区 案内図



01	教育学部 音楽棟	12	社会文化科学研究科	23	プール
02	教育学部 実習工場	13	五高記念館	24	文化部室
03	教育学部 研究講義室	14	化学実験場	25	学生会館
04	教育学部	15	大学教育機能開発総合研究センター 講義棟	26	赤門
05	教育学部 東棟	16	テニスコート	27	生涯学習教育研究センター, 情報プラザ
06	外国人教師宿舍	17	大学教育機能開発総合研究センター	28	保健管理センター
07	知命堂 (宿泊施設)	18	学生部, 留学生センター	29	附属図書館
08	くすの木会館	19	運動場 (武夫原)	30	門衛所
09	文学部・法学部 講義棟	20	体育会部室	31	北地区食堂
10	文学部・法学部	21	体育館	32	環境安全センター
11	文学部・法学部 南棟	22	トイレ		

* 本学会の会場となっております「大学教育センター棟」は、上図中 15,17,18 番の建物全体を指します。

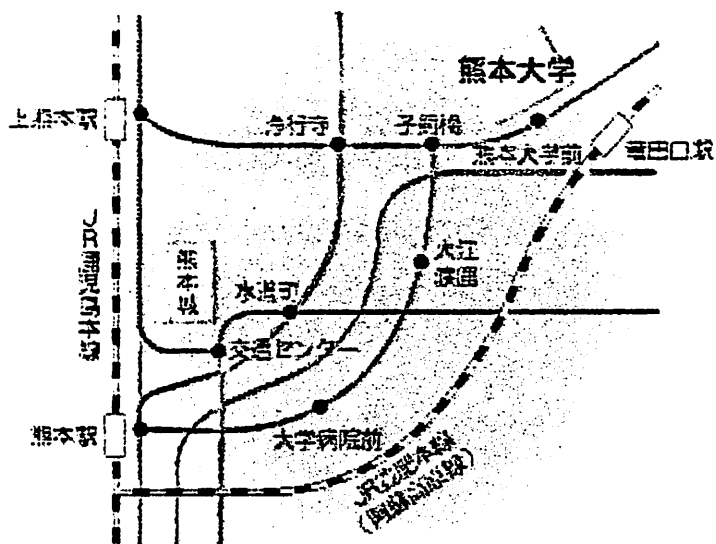
熊本大学 黒髪キャンパスまでの交通案内

熊本空港から	産交バス：熊本市内交通センター行き乗車。交通センターにて乗り換え。交通センターまでの所要時間は約 40 分（670 円）。
	タクシー：熊本大学黒髪キャンパスまで約 35 分。 約 3,700 円。
J R 熊本駅から	市営バス：竜田口駅、楠団地・武蔵ヶ丘行（8 番系統）乗車。「熊大前」まで 30～40 分。 第一環状線（渡鹿（とろく）経由，1 番系統）に乗車。「子飼橋（こかいばし）」まで 20～30 分。「子飼橋」から徒歩 10 分。
	産交バス：大津（三里木，陣内経由）武蔵が丘方面行き乗車。「熊大前」まで 30～40 分。
	タクシー：熊本大学黒髪キャンパスまで約 20～30 分。 約 1,700 円。
J R 上熊本駅から	市営バス：第一環状線（京町本丁経由）に乗車。「子飼橋（こかいばし）」下車。徒歩 10 分。
	タクシー：熊本大学黒髪キャンパスまで約 12 分。 約 1,000 円。
交通センターから	産交バス：大津（三里木，陣内経由）武蔵が丘方面行き 13 番乗り場にて乗車。「熊大前」まで 20～30 分。
	市営バス：竜田口駅，楠団地・武蔵ヶ丘行（8 番系統）12 番，13 番乗り場にて乗車。「熊大前」まで 20～30 分。
	タクシー：熊本大学黒髪キャンパスまで約 15 分。 約 1,400 円。

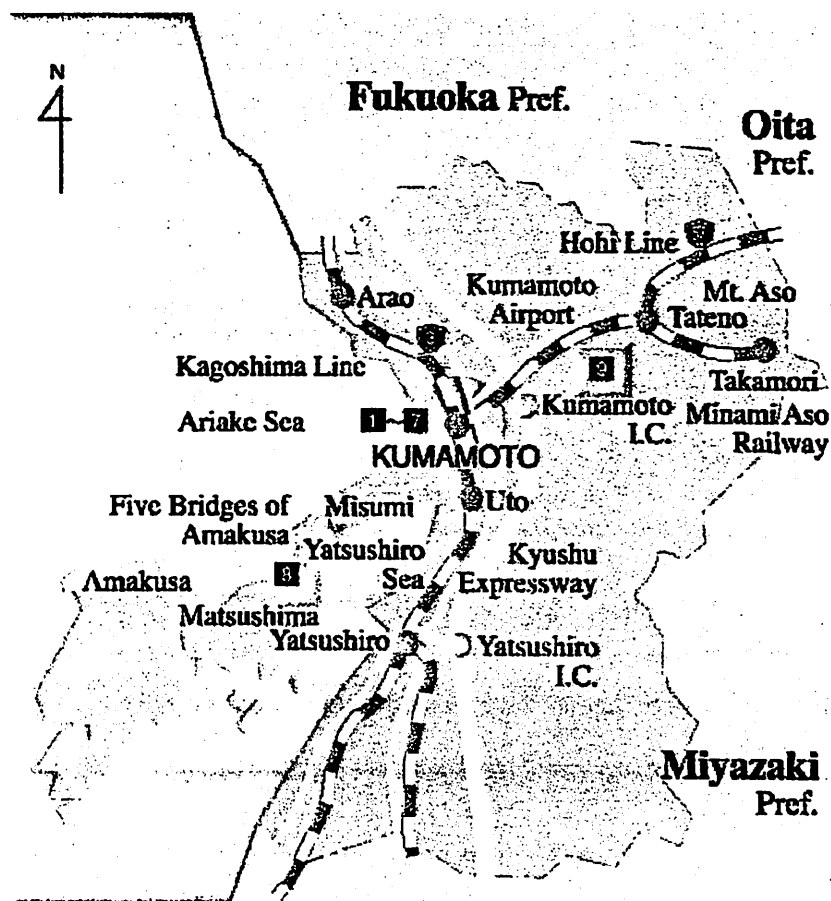
- * タクシー乗車の際には、「熊本大学の黒髪キャンパスまで」とお伝え下さい。
- * 熊本空港からバスに乗り「交通センター」に着く前に、「通町筋」で下車して、路線バスに乗り換えると、時間の短縮ができます。「通町筋」バス停から、横断歩道を渡り、「水道町」バス停まで少し歩き、「熊本大学前」行きのバスに乗車します。
詳細は、<http://www.kumamoto-u.ac.jp/campus-map/photo-air.html> で。
- * 次の Web ページは、写真入りで、熊本大学黒髪キャンパスまでの行き方を案内いたしております。ご参考になさって下さい。

<http://www.kumamoto-u.ac.jp/campus-map/cmap-1st.htm>

熊本市中心部略图



熊本県略图



お 知 ら せ

1. 受付は、大学教育センター棟1階玄関ホールで行います。
2. 大会参加費は、会員が1000円、非会員が3000円です。
3. 昼食は、学内の食堂または大学近隣の食堂をご利用ください。
4. 懇親会費は5000円です。会費は、学会当日、受付でお支払い下さい。
会場は、熊本大学くすの木会館レセプションホールです。
5. 大会参加・懇親会の申込締切を、10月10日まで延長いたします。ただし、宿泊・航空・オプションツアーの申込みは、9月末日までの受付といたします。既に、配布いたしております申込書を使って、JTB熊本支店ECセンターまでお申込み下さい。