

JLTA Newsletter
外国語教育評価学会
The Japan Language Testing Association

JLTA Newsletter No. 8 発行代表者：大友 賢二 2001年(平成13年)2月1日発行
発行所：外国語教育評価学会 (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>



2001年度 JLTA の課題に備えて

JLTA 会長 大友 賢二 (常磐大学教授・筑波大学名誉教授)

JLTA の仲間の中で研究テーマの話題として出ていることが数多くあります。そのなかで、2001年度の研究・活動の課題の一つとして、CODE OF PRACTICE はどうかということがあります。これは、LTRC 99 のテーマでもあった「21世紀における言語テストの社会的責任」とも直結する重要な課題です。言語テスト作成者・実施者・受験者はどんな事に留意してその社会的責任を全うすべきかということです。この課題に備えて、以下の2つのことをここでは取上げたいと思います。ひとつは、Fredric M. Lord 氏の逝去のことです。もう一つは、1999年版の *Standards for Educational and Psychological Testing* の発行についてです。

第1は Lord 氏の逝去について考えることです。2000年5~6月にわが国の8会場で行なった JLTA Language Testing Workshop では、ETS (Educational Testing Service) と協力して、TOEFL の CBT (Computer Based Testing) 化への課題解決に全力を尽しました。そして、CBT を可能にしているのは項目応答理論 (IRT: Item Response Theory) であることの認識を新たにいたしました。その IRT 開発に中心的役割を果たしていた Frederic M. Lord 氏は2000年2月5日(土)、Florida の Naples においてその生涯を閉じました。87歳でした。この Lord 氏の研究とはまったく独立して、かの Rasch Model を開発されたデンマークの Georg Rasch 氏もすでに1980年に79歳で亡くなっています。お二人とも見事な人生でした。Lord 氏は1944年から1982年までの38年間、ETS においてその道一筋に全力を注がれました。この機会に、IRT を原点に戻って考え直してみることをお勧めします。氏の主要著書には、Lord and Novick, *Statistical Theories of Mental Test Scores*, Addison-Wesley Publishing Company, Inc., 1968. や Lord, *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates, Publishers, 1980 があります。

第2は、1999年版の STANDARDS のことです。言語テストの CODE OF PRACTICE に最も関連する文献の一つは、American Educational Research Association, American Psychological Association, National Council on Measurement in Education の合同委員会が作成している *Standards for Educational and Psychological Testing* です。1985年版よりかなりの年月を経て、1999年版が出版されています。1985年版では98頁であったのに対し、1999年版では194頁にも及んでいます。テスト作成者・実施者・受験者が新しい視点から考えるべき重要な項目が数多く見られます。研究課題の参考文献の一つとして、ぜひ、一読を薦めたいものです。

第4回 JLTA 全国研究大会 報告

2000年10月8日(日)、第4回 JLTA 全国研究大会が、「コンピュータ使用による言語テストの理論と実践」(Computerized Language Testing: Theory to Practice)を大会テーマに掲げ、東京経済大学に於いて開催されました。以下、全国研究大会の報告です。(文末に報告者が明記されていないものは、発表要旨を基にした概要で、文責は事務局にあります。)

基調講演

「コンピュータテストの理論的背景について」 講師 前川 眞一 (大学入試センター)

わが国においても TOEFL にコンピュータが導入され、いよいよコンピュータテストが実用的な段階に入ったが、前川眞一大学入試センター教授による基調講演では、コンピュータテストの理論的基礎となる項目反応理論の解説とテスト作成のための基本的技術が紹介された。

項目反応理論では、項目反応の持つ特性を数量化して、困難度パラメータや識別力パラメータが利用される。項目困難度や項目識別力という概念は古典的テスト理論でも項目分析において用いられるが、項目反応理論においては、これらの概念は、純粋に項目の特徴として考え、テスト項目の特性(困難度や識別力等)は、受験者集団の特性(能力、学力、性格等)に依存しな

いものと定義されている。ただし、このことが可能となるためには、項目パラメータの分かっている項目を収集した項目バンクが必要である。

講演では、2値項目からなる基本的な項目反応の数学的モデル化を、体重測定という身近な例を取り上げて、項目特性曲線、局所独立性、特性値の推定、テスト特性曲線など、理論の理解に必要な基本的概念がていねいに解説された。受験者が同じテストを受けるのではなく、個人の特性値の大きさに見合ったテストを受けることが最も精度がよい測定であり、そのためにコンピュータが有効に用いられるのである。

(報告者：浪田 克之介 北海道大学)

パネルディスカッション

「コンピュータ使用による言語テスト (CAT/CBT) が外国語教育に与えるもの」

峯石 緑 (広島国際大学)

地方在住者が被りがちなデメリットが解消され、より多くの受験者が、より整った同様の環境下で受験できるという可能性も見られるが、コンピュータ・リテラシーが低い受験者の問題、コンピュータによる試験と筆記試験の並行管理の問題、CALL 教室の充実など授業の中での指導に関する問題などが課題として残る。

伊藤 彰浩 (愛知学院大学)

コンピュータを利用した言語テストを用いれ

ば、項目応答理論を用いた英語学力の測定と評価がより容易に実施できるようになる。しかし、コンピュータの整備がなされても、教師と学習者の測定や評価に関する意識がかわらなくてはコンピュータ化されたテストを利用することは混乱を招くことになりかねない。コンピュータを用いたテストは、言語産出をどの程度測定できるか疑問がのこる。よって、コンピュータ化されたテストとアナログ型の測定・評価法を折衷して利用することが必要なのではないか。

智原 哲郎 (大阪女学院短期大学)

Paper-and-Pencil Test から Computer-Based Test へのシフトは、新たな問題が出てくるものの、従来抱えていた問題を解決するものが多くなる。

CBT では、より authentic なテスト材料が使用され、より正確な言語能力の測定が可能になることから、CBT が授業のツールとして機能することによって、4 技能をバランスよく訓練しながらより現実に即した言語活動を中心とした方向に向かい、スキーマの増長を目指した内容重視型の授業に切り替わっていく。

一方、カリキュラムやその運営においては、学生/英語教員へのコンピュータリタラシーの必要性、より多くのクラスへのコンピュータの導入、ソフトの開発、など人的資源の調達や設備の整備が必要になる。

研究発表 I・II

I-1-1: Adapting the Levels Test into a Japanese Translation Format

Yoshimitsu Kudo (*Rikkyo University*)

Nation's vocabulary test was found reliable and valid, studies on the test's validity went so far as to revise the 2000 Word Level and the University Word Level of Nation's test.

However, no researcher has adapted Nation's test taking relatively low-level learners into account. Therefore, a new translation definition-matching format test was developed for criterion-referenced purposes that includes five lower levels, namely, the 1000-, 2000-, 3000-, 4000, and 5000-word levels.

The pilot study examined its dependability and validity, and the problem was identified as loanwords included in the 5000-word. In the main study, those loanwords were replaced by non-loanwords. This new, dependable, valid, and practical test can be used for both diagnostic and achievement purposes, and has a considerable

根岸 雅史 (東京外国語大学)

「無駄な項目」の削減によるテストの信頼性向上と結果のフィードバックの即時性が改善点としてあげられる。一方、言語材料やトピックなどの組み合わせが「偶然に」任されること、サブスキルの組み合わせが「偶然に」任されることに対する危惧や、CATの原理を知る受験者の心理については問題がある。

今後は、コンピュータの音声認識技術が高まることにより、スピーキング・テストがネットワーク上で行われる可能性がでてくる。また、発表技能のコンピュータによる自動採点の可能性や、「診断」から「学習」へと有機的につながったプログラム診断的機能の高いフィードバックが実現する可能性も考えられるであろう。

possibility of being developed to a computer adaptive test by making an item bank.

I-1-2: What a Computerized Reading Test Can, but a Pencil-and-Paper Reading Test Cannot, Do

Shizuka Tetsuhito (*Institute of Foreign Language and Education Research, Kansai University*)

The paper reports on the development and the evaluation of a computerized reading test. The first feature of the test is that its items are sentence-based, not discourse-based. The second feature of the test is that it records the examinee's reading speed. The third feature of the test is that it records confidence levels the examinee has in each response he/she makes.

Some 210 Japanese EFL college students took this test as well as the reading section of Cambridge First Certificate in English (FCE), which served as the criterion measure of the subjects' reading proficiency. The results clearly indicated that

incorporating confidence and speed resulted in substantially higher reliability and validity than when scoring was done based on 1/0 correctness alone. Even more remarkable was the fact that reliability and validity levels attained by combining the three measures at the point at which only the first four items were attempted exceeded those achieved by scoring exclusively on 1/0 data at the point at which all 40 items were tried.

I-2-1: Modifying an Existing Test to Allow it to be Delivered by Computer

Randolph Thrasher (国際基督教大学)

Jack Upshur and the presenter developed a two-stage test, the Businessman's English Test and Assessment (BETA) in 1976. The first stage (BETA I) consists of 210-item multiple-choice questions of English structure, vocabulary, listening comprehension and reading comprehension, and the second stage (BETA II) consists of 5 writing tasks and 2 face-to-face interviews. It has been used for 25 years but recently economic considerations have necessitated plans be made to permit it to be delivered by computer and make part of it computer adaptive. Prof. Thrasher pointed out two problems: computer based questions and maintenance of security. The dilemma of modifying BETA lies in how to keep reliability and validity high while reducing the cost of administering and scoring the test. Test items should be treated in line with the Item Response assumptions, and also a broad pool of test items is needed. In modifying BETA II, telex questions WILL BE replaced with e-mail tasks. Other issues about computer based testing include keyboard skills, speaking TASKS, how to inform test takers of the passage of time etc. After the presentation there was question about the correlation between BETA and TOEIC from the floor.

(報告者：望月 昭彦 筑波大学)

I-2-2: Peer Ratings in EFL Writing Course: User Acceptance and Characteristics

Hidetoshi Saito (Hokusei Gakuen University)

Tomoko Fujita (Rikkyo University)

In order to achieve two purposes, that is, to examine characteristics of peer writing ratings and to study attitudes toward peer ratings, two presenters addressed three research questions: 1) What are similarities and differences among peer, self-, and teacher ratings of EFL writing?; 2) Do students like or dislike peer ratings?; and 3) Is students' acceptance of using peer rating influenced by receiving high ratings from teachers and peers? The presenters conducted an experiment on 68 college students studying English writing at a Japanese college. The students took a rater training session and wrote 2 essays in one semester, each of which was rated by 2 teachers, 3 peer students, and the writers themselves with a 4-point rating scale. Students made final revisions using peer and teacher ratings and comments. They also completed a questionnaire on the utility of peer rating. The result shows that whereas peer and teacher ratings were highly correlated, peer-self and teacher-self ties were almost negligible. Students' positive attitudes toward peer ratings were revealed. Neither teacher nor peer ratings influence students' acceptance of peer ratings. After the presentation, questions about rater training sessions and topics of essays were given by the floor.

(報告者：望月 昭彦 筑波大学)

I-3-1: 英語能力簡易測定のための語彙テスト

－ 開発されたテストの妥当性・信頼性の検証

片桐 一彦 (東京学芸大学 博士課程)

日本の高校生・大学生の英語力を短時間で推定する語彙テストの作成を目的とする発表がなされた。英語能力推定簡易テストの先行研究について詳しく述べられ、発表者が過去に行った2つのプロジェクトにもとづき、第3番目の研

究の結果が発表された。10分のできる「項目困難度を基軸にした語彙テスト」の3つの版が作成され、3つの版が平行テストであるか、10分で実施できるかどうかに関して検証された。

(報告者：山崎 朝子 武蔵工業大学)

I-3-2: パブリックスピーキングテストの分析： ラッシュモデルが教えてくれるもの

中村 優治 (東京経済大学)

本発表は、パブリックスピーキングの測定と評価に関する研究報告である。授業中に行われたパブリックスピーキングのテスト結果が、ラッシュモデルを用いて分析された。被験者(大学生15名)、評価者(5名)、評価項目11項目に関して細かい分析と検討が行われた。

(報告者：山崎 朝子 武蔵工業大学)

II-1: Aパラメーターから見た言語テストの弁別力

湯浅 克己 (ILC BETA テストセンター)

ILC-BETA テストセンターで開発された10言語の語学テストから、受験者が多い5言語(英、仏、独、西、中)テストのサンプルを抽出。IRT 2pモデルで各項目の弁別力(Aパラメーター)を探り、更に素点の分布を比較。結果、英語テストのAパラメーター値が他の言語テストより低く、後者の素点が下位層に偏る傾向を確認。英語に比べて初心者と熟達者の差が大きく中間者層が薄いことが、他言語テストの項目弁別力を高めた要因である可能性を指摘。

(報告者：法月 健 静岡産業大学)

II-2: Test Data Analysis Program: TDAP Ver. 1.0E について

中村 洋一 (常磐大学)

TDAP Ver. 1.0 (大友・中村, 1996.) は、『項目応答理論入門』(大友, 1996. 大修館.) で示された古典的テスト理論及び項目応答理論のデータ分析方法に沿って、テスト・データの処理を行うPCプログラムである。

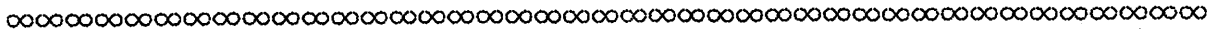
TDAP Ver. 1.0 はテスト・データの分析をPCで容易に行うことを可能にした点では画期的であったが、動作環境が限られ、現在のPC環境での稼働は事実上不可能となった。そこで、Windows版Excelのマクロ機能を用いた改良版のTDAP Ver. 1.0Eを発表した。Ver. 1.0Eでは取り扱い可能なデータセット量が増えた。前回の第一歩からさらに大きな第二歩を踏み出した、と言える。現在、ブラウザを使うHTML版を開発中とのことで、さらに使い易くなり、テスト・データ分析に貢献することが大いに期待される。

(報告者：飛渡 洋 国士舘大学非常勤)

II-3: Classical and IRT Approaches to System Feedback on Language Testing for University Entrance

Steven Ross (関西学院大学)

Feedback systems for high-stakes examinations such as those commonly used in Japan are largely dependent on non-empirical, impressionistic critiques. These critiques are often given as a form of test pre-administration moderation by committees of language specialists, who often lack expertise in SLA, language assessment, or psychometric theory. Another approach is to invite post-administration critiques from secondary school teachers or their representatives, as well as those from cram schools. While these types of critiques offer a modicum of feedback, they cannot suffice to aid the creation of comprehensive positive and negative feedback systems for the cumulative betterment of entrance examinations. The present paper outlines the use of classical test item analysis and item response theory approaches for the purpose of creating synchronic and diachronic composites of language tests on university entrance examinations. These composites, it is argued, are essential for diagnosing test malfunction, item weight determination, and test tailoring.



第 12 回 JLTA 研究例会 報告

第 12 回 JLTA 研究例会は、2000 年 12 月 2 日、神戸松陰女子大学で開催され、約 30 名の参加者で賑わった。以下はその報告である。

Plenary Speech では Prof. Steve Ross による **Foreign Language Education Macro-policies in Japan and Korea: Steps Toward an Impact Analysis** と題する講演がなされた。その概要は：国家レベルの外国語教育政策はマクロ経済的な考え方によって誘導される場合と、ミクロ経済的政策が優先される場合があるが、いずれにしても、その政策の結果が十分に評価されていないことが多い。Ross 氏は、TOEFL、TOEIC の結果の分析により、大学入試改革を通して英語教育改革を行っている韓国での教育政策と対比させながら、日本の JET プログラムの評価を行った。彼は、多額の金を費やしているにもかかわらず、JET プログラムによる語学力向上の成果はあまり期待できるものではなく、むしろ大学入試に Listening test を導入することが先決だとしている。

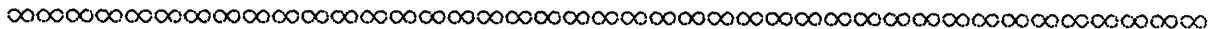
引き続き、研究発表となり、まず、Prof. Thomas Robb による **On the Estimation of TOEIC Gain in a Four-Year Program with Partial Data** では、コ

ース終了時の TOEIC の結果を基に、学業成績などを使用しながら、受験しなかった学生の TOEIC の得点の算出が可能であり、これにより、コースの成果を見る手段としての TOEIC の有効性があると論じた。

次に、Prof. Elizabeth Hiser による **Development of an Oral Proficiency Test for English as a Foreign Language** の発表があった。Carroll and Hall (1985) から得た手順を使い、口頭能力テストの開発過程を示した。このテストは、妥当性、信頼性ともに高く、TOEFL や Michigan Test よりも学力を正確に測定できるとしている。

最後に、Prof. Parrill Stribling による **A two Year Study of Japanese Learning Approaches** では、日本の短期大学生をサンプルとし、2 年間にわたる学習アプローチを測定し、この結果、Learning Process Questionnaire と Study Process Questionnaire の妥当性と信頼性が証明されたとしている。

(報告者：智原哲郎 大阪女学院短期大学)



計量心理学会国際大会のご案内

International Meeting of the Psychometric Society: IMPS-2001

計量心理学会国際大会 (IMPS-2001) が、2001 年 7 月 15 日(日)~7 月 19 日(木)、大阪大学コンベンションセンター (吹田キャンパス) にて開催されます。本学会は、IMPS-2001 の協賛学会になっておりますので、参加費については「会員扱い」となります。

次ページに、IMPS-2001 の情報をご案内いたしますので、ふるってご参加下さい。最新情報につきましては、IMPS-2001 のホームページ (<http://www.ir.rikyo.ac.jp/imps2001/>) をご覧下さい。

International Meeting of the Psychometric Society (IMPS-2001)

From July 15(Sunday) through July 19 (Thursday), 2001
Osaka University Convention Center, Osaka, JAPAN

Meeting schedule

July 15 (Sunday)	Evening: Welcome reception
July 16 (Monday)	Sessions
July 17 (Tuesday)	Sessions
July 18 (Wednesday)	Sessions PM: Excursion to Kyoto Evening: Banquet in Kyoto
July 19 (Thursday)	Sessions
July 20 (Friday; holiday)	Optional tour to Nara

Registration fees

	On or BEFORE May 1, 2001	AFTER May 1, 2001
Members*	20,000 Japanese Yen	25,000 Japanese Yen
Non-members	25,000 Japanese Yen	30,000 Japanese Yen
Students**	10,000 Japanese Yen	15,000 Japanese Yen

(Currency Conversion by Yahoo.com)

* include members of the societies sponsoring IMPS-2001. (We, JLTA, are one of them.)

** A copy of student ID is required upon registration for Student registration.

Conference site

Osaka University Convention Center (about 20km from the downtown Osaka)
One hour from the bullet train station (Shin-Osaka) by subway and monorail
Thirty minutes from Osaka (Itami) Airport (domestic) by monorail
Two hours from Kansai International Airport by train and monorail.

Registration and Accommodation

Please find information on website <http://www.knt.co.jp/intltyo/imps/>.

Correspondence

National Center for University Entrance Examinations
2-19-23, Komaba, Meguro, Tokyo, 153-8501 Japan

E-mail: imps@rd.dnc.ac.jp, Fax: +81-3-5478-1297
Haruo YANAI, Chair of the Organizing Committee

*For further information, visit the website at <http://www.ir.rikkyo.ac.jp/imps2001/>.

≡≡

学 会 短 信

1. 学会名称の変更について

2000年の総会報告でお知らせいたしましたが、2001年4月より、当学会の日本語名称は、外国語教育評価学会から、「日本語テスト学会」に変更されます。

会員各自において必要と思われる事務手続き等についてご注意ください、ご準備下さい。郵便物の送付元、会費の納入等については、事務局で変更の手続きを進めておりますので、ご承知置き下さい。

2. 会費の納入について

2000年度の会費を未納の会員宛、昨年の暮れに、郵便振り込みの用紙を送付させていただきましたが、それぞれお手元でご確認いただき、未納の会員は大至急納入して下さい。

3. 住所・所属・銀行等の訂正・変更

名簿の訂正、あるいは住所・所属・メールアドレスなどの変更が必要となった会員は、正しい、あるいは新しい住所・所属を、事務局までご連絡下さい。

また、銀行振込で会費を納入している会員の中で、合併・吸収等により、銀行名あるいは口座番号等の変更がある場合には、至急、事務局までご連絡下さい。

4. その他

JLTAの活動に対するご意見やご要望、Newsletter等への掲載希望記事などがありましたら事務局までお申し付け下さい。

より充実した活動ができるよう、皆様のご協力をよろしくお願いいたします。

≡≡

外国語教育評価学会事務局

〒389-0813 長野県埴科郡戸倉町芝原 758

TEL 026-275-1964 FAX 026-275-1970

e-mail: youichi@avis.ne.jp

URL: <http://www.avis.ne.jp/~youichi/JLTA.html>

