

JLTA Newsletter

外国語教育評価学会

The Japan Language Testing Association

JLTA Newsletter No. 5 発行代表者：大友 賢二 2000年(平成12年) 1月15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>



2000年における3つの願い

JLTA 会長 大友賢二 (常磐大学教授・筑波大学名誉教授)

1999年、JLTAにとって画期的な、しかも、最も記念すべき年の1つになりました。13カ国から198名の参加者を得た、つくば国際会議場でのThe 21st Language Testing Research Colloquium (LTRC 99)の開催を成し遂げたからです。生まれてまだ3年という若いJLTAが、わが国の外国語教育の歴史に残るこの大事業をなしえたのは、多くの方々のご指導とご協力の賜物と深く感謝するものであります。この会議のテーマであった「21世紀における言語テストの社会的責任」をわが国で果たせるかどうかは、まさに、私どもJLTAのこれからの活動にかかっています。

2000年のJLTAの研究活動を開始するにあたって、私には3つの願いがあります。第1は、わが国での言語テスト研究の地固めです。第2は、テスト研究分野に関する私たちJLTA会員の視野の拡大です。そして第3は、JLTA会員同志のますますの情報交換と親睦を図ることです。

第1のわが国の言語テスト研究と実践の地固めのためには、本学会設立後4年目を迎えるので、いよいよ、研究・実践活動の目標設定を始めなければなりません。いま、わが国の外国語教育の中の測定と評価の分野で、何がもっとも重要かということの検討です。多くの課題の中からの厳選作業が必要です。そして、その選ばれた課題を解決するための研究部会を結成することも必要でしょう。

これと平行して行わなければならない第2のことは、JLTA会員のテスト理論に対する視野の拡大です。LTRC 2000でも取り上げられる「構造方程式モデル」(Structural Equation Model)、speaking や writing などの performance assessment で必要な「一般化可能性理論」(Generalizability Theory)、テスト得点の抜本的再考に必要な「項目応答理論」(Item Response Theory) などの研究はどうしても必要です。このテスト理論の急速な発展を言語テストではどう対処するかを考えることです。新しいテスト理論の動向を知るための国際的なワークショップの開催は視野拡大のための一つの方向でもあります。

第3は、会員間のますますの情報交換が必要です。そのためには、誰が、どんな研究をしているのかを知ることが大切です。たとえば、JLTA会員の研究業績等をまとめること、言語テスト研究のための海外留学・在外研究に関する情報交換を行うこと、どの大学ではどんな研究分野の人を求めているか等に関する情報を交換すること等も考えられます。それが、JLTA会員の親睦につながり、ひいては、わが国の言語テスト研究のさらなる発展へと結びつくことを切に願っているものです。

第3回全国研究大会・第21回言語テスト国際会議レポート



1999年7月28日～7月31日、外国語教育評価学会の第3回全国研究大会、及び第21回言語テスト国際会議が、茨城県つくば市の、つくば国際会議場で行われました。以下は、各セッションのレポートです。

《Wednesday, July 28th》

JLTA Annual Conference

Plenary Speech One: Hiroshi Ikeda

第21回言語テスト国際会議の直前に行われた、第3回外国語教育評価学会全国研究大会で、池田央氏（立教大学名誉教授）が「What we need for research on language testing in Japan - A psychometrician's view (外国語試験の研究において何が重要か - ある計量心理学者の視点から)」という基調講演を行った。講演に先立ち、司会者から、池田氏がこれまでにテストに関して20冊以上の著書を執筆されているテスト研究の第一人者であること、また、最近、20世紀の初頭から21世紀にかけて、10年ごとの区分を設けて世界のテスト研究がどのように進展していったかという興味ある論文を書かれていること等の紹介があった。

池田氏は講演の中で、外国語試験の研究を推進していくためには、

- 1) コンピュータ技術を含む近代技術の利用（現行の紙筆検査をコンピュータ検査に変革させること）
- 2) 近代テスト理論（項目反応理論、一般化可能性理論）の適用
- 3) データベースおよび項目バンクの構築
- 4) 妥当性研究の推進、および信頼性の有無の判定に関する標準的目安の確立
- 5) 研究者、外国語担当教師、テスト作成者間の交流の促進、

等が必要であることを力説され、さらに、外国語の試験を受けることなくして、言語能力は進歩しないこと。さらに、テスト作成は芸術ではなく、むしろひとつの技術で個人の手に負えるものではないこと。テストに関する情報はむしろ積極的に公開して大衆の共有の知識とすべきであること。こういった諸改革が実現できれば、現行の外国語教育が改善されるであろうと述べ、テスト研究は一つの科学であることを強調して、講演を結ばれた。

(報告者: 柳井晴夫 大学入試センター)

Invited Paper: Hossein Farhady

Prof. Farhady introduced his talk by explaining that constructs in language testing theory formerly based on structuralist linguistic models have been challenged and

become more informed by Oller's insights into the integrative nature of language, notional-functional models and Cummin's work on the contexts of language use on through to the work of Bachman on the construct of communicative competence. Farhady observes, however, that proficiency remains "a vague construct" and that even if the construct were known, there remains the job of operationalizing and measuring it as test users continue to struggle with the relationships between grammar, tasks and performance, of the choice to use one test or several, of quantitative vs. qualitative assessment, and of norm referenced vs. criterion referenced testing. There has been some progress, he states, but "not much change in practice", which has questionable consequences on educational goals and educational equity.

Prof. Farhady illustrates this with a study conducted with 200 subjects who were pretested with the TOEFL, which is the most widely used instrument for the purpose of assessing the ability to use English in academic settings today. Half the subjects then received instruction which featured TOEFL-like tasks while the others were instructed in explicit rules for reading, writing, vocabulary and structure. Both groups were given an additional task-based examination (IELTS) along with the TOEFL as post-tests. The results showed an average gain by those who had received TOEFL coaching of 73 points on the TOEFL with significant gains among those who had scored from 350-400 on the pretest, but little gain among those who had scored 500 or above. Many with over 500 on the TOEFL had problems finishing the IELTS reading section in the allotted time, and the correlation between the total scores on the two examinations was only $r = .27$ with the greatest differences in the respective reading sections ($r = .48$). The correlation between the two reading sections was less than that between the reading and writing sections of the TOEFL ($r = .72$).

Prof. Farhady pointed out that when there are cultural differences in how education and testing interrelate and students are motivated by "beating the test", the impact of using the present form of the TOEFL rather than task-based

assessment can lead to truncated ability among certain groups to perform the tasks required of them in the real world, such as when studying abroad.

Plenary Speech Two: Nancy Cole

Nancy Cole characterized language assessment of the past and today as aimed toward limiting context to the general and balancing content, as primarily for the purpose of ranking students rather than diagnosing weaknesses, and that recognizing correct usage or explicit and implicit information from written text in test booklets has been the mainstream paper-pencil feature, with other skills being assessed through supplementary procedures. She explained that advances in language theory that recognize the various modes of speech, listening, reading and writing as integrated in language use, the relationship between language and context, and "language use in different life spheres", have all led to a shift in the target of assessment away from language structure toward language use.

The limitations of past technologies made it difficult to measure these new dimensions, Dr. Cole explained, but new assessment technologies now offer opportunities to assess speaking, writing and listening tasks integrated with each other or in isolation. Computer-assisted testing (CAT) has enabled tests to present tasks and record writing or speech electronically. There are new possibilities for simulating varying contexts as well as the ability to branch from one task to another.

Beginning with the TOEFL 2000 project, ETS has been actively exploring these new avenues, and will continue to pursue what she calls "evidence-centered design". This approach to test design would first work with test-users to determine the inferences they need to make, such as differences between scores e.g. 550 vs. 580 or where additional instruction might be needed, and then use new technologies to build tasks that would supply convincing evidence to support those inferences. In the discussion that followed the presentations several points were raised by members of the audience, including question of how accessible the test would be, the dangers of going off the "relativistic deep end" when individualizing tests risks making the results difficult to generalize, of surrendering test constructs to test users without rigorous scientific and theoretical basis, and of attempts to amalgamate the essentially different natures of norm-reference and criterion-reference test development. Dr. Cole predicted that accessibility would grow dynamically

by the year 2010. As to the other points, she stated that during this ongoing process of change, ETS is "getting on it" with all its resources and its work would move toward making assessment a more integral part of the learning process.

(Reporter: Jeffrey Hubbell Hosei University)

#####

LT RC 9 9

Plenary Speech One (Messick Memorial Lecture):

Tim McNamara

Samuel Messick 氏はアメリカの ETS (Educational Testing Service) でかつて活躍した人であり、今回の LT RC には来日が予定されていたが、1998 年半ばに突然亡くなられた。そこで ILTA の前会長である Tim McNamara 氏が追悼講演を行った。

講演の概要は会議要綱の p. 11 に記されているが、要約すると次のようになる。

Messick 氏は 2 つの点で言語テストの研究に影響を与えている。1 つはテスト結果の解釈の仕方であり、もう 1 つは妥当性 (validity) という概念を打ち立てたことである。さらに、Messick は妥当性の考えを行動科学や社会科学と結びつけて、認知的に考えようとした。

当日配布された印刷物には講演の資料が載せられていたが、その 1 つに Messick が、法廷における裁判と学力評価との比較をまとめた表がある。だれが評価に関係するかは、裁判では裁判官や陪審員、学力評価では研究者やテスト作成者、決定の根拠は、裁判では種々の証拠や法廷での論争、評価ではテストの成績とその分析としている (あとの項目は省略)。

Messick のテストに関する考えが具体化されたものとしては、オーストラリアにおける各種のテスト制度、カナダの Benchmarks Project、EC のスケール化の試みなどがある。

また、Butler, J (1990) の著書の中の、社会での性の形成は、人間に内在するものが外面的な行為となって表れ、その評価が社会における役割を作り上げる、という説にも影響が見られる。

Memorial Speech というのもあつてか、McNamara 氏は原稿を読んだだけで、自分の言葉でしゃべることが殆ど無かったため、分かりにくかったことは否定できない。しかし、氏はスライドを見せるときは、マイクから離れるので話しをせず、こういう場面での話しにいかにも慣れていているという感じであった。また、用意してくれた印刷物には要点がたくみにまとめられているので感心させられた。

(報告者: 羽鳥博愛 文教女子短期大学)

《Thursday, July 29th》

Student Research Reports:

Yoshihiro Sugita

Sugita presented his research on constructing rating scales for EFL writing tests. He explained how he and his co-researchers developed assessment procedures that would elicit discourse.

Sugita noted that standard rating scales require the matching of examinee performance to a verbal description. Thus, in spite of the fact language is assessed in discourse, there is no specific focus on discourse features. Sugita described the analytical test of organizational ability that they developed and discussed the results in terms of strategic discourse planning.

Sugita further offered reasons to explain why this type of scale may improve reliability and validity problems associated with standard rating scales.

Dongil Shin

Shin discussed validation of a diagnostic ESL test for measuring lecture comprehension ability and explained that his model was operationalized to diagnose students' weaknesses and strengths in a lecture comprehension test. He reported that four types of validity evidence argue for and against inferences that variance in test response can be attributed to lecture comprehension ability: content evidence, empirical item analysis, correlational research, and internal test structure. Shin asserted that this study offers a principled approach to theory, test construction, and validation study in ESL listening.

Amy Yamashiro

Yamashiro considered the use of structural equation modeling to validate a rating scale. She explained that the rating scale which was developed as the criterion-referenced assessment of the EFL speaking course objectives has three traits: non-verbal delivery, verbal delivery, and organization/purpose. Peer and self-assessments were included in the course design to promote deeper understanding of the course objectives. Yamashiro used structural equation modeling to analyze the performance assessment to determine the extent to which teacher, peer, and self-assessment may be reliable and valid. Yamashiro further discussed the study's limitations and implications.

(Reporter: Matsuo Kimura Aoyama Gakuin University)

Panel Discussion:

“Language Testing: Yesterday, Today and Tomorrow”

Bernard Spolsky, Elana Shohamy, Alan Davies

言語テストの過去、現在そして未来を取り上げた本パネル・ディスカッションでは、まず Spolsky 教授が前世紀から今日に至る間のテストが担う意義を取り上げた。19 世紀の西欧では、テストは学校教育制度及び社会における人材選抜の手段として重視されたが、今世紀に入り、統計学の発達とともにその方法論上の欠陥が指摘され、信頼度の向上に力が注がれるようになった。このことは言語テストについても当てはまり、今日ではその使用上の倫理責任まで視野に入れる必要があるのではないかと指摘がなされた。

Shohamy 教授は、今日言語テストが抱える理論上、研究上、及び実施上の諸問題を分析し、その有用性を検討した。そのために、言語テスト及び関連分野の代表的研究者 (M. Swain や A. Cohen など) 10 名の見解を対象に考察し、言語テストのあるべき姿を追求しようとした。

最後に発言した Davies 教授は、一般に物事の未来を予言することに伴う限界に触れ、同じことが言語テストについても当てはまることを指摘した。言語テストの中心的課題であるその目的と使用に関して、雑誌 *Language Testing* の既刊分 16 巻を概観しても、その根底にある言語観、言語能力の定義に見解の一致を見いだすことができない以上、今後予想されるコミュニケーション工学の発達にもかかわらず、言語テストの未来に関する見解の多様性を予測した。

本パネル・ディスカッションの発題者は、それぞれ ILTA の前会長、現会長及び次年度会長就任予定者であり、また司会は JLTA 会長が担当したもので、参加者との質疑応答と相まって、内容の充実度の点でも今大会の中心的プログラムとなった。

(報告者: 浪田克之介 北海道大学)

Research Reports Session One:

このセッションでは 4 つの研究発表があった。

1. Developing and Cross-Validating a Set of Cognitive and Linguistic Attributes Applicable to Multiple Test Forms on a Multiple-Choice Listening Test (ETS, Gary Buck など)

これはルールスペース (rule-space) 法によって抽出される認知特性と言語特性の一般可能性の研究である。The Test for International Communication の聴解テストの 5 つのフォームから抽出された特性と他のフォームとの妥当性を検証し、一般化可能な特性が抽出されたとする。

2. Analyzing the Dimensionality of a Second Language Reading Test from Cognitive Perspectives (Mission 大学、Hisami Saito-Scott など)

これは項目応答理論の前提としてのテスト項目の一次元性を言語テストに適用するにあたっての課題を研究したものである。ルールスペース法を使って TOEFL テストの読解テストを分析し、多次元的な言語テストに対して一次元的項目応答理論が働く条件を検討した。その欠課、24 の次元が読解テストに見られること、学習者の反応はルールスペースの複数の認知経路に分類できるなどが分かったとした。

3. Validation of Test-Dependent and Task-Independent Ratings of Performance Assessment (Hawaii 大学、J D Brown など)

タスクの困難度に基づいて開発されたパフォーマンスタスクを測定する基準としてタスク依存度とタスク独立性による方法をすでに提案しているが、この研究発表ではその測定基準の妥当性を 90 名の学習者から得たデータを分析したものである。

4. Examining Passage Related Local Dependence (LD) Using Q3 Statistics in an EFL Reading Comprehension Test (ETS, Yong-Won Lee)

これは読解テストの局所依存性の研究である。1857 名の韓国の高校生を対象とした読解テストの 40 項目を Q3 法を使って分析し、文章形式の読解テストにはかなりの程度の局所依存性が存在すると報告している。

(報告者: 和田稔 明海大学)

Research Reports Session Two:

Catherine Burrows

オーストラリアの AMEP (Australian Adult Migrant Education Program) では、新たに「授業観察に基づいた評価システム」classroom-based assessment system を採用することとなった。本調査はこの新システムの「波及効果」washback を考察したものである。授業観察及び面接によりデータを収集し、数量的分析、及び定性的分析を行った。その結果、同一の評価システムを採用しているにも関わらず、その波及効果に関しては教師による個人差が大きく、必ずしも一様でないことがわかった。この評価システムに対する教師の態度は大きく三つのタイプに分れるようである。すなわち評価を規定通り忠実に実行しようとする adopters、自分なりに改変して採用している adapters、意識して影響を受けないようにしている resisters である。この結果から、テストは教育現場に一律の結果をもたらす、といった「刺激 - 反応モデル」stimulus-response model

は変更を強いられることになる。より現実に近いモデル(「カリキュラム革新モデル」curriculum innovation model)は、一方にテスト・評価システムを置き、次にこれを使用する教師のピルーフや知識等があり、それを仲介要因として、多様な教育上の効果をもたらされるというものである。

同主旨の調査はこれまでスリランカ、アメリカ、香港、日本などで行われてきたが、今回の調査はこれらと同様の結果を示しているのは甚だ興味深い。すなわち、信頼性、妥当性に優れたテスト・評価システムを使うだけでは必ずしも望ましい教育効果は得られない、ということである。今後はテスト・評価システム以外にどのような要因がどのように関わっているのかを明らかにすることにより、より有益な情報が得られると期待される。

Jane Andrews and Richard Fay

現在、英語教育・学習が対象としているのは「外国語としての英語」EFL/ESL ではなく、「国際コミュニケーションの手段としての英語」ELWC である。この傾向は今後ますます強まることであろう。ここで特に重要になるのは、口頭試験の結果を評価する際、特定の変種の英語に偏らない基準を採用するということである。本調査の最終目的は、試験官が採用している基準を明示するところにあるが、今回は目的達成のために最も効果的な調査手続きを探ることにあった。調査は 3 段階に亙って行われた。第一段階では、試験官に Paired Oral Test (非母語話者同士に会話を行わせたもの)の録画ビデオを見せ、評価のプロセスを think-aloud してもらった。その結果、被調査者は think-aloud の手続きについて戸惑うことがある、途中調査者が干渉すると自然な思考過程の妨げになる、受験者同士のやりとりで注意が向くあまり個人の受験者の評価が疎かになることがある、などの問題点が挙げられた。第二段階では特にこれらの問題を解決すべく詳しい指示を与えた。しかし新たな問題として、通常の評価過程ならば行わないようなことを行っていることがある、という指摘があった。そこで、第三段階では、なるべく試験官の自由度を増し、ビデオを自由に操作してもらうようにした。評価の基準に関して、有益な結論を導き出すまでには至らなかったが、think-aloud は試験官が評価を行う際のプロセスを検証するに極めて有効であり、多くの有益な情報が得られるとの確信を得た。

更に、試験官同士の paired think-aloud を採用する、受験者に面接調査を行う、試験官の文化的背景を明確にする、などを考慮することによって、より実り豊かな成果が得られるものと期待される。

Laura MacGregor

「英検」の受験者数は今や約 3 百万、極めて大規模なテストとなっている。しかしながら、テスト開発のプロセス、採点の基準などに関する情報は入手しにくく、受験者、使用者、作成者同士の情報交換の機会も極めて限られているのは遺憾である。本研究は特に面接試験の問題点を明らかにしその改善策を探った。1998 年 7 月に行われた 3 級、準 2 級、2 級の二次試験（口頭試験）に参加した 3 つの試験場の受験者及び試験官にアンケートと面接調査を行った。以下はその結果と考察である。第一にウォームアップの質問に関して、試験官の半数は有効である答えているものの、特に 3 級の受験者の場合は反って緊張させてしまっているようだという意見もあった。受験者は過半数が満足しているという回答であったが、むしろ緊張が高まった、本テストと関係がはっきりしない、などの批判もあった。より効果的な質問の条件を探る必要がある。第二に、試験に関して、試験官からは受験上の指示 “Please look at the passage.” について、passage を知らない受験者も多いので変えるべきであるとの要望があった。受験者は全般的に満足していると回答したものの、本テストの後に自由会話をし、それも考査対象とすべきである、などの提案もあった。第三に、テストの採点に関して、試験官からは最高点の 5 の解釈、satisfactory などの主観的な評価尺度の解釈などについて疑問が出され、受験者からも採点基準が必ずしもはっきりしないとの批判があった。採点の尺度をより明確にし、できれば事前に公表する必要があるかもしれない。

英検は多くの受験者の将来に影響を与えうる。このようなテストに関してはできるだけ多くの情報を公開することが望ましい。本調査は特に面接試験の妥当性、信頼性を高めるために有益な調査であった。

Junko Matsui

語彙力と背景知識は英語の聴解を促進するか、この疑問を解決すべく行ったのが本調査である。調査対象は中級 intermediate、準中級 lower intermediate level の英語学習者。語彙に焦点を置いて指導した被験者のグループ、背景知識に焦点を置いて指導した被験者のグループ、それぞれにディクテーションと翻訳を行わせ、結果を比較検討した。分析の結果、聴解には背景知識よりも語彙が聴解力に大きな役割を果たしていることがわかった。しかしながら、更に TOEFL のリスニングテストの結果を考慮したところ、聴解力に優れている被験者と、劣っている被験者との間に差が見られた。すなわち前者の場合は語彙及び背景知識を得ることで聴解は促進されるが、劣っている被験者にこの

ような傾向は見られなかったのである。聴解には、音声言語の速度をフォローする能力、個々の音声をはっきりと聞き分ける能力、音声の弱化 reduction、連結 linking、同化 assimilation などを聞き分ける能力、子音と母音を聞き分ける能力などが必要であり、これらの基礎聴解能力があつてはじめて語彙、背景知識が有効になるのであろうと推論される。

テストから得られたデータに加え、語彙及び背景知識に関して実際にはどのような指導が行われたのか、ディクテーションテストに使われた英語の特徴（速度、発音等）などに関する情報も考察の対象とすることによって、より有益な情報が得られると期待される。

Tetsuhito Shizuka

多岐選択問題は 0 か 1 か、すなわち all or nothing を基本とする。したがって、応答に際して、受験者がどのような判断を下しているのかは得点に反映されないのが普通である。しかし、実際には、解答を確信している場合 know、恐らくこれが正答だろうと判断して答えている場合 think、当てずっぽうで答えている場合 guess、などのケースが考えられる。能力を正しく反映しているのは、単なる推測で正答を得た場合よりも、確信して正答を得た場合だと言えるであろう。受験者の能力を正しく測定するためにはこの点を考慮すべきである。

本調査では応答において受験者の確信性 confidence がどう関わっているかを考察した。解答と確信性の両者を考慮した採点システムに基づき、更に予備データの分析結果を考察した。1/0 応答と 3/2/1 に互る確信の度合いを示す尺度 scale を採用し、計 6 ポイントに互るデータを得た後、高い確信度をもって正しく応答した件に関しては高い得点を与えた。結果、確信性を考慮した場合はそうでない場合よりもより高い信頼性を得ることができた。更に risk-takers と risk-avoiders を想定し、妥当性をも検証できることを示した。

どのような受験上の指示 — 推測を促すのか、減点の対象とするのか — などの要因を考慮する必要があるだろう。また確信性という要因には性別も関与しているかもしれない。これらテスト項目以外の要因も考慮することによってより実り豊かな成果が期待できる。

(報告者: 渡部良典 国際基督教大学)

《Friday, July 30th》

Poster Session One:

Poster Session One took place on July 30. Starting at 8:15 a.m. in the convention hall, nine individual and group presenters spent maximally five minutes respectively to outline their projects, followed by a one-hour poster

presentation and discussion held in a smaller conference room. The topics covered in this session were roughly the issues on (1) innovative ESP or EFL tests (Briggs and Dobson, Wylie, Nakamura), (2) performance or communicative tests (Halleck and Reed, Mochizuki and Yamada), (3) Japanese Language Proficiency Tests (Noguchi, Saegusa et al.) and (4) qualitative analyses of examinations and washback effects (Huhta et al., Watanabe).

Many of the presenters looked somewhat nervous during the five-minute presentation time. Some presenters made a good, concise summary of their projects, while those who went into specific details were constrained by the time limit. As a timekeeper, I was in a dilemma, having to show two or three time-warning notes to grim-looking presenters and wishing I could allow them an extra few minutes!

In contrast to the first part of the session, a more relaxed and lively atmosphere prevailed in the following poster presentation and discussion time. Unlike similar sessions in other conferences that I have seen, this two-part poster session was placed in a specific time slot, set apart from other types of presentations, such as research reports and symposiums. Thus, in the second part of this session, with a lot of time reserved for presenters and audience alike, all the participants were actively and uninhibitedly engaged in exchanging their views and ideas on language testing and assessment.

In summary, Poster Session One has successfully informed us of language testing and assessment projects that are in progress in various sociocultural contexts. It has ascertained that those involved in language testing should take account of resources available to innovate their practices, different sources of variance that affect language test performance, the need to develop new test methods to meet their social or educational demands, findings of testing research beyond their immediate linguistic and social concerns, and roles of qualitative analyses as well as quantitative analyses, just to name a few. The issues are far more complex than they appear to be, and the overall picture shared by language testers still looks dim. However, as Bachman (1990) notes, "perhaps we have benefited from the confrontation with these issues, and will be able to address more wisely those that lie ahead" (p.355).

(Reporter: Ken Norizuki *Shizuoka Sangyo University*)

Research Reports Session Three: (Cancelled)

Plenary Speech Two: Ikuo Amano

"Admission Policies in Japanese University"

OECD では大学入試制度を社会的出生 (Social Birth) と位置付けているが、入試制度に関する限り日本は世界の中でも先進国であり、それだけ入試制度が厳しいと言える。しかも大学入試においては英語は、最重要科目の一つである。

入試制度について考えるとき、日本では近代的な国家改造以後の問題として考える必要がある。そもそも筆記試験の始まりは、7世紀中国の科挙 (The Keju) の制に遡ることができる。日本にも科挙の制は中国文化の一つとして入ってきたが、定着しなかった。この制度をヨーロッパに最初に伝えたのは、16世紀に中国で布教活動をしていたイエズス会の宣教師マテオ・リッチ (Matteo Ricci) であった。官僚の任用試験であった科挙の方式は、18世紀のドイツにおいては大学への入学試験の方式 (Abitur) として採用され、19世紀のフランスにおいてはナポレオンの教育改革の一つとしてグランデコール (grandes école)、後のバカロレア (baccalauréat) の制度として生まれ、いずれも大学入学資格試験として位置付けられた。この制度が1870年代に日本に入り、義務教育以後の入学試験として使われるようになったのであり、近代化の遅れを取り戻すための手段の一つとして使われるようになった。

問題は、欧米においては入学試験の他に卒業試験等の資格試験が課せられたのに対して、日本では入学試験が高学歴を獲得するための手段となり、大学入学後に資格試験等が課せられなかったことである。これは最近のように定員割れが生じたり、入試制度が多様化したりすると大きな問題となる。今後の課題としては、入試制度そのものの質を変え、資格試験等の導入が考えられる。

フロアからの質問は、1) 大学における男女の比率、2) マイノリティの扱い、3) 入社試験の実態などの3点についてであった。

(報告者: 斉藤誠毅 *神奈川大学*)

Symposium One:

"National Language Testing Reform: Possibilities and Constraints"

カリフォルニア州立大学のクーン教授の司会のもとで、定刻に、シンポジウムが始められた。オーストラリア、ハンガリー、香港、イスラエル、中国、エジプトにおいて、現在どのように言語テストに関して改革が進められているか、どのような問題点が生じているか、今後の方向について発表が行われた。

タイトルと発表者 (敬称略) は下記のとおりである。

1. Issues and Problems in National Language Testing Reform (in Australia) (Brindley)
2. Language Testing Reform in Hungary (ランカスター大学の C. Alderson 教授が欠席のため、Brindley 教授によって、原稿が読まれた)
3. Language Testing Reform in Hong Kong (Bachman, Cheng, Hamp-Lyons)
4. Language Testing Reform in Israel (Spolsky)
5. Setting up a dynamic language testing system in national language test reform: the Public English Test System (PETS) in China (Yurong, Taylor)
6. A Training Model for Egyptian Language Testers (Kunnan, Kawamoto)

このシンポジウムへの参加者は、110 余名に達する。2 時間の中で、6 つの発表が行われたため、各発表者に与えられた時間は 15 分 (3 人の発表者の場合は 1 人、5 分) であったが、その中で、熱のこもった発表が行われた。4 つめの発表の後と最後に、活発な質疑応答が行われた。参加者は、それぞれの国で改革に取り組んでいる実態を、如実に実感出来たと思われる。

(報告者: 山崎朝子 武蔵工業大学)

Research Reports Session Four:

Charles Alderson

このリサーチ・レポートは、発表者の Charles Alderson が参加できなくなったために Caroline Clapham により、代読された。司会は Janna Fox である。

このリサーチは、ハンガリーのテスト改革の一部として、リーディング能力の測定のために用いられている「並べ替え問題 (sequencing)」について、実証的に研究したものである。

発表の中では、文の並べ替え問題について、実際の答案例を見ながら、原文以外にも、意味の通る順番が存在することを示し、採点方法による得点の違いを見た。採点方法は、exact match (exact) と classic と added-value (addval) と full pair である。

さらに、このリサーチでは、複数の採点方法による違いをデータの点から分析している。全体的能力との相関では、exact > classic > addval > full pair の順で相関が高い。また、採点方法間の相関では、classic : addval > exact : classic > exact : addval の順で相関が高いが、いずれも相関係数は .79 から .99 と非常に高い。したがって、exact match が簡便で、他の採点方法とも大きな違いがないことが分かった。

(報告者: 根岸雅史 東京外国語大学)

Poster Session Two:

このセッションでは、発表者が各自 5 分間の持ち時間で研究の要旨を発表し、その後、別の会場で一堂に会して、発表者は各自の所定の場所に位置し、参加者からの質問や意見の交換が行われた。わずか一時間たらずではあったが、会場は発表者と参加者との真剣かつ熱心なやりとりで「熱い」雰囲気にもまれていた。発表者と個別に話が出来るのがこのポスターセッションの特徴であろう。

それぞれの内容は：

- ・言語テストコースでの実践部分としての口頭発表能力の評価法の開発
 - ・日本の大学入試問題の信頼性・妥当性
 - ・日本と韓国における大学入試問題の相違がそれぞれの国の高校生の英語の成績に及ぼす影響、彼らの英語学習に対する態度や聴解力テストの結果の相違の要因
 - ・日本人学習者の TOEFL セクション II (文法) の成績の分析・検証
 - ・Rule Space Methodology による日本語語彙テストの検証
 - ・ノートパソコンを使用した日本語ディクテーションテストの開発
 - ・日本の大学生 (初級・中級レベル) のオーラルコミュニケーション能力診断テスト
 - ・Guam Educators' Test of English Proficiency (GETEP) の開発
 - ・ビジネス用英語ライティングテストの開発とその採点基準
- などで、多彩で興味深いものであった。

(報告者: 智原哲郎 大阪女学院短期大学)

《Saturday, July 31st》

Research Reports Session Five:

わたしがこのような会議に不慣れで、関係者にご迷惑をおかけしましたが、Chair の Ms. Elaine Wylie の機転によって、円滑にかつ成功裏に発表が行なわれたことをご報告申し上げます。原因は私がタイムカードの存在をセッションの前に確認しなかったことにつきます。そのために、最初の発表者の Dr. John Read の途中になってから、Chair と Co-chair の 2 人がタイムカードの必要性に気づきました。そこで、プラスチックケースを確認すれば良かったのですが、それをせずに地図の裏の白紙の部分に、残りの時間を書いて発表者に示すことにし、ことなきをえました。Chairs が発表者の右手にいますので、OHP の右側にいて熱心に説明する発表者にとっては、どちらかという背中側に

あたり、Chair がタイムカードを示しても、なかなかそれに気づいてくれませんでした。すると、発表者のまんにいた男性の参加者が指で数字を示し、発表者に残りの時間を教えてくれました。興味深かったのは、発表時間が過ぎてしまった場合に、自分の首の部分を手で切るしぐさで「時間切れ」を示したことでした。

John Read

最初の発表は、NZ の移民政策についてでした。内容は、英語が話せる移民を引きつけるために、英語のテストを行なっているが、現行のテストのせいで、現実的に移民が減っているのです、その代替案をしめしたものでした。このテーマは非常に興味深いもので、将来、日本が日本語が話せる移民を受け入れるようになる段階で参考にしなければならないテーマであったと思います。

Jared Bernstein, Maxine Lipson, Gene Halleck, Martinez-Scholze

次の発表は、スピーキングテストの併存的妥当性についてのもので、日本のテスト研究でも今後研究の焦点が向けられるところである。日本では OPI テストについてはよく紹介されているが、Phone Pass Spoken English Test, 10-minute form についての説明があり印象的だった。

(報告者: 酒井志延 桜の聖母短期大学)

Symposium Two:

“The Rights and Responsibilities of Testers and Test-takers in Language Testing Settings: Ethics, Policy, Practice and Research”

今回の国際会議を締めくくるシンポジウムは最終日の午前中に行われた。

まず Janna Fox が総括的導入を行い、続いて 7 名の発言者がそれぞれの立場から提言した。最後に司会者である Bernard Spolsky がまとめのコメントを読み上げた。

最初の発言者である Elena Shohamy は、彼女の critical language testing という視点からこの分野に関するこれまでの研究を 5 つのアプローチのもとにまとめ、テストがどのような使われ方をするかを認識することが重要であるとした。

2 番目と 5 番目の発言者である Tim McNamara と Cathie Elder は、いずれも Samuel Messick の議論を受け継ぎ、その後の研究の進展に沿って (例えば communicative competence の概念の改訂)、テストの妥当性に社会的影響力という側面をより反映させた新たな価値観が必要であると主張した。また、テストに関わる者によって判断基準やニーズが異なり、その結果

テストの意味が変わるのなら伝統的な心理測定法に基づいた妥当性の診断には限界があると指摘した。

3 番目の発言者 Geoff Brindley は、テストに関わる様々な人々 (テスト開発者・テスト実施者・受験者・親・行政当局など) の間で意志疎通を高め、“assessment culture” を向上させるべきだと提言した。これに関連して 6 番目の発言者 David Nevo は、テスト実施者とテスト受験者の間の対話と相互理解を通して、評価者が負う責任分野である (テストの) accuracy, utility, propriety, feasibility が達成されると考えた。

更に 4 番目の発言者 Janna Fox は、特定のテストに対する受験者の評価がその能力により異なるという研究を報告し、受験者に公平であると意識されるテストを開発することが重要であると主張した。

最後の発言者 Alan Davies は、テスト実施者の倫理条件を考えた。組織の中で倫理に悖る行為をする者には法的な制裁が加えられるべきであるが、それが存在しない状況下でも組織構成員としての自覚や同僚の「目」の意識といった倫理規範を高めることで社会的責任の達成に繋がると述べた。

このようにシンポジウムの内容は多岐に亘り、発言は個別のテスト作成の問題からテスト作成者の職業倫理や政策決定の問題にまで広がった。Fox や Nevo が主として実際のテスト作成環境を取り上げたのに対し、Brindley, Elder, McNamara は誰にどのような社会的責任があるかを考えようとした。Shohamy が今後の研究の指針を提唱したのに対し、Davies は今後の実践のあり方を考察した。

このような多様性あるいはまとまりのなさは第 19 回 LTRC (1997, Orlando) で見られた盛り上がりとは対照的である。このときは一元的な評価から学習者の個別の適性・能力・環境に応じた多面的な評価にパラダイムが移行し、CBT や structural equation modeling がそれを具現化するものとして期待された。今回の社会的責任というテーマは Fox が言うように、fairness 追求の当然の帰結である。その結果 Spolsky が最後に指摘したように我々テスターの本務からかけ離れた議論になりかねない。実際、会場から「もっと内容の質向上に身を入れるべきではないか」(Bachman) との意見も出て、McNamara との間で興味深い議論の応酬が始まりかけたが時間切れとなった。

私見では、日本の実状に基づいた研究や報告がなされるべきであると思った。多くの語学教師がカリキュラムなどの意志決定に参画し得ない状況では、責任の取り方そのものに問題がある。さらに、特定のテスト (例えば国内最大の英語検定試験) がテストの社会的意味づけに及ぼした功罪も検証されるべきであろう。

第 11 回 研 究 例 会 報 告

JLTA 第 11 回研究例会が、1999 年 11 月 23 日(火)、北海道札幌市の、北海道大学言語文化部で、JACET 北海道支部との共催により開催されました。

講演 大友賢二 (常磐大学)

「言語テストと言語教育—テスト得点からの視点」

諸外国に比べ日本は言語教育の中でとりわけ言語テストの現状が立ち後れていることを氏は、指摘した。このような点から、今年の 7 月 28 日～31 日にかけて筑波国際会議場で開催された LTRC (The Language Testing Research Colloquium) '99 は 21 世紀に向けて世界の言語テストの動向を知る上で、又、言語評価という側面から我が国にとってはこの上もない機会であった。

現状の問題を解決していく上で大事な点を 3 つ挙げられた。一番目は、現実の課題を直視することである。すなわち現状テストの得点処理が適切に行われているかどうかを見極めることが大事である。Scale の違いによって被験者のテスト結果は変わってくる。結果的に被験者の能力を客観的にそのテストが計測しているかは疑問が残る。二番目にこのような問題の所在を明確にすることである。すなわち、データ処理の中にどのような問題が潜んでいるのかを明らかにすることであると指摘された。例えば、テスト困難度と受験者集団の問題、テスト弁別力と受験者集団の問題、テスト得点と受験者能力の問題、そしてテストの信頼性の問題である。3 番目に、これらをどのように解決していかなければならないかを挙げられた。すなわち異なるテストを受験者が受けても公平な測定ができるテスト方法を将来に向けて採用していかなければならない。具体的には今までの NUMBER RIGHT SCORE から欧米では常識になっている LOGIT SCORE を採用した項目応答理論を日本でも多く採用していかなければならない。更に日本は Rasch model、CAT (Computerized Adaptive Testing) に関しても研究を押し進めて行かなければ、今後我が国でも導入が決定されている TOEFL の CBT (Computer-Based Test) 等の欧米のテスト分野での立ち遅れが懸念される。

我々の日々行っているテストにもっと目を向け、直感ではなく科学的な評価を受験者に課していくことで本来の言語教育の発展に繋がるものであると示唆された。

JLTA 評議員の浪田克之介教授(北海道大学)の挨拶に続き、大友賢二 JLTA 会長による講演と中村優治 JLTA 副会長による Workshop が行われました。

Workshop 中村優治 (東京経済大学)

「スピーキング能力測定のあり方」

スピーキング能力とはいったいどのようなものなのかを氏は、定義づけることの重要性を説いた。「スピーキング能力」と一口に言っても簡単に定義づけることは難しい。例えば言語学的な側面からみれば words から pragmatic まで Microlevel から Macrolevel まであり、テスト方法の面でも代表的なもので interview test, group interview test, tape-mediated test, discussion test があり、場合によっては 60 を超える方法も存在していて、果たして同じ能力を測定しているのかという疑問も起こる。また、スピーキングが成立する対話の人数の面でも状況によって変わるので設定をどうするか、更に、設定された状況に対応するリスニング能力との関連等多くの要因がスピーキング能力を測定する上で立ちほだかっている。

氏は対話の設定を 6 つの範疇(例えば、One-Way Monologue, Face-to-Face Dialogue, etc.)に分け、そのテスト結果の分析を Rasch Based Analysis of Polychotomous Data により行い、被験者の能力と困難度を測定した。更に別の実験で 35 項目のテスト項目を被験者に課し、評価者の 4 点法をデータに加え同様に Rasch Based Analysis using FACET Model を用いて能力、困難度、評価の厳格さを同じスケールの中で測った。これにより評価者の評価の仕方が浮き彫りになり評価の基準を考える上で役に立つことが判明した。

最後に、今後この分野での発展に欠かすことができないのは、スピーキング・テストは生のスピーキングの状況をどれだけ推測できるのかを測れるものでなければならぬ。この点から Validity, Reliability, Practicality の 3 つが調和したものでなければならぬ。また、これを補う上でも被験者の能力を多面的にテストすることも重要である。この観点から、スピーキング能力の定義を厳密にしおくことが今後の課題と言えると指摘された。

(報告者: 竹村雅史 札幌北高等学校)

1. 研究紀要・会員名簿について

研究紀要第2号の発行後、研究紀要、及び会員名簿の今後の発行方法について役員で検討し、以下のよう
に決定いたしましたので、報告いたします。

- 1) 研究紀要は、次の事項を含み、必要な場合は、
会員以外の方にも配布・頒布する。

論文・研究紀要執筆要項・会則・ JLTA 役
員名・名誉会員名・年次活動報告（予算、
決算を除く）・編集後記・編集委員名

- 2) 会員名簿は、次の事項を含み、原則として、会
員のみ配布する。

JLTA 役員名・名誉会員名・全名簿（現行
どりの情報を掲載する）・会則・総会報告
事項（予算、決算を含む）

2. 新規加入会員

(1999年9月16日～2000年1月15日)

賛助会員 1 口

財団法人 国際ビジネスコミュニケーション協会

一般会員



現在、JLTA の一般会員は 104 名、賛助会員は 6 社
です。

3. LTRC 99 Program & Abstracts

LTRC 99 Program & Abstracts 冊子の余部が事務局に
若干あります。また、当日配布されたハンドアウトも
若干あります。ご希望の方は、事務局までお問い合わせ
下さい。

なお、LTRC 99 Program & Abstracts は、ILTA のホ
ームページ(<http://www.surrey.ac.uk/ELI/ilta/ilta.html>) か
らリンクする、Program Book のサイト(<http://www.surrey.ac.uk/ELI/ilta/ltrc99.html>) でも見ることが出来ま
す。ご活用下さい。

4. 『研究紀要』 第3号 原稿募集

『研究紀要』第3号の原稿を募集しています。ふる
ってご応募下さい。

提出期限は 2000 年 4 月 30 日(日)です。執筆要項の
詳細については、『研究紀要』第2号の 116-117 ペー
ジをご覧ください。

5. 住所・所属・銀行等の訂正・変更について

名簿の訂正、あるいは住所・所属などの変更が必要
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、できるだけ早くに、事務局まで必
ずご連絡下さい。

6. LTRC 2000

2000 年 3 月 6 日(月)～11 日(土) の間、カナダのバン
クーバーで LTRC 2000 が開催されます。

大会テーマは “Interdisciplinary Interfaces with
Language Testing” です。

詳しい情報は LTRC 2000 のホームページ
(<http://www.housing.ubc.ca/conferences/events/LTRC2000/homepage.htm>) で公開されています。

7. その他

JLTA の活動に対するご意見やご要望、ホームペー
ジの掲示板、Newsletter 等への掲載希望がありましたら、
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。

≡≡≡≡≡≡≡≡≡≡≡≡≡≡≡≡

