JLTA Newsletter
日本言語テスト学会
The Japan Language Testing Association

JLTA Newsletter No. １６　発行代表者：大友 賢二　2002 年 (平成 14 年) １１月３０日発行
発行所：日本言語テスト学会 (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

4. ＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

A Relay Essay (1)

Randy Thrasher
(Professor Emeritus, *International Christian University*)

An Apprenticeship in Language Testing

I became a language tester, as opposed to a language teacher who sometimes writes tests, practically by accident. When I entered graduate school after three years of teaching in Okinawa I needed a job, but initially, there were no teaching jobs available. However, Jack Upshur, who was then director of the University of Michigan ELI Testing Office offered me a job as an item writer for the Michigan Test of English Language Proficiency. It was essentially an apprenticeship. I would write items and Jack would tear them to shreds. But he didn't just tell me what was wrong with the items but also explained why what I had done was wrong. This learning by making mistakes helped me to begin to understand both the art of item writing and also fundamental measurement theory. Jack introduced me statistics and the practical constraints on large scale testing programs. He also gave me the job of designing the achievement tests the ELI needed and even got me involved in writing items for the first form of the TOEFL. But probably the most rewarding part of my apprenticeship were the long lazy afternoon discussions of both test theory and practical testing issues. My experience working with Jack enabled me to get more practical test writing and test design experience after returning to Japan. The YMCA gave me the opportunity to design and build an English proficiency test and with Y colleagues, designed and produced both a language aptitude test (published by Taishukan) and a test of Japanese language proficiency. When I returned to Ann Arbor to finish up my PhD, Jack gave me a desk in the testing office and my testing education continued. Even when I returned to Japan after getting my PhD, Jack and I were able to collaborate in designing the Businessman's English Test and Assessment (BETA).

Now there are many language testing texts and a huge number of articles in the field. They make great reading, but I am thankful that I learned both the art and science of testing by serving an apprenticeship under a master in the field.

第5回アジア英語テスティング会議報告

2002（平成14）年10月4日（金）から5日（土）まで、アルカディア市ヶ谷で第5回アジア英語テスティング会議が開催された。テーマは *Increasing Reliability and Validity of English Language Testing in Asia*。香港、中国、台湾、日本、韓国、シンガポールの各国から、延べ18名の発表があった。

宮田光朗氏、金口恭久氏の開会挨拶、小池生夫氏の基調講演に引き続き2日延べ12件の研究発表が行われた。Urmstonは香港の教員英語能力測定のためのテスト開発について、Kimura and Ougiは英検のスピーキング・テストについて母語話者と日本語話者の採点官の違いについてであった。Wuは中級受験者向けのスピーキング・テストの信頼性を相関係数、IRTなど多角的に検証したものであった。Yan and Lingliはスピーキング・テストのテスト不安について、retrospective account定性的方法を用いて分析したもの。

（報告者：　渡部良典　*秋田大学*）

午前のみの2日目は、4件の発表があった。中国のYan Huizhong教授とJin Yan博士は、National College English Syllabusに基づき作成されたCET testで40%の比率を占める読解部門に焦点を当て、2種のレベル(CET-4, CET-6)の問題を様々な角度から分析された。続いて、韓国のDongil Shin博士は、高校生を対象に、スピーキングおよびライティングの評価基準記述(descriptor)の妥当化を試み発表された。また、台湾のHengsyung Jeng教授は、5種のテスト(GEPT, CEECなど)について、Dale-Challのリーダビリティー公式(1995等)を用い、多角的に分析された。最後は、韓国のWon-Key Lee博士とOryang Kwon博士の発表で、多肢選択問題における選択肢の数について焦点が当てられた。限られたサンプル数ではあったが、平均値(mean score)、項目容易度(item facility)、弁別指標(discrimination index)の全てにおいて5択、3択、4択の順に高く、聴衆の多くが高い関心を寄せていた。

（報告者：卯城祐司　*筑波大学*）

Rejenthiran Sellen, Sued Mohamed and Low Ying Pingの発表は、シンガポールの小学校教育（4年間のFoundation Stageおよび2年間のOrientation Stage）の終了時に実施されるthe PSLE (Primary School Leaving Examination)に関するものであった。この試験は、各人の能力に合わせ、適切なレベルの教育の提供を目的とするシンガポールの教育システムの中で重要な役割を果たすものである。小学校4年生終了段階で、すでに言語教育の内容やレベル別にEM1・EM2・EM3の3コースに分け、複線型の教育を施しているが、よりアカデミックなコースであるEM1とEM2に属する学習者には、PSLEの中のthe English Language Examinationを実施し、EM3に属する約10％に対してはthe Foundation English Examinationが実施される。発表は、後者を中心に、前者と比較しながら4技能、文法、語彙などの測定領域や形式、配点等が紹介された。フロアーからは低学年の段階で能力別教育を行うことへの是非などの質問も出、言語政策および教育施策に対する各国の風土や状況の違いを感じさせられた。

（報告者：清水裕子　*立命館大学*）

第 6 回（2002 年度）全国研究大会が、2002 年 10 月 20 日（日）、東京経済大学で行われた。テーマは Language Testing and Second Language Acquisition Research。

＜基調講演＞
青木昭六（*愛知学院大学*）
「実践的コミュニケーション能力」の達成度設定と評価のために」

本講演の議題は 3 つあった。第 1 は、応用言語学におけるコミュニケーション能力やその構造(構成概念)についての先行研究に基づき、実践的コミュニケーション能力の評価を考察すること、第 2 は、達成目標とレベルの設定について提案すること、第 3 は、テストの課題が言語使用の側面に与える影響を考えることであった。

第 1 の議題について、コミュニケーションとは人とコンテクストとの相互作用であり、時間的制約の中で起こり、"continuous evaluation and negotiation of meaning"などが特徴である(Canale, 1983)と述べられた。また、実践的コミュニケーション能力の構成概念は、①実際の言語の使用場面(=実生活)で英語でコミュニケーションができる能力、②実際の英語の使用場面にふさわしい「言語の働き」を使える能力、③実際の言語の使用場面で遭遇する未知の状況で英語を創造的に使える柔軟な対応力から成ると仮定される(和田, 2002)ことも紹介された。具体的な評価の側面からは、国立教育研究所教育課程研究センター(2002)に基づき、実践的コミュニケーション能力を、正確さ・適切さ・流暢さ・複雑さの点から評価することが提案された。その他にも、規準から基準を具体化すること、評価者間の指導の大切さが示された。

第 2 の議題は 2 つの事例を用い、講演が進められた。1 つめの事例は、ACTFL Performance Guidelines for K-12 Learners であった。このガイドラインの特徴として、①小・中・高の外国語学習の目標であり、②成人ではなく、認知発達の途上にある小・中・高の学習者を対象とし、③教室での外国語学習を対象とすることが挙げられた。また、このガイドラインで興味深い点は、言語使用能力を 4 技能ではなく、(対人関係的・解釈的・発表的)×(発表力・理解力・言語構造の駆使力・語彙力・コミュニケーション方略・文化意識)という点から捉えていることであると指摘された。そしてこのガイドラインを参照すれ

ば、どの状況下で、何がどの程度できるかという発達の過程が概略的に理解でき、実践的コミュニケーション能力の達成度を具体化するときに参考になるということであった。2 つめの事例として、イギリスの National Curriculum Statements of Attainment for Modern Foreign Languages が示された。日本の中学レベルではレベル 3 もしくは 4 を、高校では 5 もしくは 6 を目標にするのがよいのではということであった。

第 3 の議題では、テストの設問がどの程度受験者に認知的不可を与えているかを明確にすることが必要だと強調された。例えば、視覚提示の有無や、時間的制約(提供する情報量・与えられる処理時間)などの諸条件は、受験者に影響を与えている(Skehan, 1998)とする研究が紹介された。

聴衆との質疑応答は 2 点あった。第 1 に、「ACTFL Performance Guidelines の文化意識の項目内の文化的適切さは、英語圏文化への同化を目指すのか、もしくは非母語話者間で共有される文化を目指すのか」に対して、「文化を定義するのは難しいので、回答するのは

難しいが、教授するときには (教授できると仮定するなら) 常に文化的事柄に注意を払うことが大切である」ということであった。第2に、「日本でもNational Curriculumを作り、到達目標を設定すべきか」に対して、「学校により地域差などがあり、文部省が作ることは難しいと思う。むしろ、日本英語検定協会などの方がデータも豊富だろうし、作ってもらいたい。また、到達目標がしっかりしていれば、学年の枠組みを無くしてもいいかもしれない」ということであった。最新の研究に基づき詳細な分析がなされており、非常に示唆に富んだ講演であった。

(報告者：印南洋　*筑波大学大学院*)

長沼君主（*東京外国語大学大学院*）

「複数の語彙レベルによるリーダビリティの測定の試み」

リーダビリティの公式には、単語をもとにしたものと、文をもとにしたものがあり、その複雑さの指標をもとにリーダビリティを算出している。このうちの単語をもとにして算出するリーダビリティの中には、FleschやGunning-Fogの音節数をもとにした公式とSpache や Dale-Chall の語彙数をもとにした公式がある。発表者は後者、Spache や Dale-Chall の公式で用いられている語彙リストは1つであり、対応するテキストのレベルが異なってくるという矛盾に着目し、複数のレベルの語彙リストをもとにしたリーダビリティの測定を試みた。

本発表ではアルクの SVL（標準語彙水準）の 12 段階を語彙レベルとしてとして用い、難易度の異なる9つのリーディングテストのテキストを分析した。それぞれのリーディングテストに項目応答理論を適応して等化を行い、項目困難度を求めた。そして、項目困難度とテキストの語彙レベルによるリーダビリティの関係を調査した。さらに同時に実施した語彙テストの結果からテキストの読解に必要な語彙力を算出し、テキストの語彙レベルとの関係も調べた。

その結果 1. Spache や Dale-Chall のリストではある一定の語彙レベルのテキストにしか効果がない可能性がある。2. 低レベルのテキストで高い語彙レベルの語彙があることが、高レベルで低い語彙レベルの語彙のないことが、テキストの読みやすさに有意な影響を及ぼすという結果となり、低レベルのテキストにおける信頼性が問題になることが判明した。

発表後、具体的な分析方法や等化リンクデザインについての質問がなされ、この研究が今後さらに進展することに対する期待が感じられた。

(報告者：藤田智子　*東海大学*)

神前陽子　(*武庫川女子大学*)

Peer, self-, and teacher-assessment in group presentation: A pilot study

本研究は、グループプレゼンテーションに対する相互, 自己, と 教師の 3 種類の評価がどのように異なるのかを調査する事を目的としている。そのために (1)学生はクラスメートのパフォーマンスに対して相対的に甘いか厳しいか? (2)学生は他のグループのプレゼンテーションを評価する時より、自分のグループの評価をする時のほうが甘いか? (3)教師の score range は、学生のそれより広いか? (4)教師評価の項目間の相対的な難易度の捉え方は、学生のそれと大きく異なるか? 以上の個別の疑問を、ラッシュモデルを使用して評価項目の困難度や 3 種類の評価者の厳しさ等、複数の項目を分析している。

被験者は、女子大の薬学専攻の 52 人で、3〜5 名の 12 グループの 5 分間プレゼンテーションを、相互、自己、教師評価している。そして、評価項目は 5 種類で、それぞれを 5 段階評価している。分析の結果、学生は、意外に公平に相互、自己評価していていること。また、教師評価は学生の相互評価より score range が広く、学生の相互評価とは異質である事がわかった。

パフォーマンス評価をラッシュモデルを利用して項目困難度や評価者の厳しさを分析する研究は近年特に盛んに行われ、注目をあびている。今回の発表はパイロットスタディーであるが、発表後、教師評価と学生相互評価を比較する時に教師を複数にすることや、この 2 つの評価を別々にして評価の厳し

さを算出すること、など建設的な提言がなされた。

(報告者：藤田智子　*東海大学*)

Alistair Van Moere (*Kanda University of International Studies*)
Validating Scores in University-Wide Group Oral Tests

The battery of proficiency tests administered to 1,600 students at Kanda University of International Studies are intended to mirror the curriculum. Students take the tests five times over their university career: first, to place them in a streamed language program, and once after each year to measure their progress. There are five separate tests of reading, listening, grammar, writing, and speaking. The presenter, Ali Van Moerne, devoted his session to examining the speaking test, which is a 10-minute instrument consisting of one conversation prompt administered to students in groups of four. The tests are videotaped and are later evaluated by two raters.

The chief questions that the presenter addressed were: Is the test fair?" and "What does the test mean?" To answer the first question, validity instruments in the form of a student questionnaire which self-evaluated their level of shyness and a qualitative evaluation from teachers as to whether each student was a high, medium, or low level performer. The second measure was compared to the actual test results and revealed discrepancies indicating that teachers' evaluations of students without the tests may not be accurate.

To answer the second question, Van Moerne did a correlation study of the Kanda test (KEPT) with the TOEFL. Using standard multiple regression, he found that there was a correlation of 0.775 between the two tests and that TOEFL scores could be predicted with 74% accuracy.

(報告者：Laura MacGregor　*学習院大学*)

AMMA Kazuo (*Tamagawa University*)
Development of multiple-choice grammaticality judgement test types

Amma Kazuo began by stating that his rationale for embarking on using multiple-choice grammaticality judgment tests was that simple "maru-batsu" tests were two easy and did not show whether students really understood English grammar. The presenter's endeavor here was to present a multiple-choice format which provided improved reliability and validity compared to a dichotomous format. He devised a 38-item error detection test which asked students to first decide if the item (sentence) was grammatically correct or incorrect. If correct, they should select one of four options which paraphrased the sentence in Japanese. If incorrect, they should select one of four options which identified the underlined part of the sentence that was grammatically incorrect. Within the group of eight options, there was only one possible answer. The test was administered twice: first, to over 1,000 students in multiple-choice format and second, to 156 students in dichotomous format. Using K-R20, he found the reliability index for the data of the first administration to be 0.934. He concluded by recommending a multiple-choice format for grammaticality judgment tests, since he found it to have a more stable reliability and better concurrent validity than a traditional dichotomous format.

(報告者：Laura MacGregor　*学習院大学*)

Chih-Hui Chang (*Da-yeh University, Taiwan*)
A Case Study on the Influence of Environemtn and Motivation on Second Language Learning in Taiwan

A total of thirteen students of Da-yeh university, three English majors and ten non-English majors, took part in the study. The participants were asked to keep language learning journal during the period of June through August 2001. The qualitative analysis of the data showed that the participants' attitudes towards the target language (English) and the language learning environment were strong indicators of success in their learning. Based on the result, a further investigation needs to be conducted as to the relationship between these factors and the scores of the General English

Proficiency Test (GEPT), which was developed by the Taiwan Government.

(reported by Yoshinori Watanabe, *Akita University*)

山崎朝子(*武蔵工業大学*)

語彙習得の測定におけるテスト

研究の目的：語彙習得に関して、偶発学習と意図学習の結果を比較すること。しかし、この発表では、その研究で用いられたテストについて報告することを目的とする。被験者：都立高等学校の 3 年生 62 人を偶発学習グループと意図学習グループに分けた。期間：4 月～9 月/4 月上旬に目標単語 150 語を選定/第一週に SLEP Test を実施/第二週に Pretest を実施/授業開始後 10 週目に Immediate Posttest を実施/Post test の 1.5 ヵ月後に Follow-up Test を実施した。目標単語 150 語は、先行研究に従い、graded readers に 8 回以上繰り返された語とした。パイロット・テストで、学生のできの悪い 90 語のうちの 30 語とまよわし 15 語で構成する Pre, post, follow-up test を作成した。3 つのテスト間に有意差はなかった。Follow-up テストの結果は、偶発と意図グループ間に有意な差は認められなかったが、速読を繰り返すことによって、語彙習得に効果があることが確認されたことに意義がある。

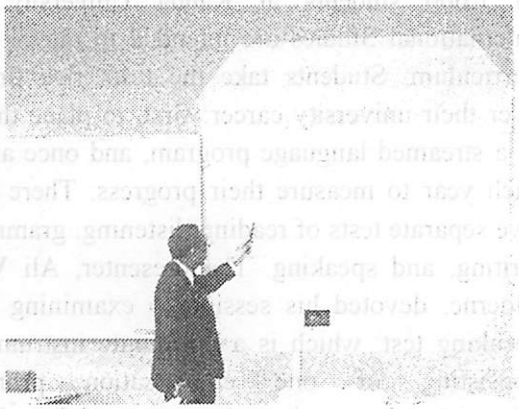
（報告者：大坪 一夫　*麗澤大学*）

池田央　(*教育測定研究所*)

TOEFL Summary Reports (1993-2001)から見えるもの

ETS の Web ページには，1993 年から現在までの TOEFL 要約情報が公開されている。この情報はデジタル化されているので，新たな観点から情報を加工しなおすことが可能である。印刷された情報は，報告者の結論を鵜呑みにするより仕方がないが，その点で，デジタル化された情報のほうが明らかに優れているといえる。Summary には，スキル別%ile Rank 換算表，平均，標準偏差。教育水準・男女受験者数。各国別受験者数，総点平均，スキル別平均などが記載されているが，その情報を利用して見てみると，合計点の平均値の順位では，PBT では，日本(以下 J)57 位，韓国(以下 K)47 位，中国(以下 C)31 位，PBT では，J:106 位，K:98 位，C:96 位であること，日本人は，聴解は苦手だが，読解・文法は得意だといわれているが，聴解のほうが読解より低い国には J,K,C が含まれ,聴解のほうを得意とする国には,フィリピン,マレーシア,インドネシア,などが含まれていること等がわかる。

（報告者：大坪 一夫　*麗澤大学*）



Shizuka Tetsuhito, *Kansai University*

Reliability/Validity of "Invisible-Blank Filling" Items

Invisible-blank-Filling test is a variation of a cloze test. The difference is that blanks are invisible. The first task is to locate the places where words are removed. Then these identified places need to be filled with the words listed at the bottom of the passage. Its statistical results were compared with those of a visible-blank-filling test of the same passage. The passage used for this visible/invisible test version was covered in class beforehand. The item discrimination mean for the invisible version is much higher than that for the other. These two tests were also compared with two c-tests. The passage of one of the c-tests was covered in class beforehand while that of the other was not. Both c-tests correlated higher with the invisible blank-filling test than the visible version. Of the two c-tests, the test whose passage was covered beforehand correlated higher with both blank-filling tests.

In addition to the statistical results described above, facility values and reliability information were provided.

At the presentation, the audience was asked to take the example test included in the handout. Finding the deleted parts was a challenging task for me. Once they were found, I tried to figure out first which part-of-speech and then what contextual information to be filled in to complete the task. Locating the missing parts itself might be tapping my L2 grammar, English.

（報告者：櫻井敏子　*神戸松蔭女子大学*）

Rie Koizumi , *Doctoral Course, University of Tsukuba*
Validating Speaking Test for Japanese Junior High School Students

Ms. Koizumi has developed a speaking test for Junior High School Students. The test consists of (a) a self-introduction task, (b) scripted role-plays, (c) a task in which the students are asked to talk about a favorite singer, TV program, or an animal and (d) a task asking students to tell the differences between two pictures. The content validity, construct validity and reliability of the speaking test were examined. This speaking test for Japanese junior high school students was developed to measure how much they have achieved in a classroom setting, to what extent they perform in outside of a school setting and how much they have developed their L2 grammar. Since a test which aims to measure these multiple purposes in a junior school setting is barely available, this research might be a good step for further study on this area.

（報告者：櫻井敏子　*神戸松蔭女子大学*）

Katagiri Kazuhiko　（*麗澤大学*）
Inferring Japanese Learners' English Ability by way of Measuring Their Vocabulary Size

本発表は、１）日本の中学３年生の英語能力において語彙知識がどのくらいの比重を持つか、２）大きな比重を持つとしたら、その"inner side of mechanism" を探る、という２つの問題意識を持った研究の報告で、たいへん興味深いものであった。発表者のKatagiri氏は、語彙サイズの測定にはMochizuki (1998)を用い、英語能力の測定にはSTEP英検３級の問題を用いている。まず語彙サイズテストと英語能力テストの間の相関を分析し、ついで英語能力テストを構成する７つのパートに関して因子分析を行っている。分析の結果として、Katagiri氏は次の２つの結論を見出している。１）日本の中学３年生の英語能力において語彙知識は大きな比重を占める。２）彼らの英語能力を支える要因の核になっているものは語彙知識と同様のものか、あるいは語彙知識を含む要因である。

Katagiri氏は、１）STEP英検３級の問題が英語能力を映し出す"mirror"になりえているか、２）本発表では、統計的なデータ分析だけであったが、主観的な評価も含まれてもよかったかもしれない、という２点を研究方法として議論の余地があるものとして挙げた。

きちんとした手続きとデータ分析の上になされた研究発表であったので説得力のあるものであった。

（報告者：塩川春彦　*北海学園大学*）

Hidetoshi Saito　（*北星学園大学*）
Rater Training Effects on On-line Paper Assessment of EFL Individual Presentations

本発表は、１）大学生のEFL授業において「ピア評価」を行う場合に、評価法についての訓練はどのような影響を及ぼすか、２）学生が「ピア評価」に否定的な感情を持っているとすればそれはなぜか、の２つの焦点をもった研究の報告であった。報告者のSaito氏は、英語のプレゼンテーションを学び実践することを狙いとした授業に出ている大学生74人を調査対象としている。この学生たちは２つのグループに分けられ、そのうち1つはプレゼンテーションの評価法について１時間の訓練を受けた。その上でクラスメートたちのプレゼンテーションを評価させた（13

項目）。結果として、Saito氏は、次の２点を報告した。１）学生間では評価法の訓練を受けた学生とそうでない学生の間に有意な差は認められなかった。しかし、教員の評価と学生の評価を比較した場合、訓練を受けてない学生グループの評価よりも、訓練を受けた学生の評価に強い相関が認められた。２）調査後のアンケートでは、ほとんどの学生は「ピア評価」に違和感はないようであったが、一部の学生には公平性、信頼性について強い疑念を抱いていたようだった。

教育的な示唆としては、１時間くらいの訓練ならしてもしなくてもよいのではないかということが挙げられた。しかしフロアーからは、点をつけるだけならその通りかもしれないが、評価法の訓練をすることには教育的な意味がある、と指摘する発言が上がった。建設的な質疑応答が活発に行われた発表となった。

（報告者：塩川春彦　*北海学園大学*）

中村　優治（*東京経済大学*）
飛渡　洋（*都立農業高校*）
「日本語と英語のライティング能力の実証的、統計的、比較研究」

この発表は JLTA 研究助成費による研究の中間報告である。この目的は日本人英語学習者の母語の書く能力がどの程度英語の書く能力と関係しているかを実証的、統計的に検証しようとするものである。具体的には、日本人英語学習者の発信能力において、書きたいことと実際にかけることのギャップが日本語と英語の作文の中にどの程度見られるのか、

また日本語と英語のライティングの表出能力の間に何か共通点と相違点が見られるか、などを分析した。被験者は大学生７６名で課題はつぎのようであった。日本文「公立学校で導入された学校週五日制について。賛成、反対その理由。」　英文 Discuss the new 5-day school week introduced in public schools. Do you agree or not?　Why?

日本語の作文は日本人３名、英文は英語母国語話者３名および日本人１名により評価した。評価の観点は全体の印象、論の構成、論理的つながり、語彙使用の適切さ、正しい文法使用、独創的内容の６項目で、１（very poor）から４（very good）までとした。

評価者間の信頼性は比較的高い数値を示した。結論としてまず英語力をつけさせなければならないグループ、英語は良いが日本語が下手なグループ、論理的なつながりの力をつけさせなければならないグループが存在することが明らかになった。

（報告者：清川英男　*和洋女子大学*）

斉田　智里（*茨城県立並木高等学校*）
「高校生のための英語学力テストの分析—学校英語テストの改善に向けて—」

この研究は茨城県高等学校英語部が県下の高校生に実施したテストの分析結果の報告で、斉田（２００２）の継続研究である。今回は1995年から2002年までの８年間のテスト A（１年生対象）、テスト B（２－３年生対象）（合計 16 種類、受験者数は 14000～33000 名）の項目分析と因子分析を行った。

項目分析の結果、テスト A とテスト B の正答率を比較すると中学校の復習事項を扱ったテスト A では学力レベルは幅広く分布していた。テスト B では中程度に集中しており、0.5 よりいずれも低かった。両テストの正答率の平均値の違いは、受験者要因としては、テスト B では英語の学力が低いことが考えられ、作問者要因としては、受験者の英語の学力の伸びを過大評価しているようであるが、IRT 分析を行わないと、断定できない。

信頼数係数（α）の値は、テスト A では 1997 年以降は 0.8 以上であったが、テスト B ではいずれの年度もテスト A より低かった。

識別力の低い項目はテスト A では発音、

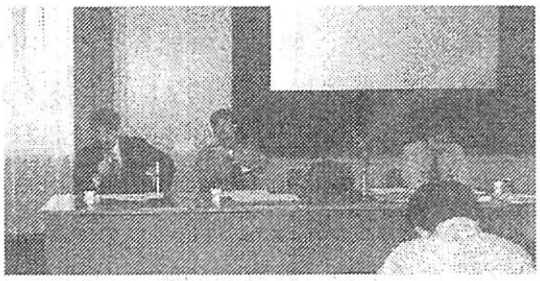
語彙、文法問題などに比較的多く見られ、テストBでは正答率が0.8を越える項目は8年間で6項目のみであった。

（報告者：清川英男　*和洋女子大学*）

＜パネルディスカッション＞

言語テスト研究と第 2 言語習得研究の相互作用

伊藤　彰浩（*愛知学院大学*）
白畑　知彦（*静岡大学*）
斎藤　英敏（*北星学園大学*）



言語テスト研究（以下LT）と第2言語習得研究（以下SLA）の相互作用について、本学会会員外の白畑知彦先生をSLAの代表としてお迎えし、LTを代表してSLAにも造詣の深い会員の伊藤彰浩氏と斎藤英敏氏の3人でパネルディスカッションが開催された。

最初に、司会兼発表者の伊藤氏が、「両者の相互作用については90年代よりずっと主張されてきていることだが、2000年代になっても議論が深まっていない印象を受ける。よって本日ここで改めて議論を行ない意味あるものにしたい。」と今回のパネルディスカッションの意義を確認し、引き続き伊藤氏が発表を行なった。今まで両研究領域を包括的にやってきた経験があること。その研究が、当初の期待ほどにはあまり評価されなかったこと。そして両研究の相互作用は難しいという印象をもったこと。等が述べられた。

次に、斎藤氏の発表が行われた。1）LTとSLAは合致しなくてもよいのではないか。2）両者はそもそも異なった測定方法（measures）で異なったものを測定対象（constructs）にしているのではないか。3）統計的結果と実際（clinical）の結果はどちらを優先するのか。4）Bachman (2002)が言うように、タスクの難易度は受験者とタスクの絡みで決まるのであってア・プリオリに決まるものではないのかもしれない。という4点が述べられた。

3人目に白畑氏の発表が行われた。LTとSLAの両方に興味があったが、とりあえず今まではSLAを中心に研究をやってきたとのことで、これまでにデータを収集する際に用いてきた方法(SLA研究で使ったテストというより、誘出法）を、それぞれの長所と短所の説明とともに紹介がなされた。

その後、フロアーの参加者も交えたディスカッションへと移った。「SLAにおける誘出法において、妥当性・信頼性はどう考えるものか。」といった質問を皮切りに、「LTとSLAそれぞれの専門家の交流が大切だ。共同研究もよい。また、これからは、両方とも学ぶのがもっとよいかもしれない。」「被験者の量をたくさんとってきて調べるとわかることももちろんたくさんあるが、しかしSLAでは、言語習得過程を知りたいのが第一の目的だから、やはり被験者が少なくても縦断的（longitudinal）研究のほうが、知りたいと思うことがよくわかる場合が多い。」「LTとSLAで共同研究をやっているのだが、やはり時々違いを感じる。LTでは『到達』と言うがSLAでは『習得』と言い術語が違ったりする。LTでは母語話者なら正解できるテスト項目でないと駄目だが、SLAでは規範文法を解答とするので母語話者ですら正解できない問題であってもよく、基本的考え方が違ったりして一口に共同研究といっても難しい。」等々、この紙面では書ききれないほどのたくさんの有益なコメントが出され、時間が足りないと感じさせるほどの白熱したディスカッションとなった。

（報告者：片桐一彦　*麗澤大学*）

＜総会＞

総会では、まず、議長選出がおこなわれ、

結果、浪田克之介理事が推薦を受け選出された。その後、中村洋一事務局長より、まず2001年度の活動報告と2001年度の決算報告がなされた。引き続き清水裕子監査委員より監査報告がなされた。審議の結果、両報告とも全会一致で承認された。

次に、中村事務局長より、2002年度の活動計画案と予算計画案が提案された。審議の結果、その両案が全会一致で了承され、今年度もますます活発に活動を展開していくことが確認された。

（報告者：片桐一彦　*麗澤大学*）

＜閉会＞

Randy Thrasher 副会長より、まず、今大会の参加者、発表者、講演者、パネリスト、実行委員、開催大学担当者ら多くの人にお礼の言葉が述べられた。そして今後とも、より一層学会活動が活発になり、研究の質と量がますます充実し、社会に対してよい影響を与えつづけ、そしてその結果、国内のテストの質が上がることを期待すると述べられた。これをもって、第6回目の全国研究大会も幕を閉じた。

（報告者：片桐一彦　*麗澤大学*）

＜懇親会＞

中村洋一事務局長懇親会にて曰く「最も力をいれて企画している催し（？）。」とのことで、食べきれないたくさんの豪華な御馳走といろいろな飲み物が用意されたすばらしい懇親会であった。報告者の個人的感想だが、この学会の全国研究大会での懇親会ほど、お手頃な懇親会費で、しかもたくさんの御馳走がでるところはないと思う。他の学会では、すぐに食べ物がなくなるということに何度も遭遇している。

恒例となった、懇親会参加者全員が円を作って一人一人挨拶と簡単なコメントを述べるという「行事」が行なわれた。「最近、ラジオ体操に週4日行っているのだが、ちっともおなかが引っ込まない。・・・」といったある先生のお話（飲み会の席での話は、匿名にしておきます。）をはじめ、昼間の堅い学問の話から離れて、和やかな雰囲気で楽しい話で盛り上がった。

（報告者：片桐一彦　*麗澤大学*）

****

国際シンポジウム
「問題作成から見る大学入試」報告

平成14年11月16日（土）、独立行政法人大学入試センター主催で国際シンポジウム「問題作成から見る大学入試」が開催された。

Yang Huizhong & Jin Yan (*The National CET Committee, CHINA*)

The Testing of Reading Comprehension – Item Setting and Test Validation for Large-Scale Public Test

中国で実施されている大学生向け英語力測定試験CETについての説明とCETテスト中特にリーディング・セクションを取り上げその妥当性の検証について発表がなされた。

CETは、1987年より始まったテストで、年2回おこなわれており、毎年多くの学生がこのテストを受験する。国家による大学英語教育シラバス（the National College English Teaching Syllabus）により、とくにリーディング能力の育成に力を注ぐことが決められており、CETもそれに準拠して作成されている。よってテストの構成もリーディング・セクションが最大の40%を占め、他にリスニ

ングが 20%、語彙が 15%、ライティングが 15%、統合問題が 10%の構成となっている。といった、CET についての説明が最初におこなわれた。

次に、リーディング・セクションの作成過程についての説明がおこなわれた。どのようにパッセージを選らんでくるのかといった話から始まり、またその際 Flesh のリーダビリティ公式を使ってパッセージの困難度を制御していること、そしてテスト項目作成時のスペックの詳細についての説明等がおこなわれた。

最後に、妥当性の検証が、波及効果の面から、計量心理学的面から、またアンケートの面からなど多方面から検証がなされた。どの観点からも、良好の結果が得られているとの報告であった。

（報告者：片桐一彦　*麗澤大学*）

W. Edward Curley (*Educational Testing Service,* USA)
The Development of College Board SAT-Verbal Questions

アメリカの SAT（言語テスト）の作成過程についての発表がなされた。

毎年 1,000 以上ものテスト項目が作成され、そして良問かどうかのチェックがなされている。ETS スタッフ 8-10 名と外部委託者（英語教師や物書き関係者や出版物編集関係者等の職業の人々）がテスト項目を作成する。もちろん質の均一のために、テスト項目の内容に関しての詳細なスペックが決まっている。その後、適切なテスト項目かどうかの検査が、15 とか 16 といった段階にもわたっておこなわれる。ここでは、何段階もの言語的・内容的審査を受け、その後、事前予備テストが実施される。事前予備テストを実施することにより、統計的観点からも項目分析がおこなわれる。各種統計数値（各項目ごとの困難度、項目困難度の平均、項目困難度の標準偏差、弁別指数の平均、DIF 値の平均）には、それぞれに対して詳細な数値許容範囲が定められており、それを満たさなければ合格とはならない。

事前予備テストの施行方法についてだが、実は本番で使われている SAT の言語テストにはその事前テストしたい「試行テスト項目」が散りばめられており、3 時間のうち 30 分の分量に相当する項目数が「試行テスト項目」である。もちろん、受験者はどれが「試行テスト項目」かは知らないし、受験者の得点算出時にはその「試行テスト項目」は使用されない。

その他、最新の IT 技術を駆使した、問題作成者に対する訓練方法等についての説明もなされた。

（報告者：片桐一彦　*麗澤大学*）

Anne Barnes (*Assessment and Qualifications Alliance,* UK. Chief Examiner)
Constructing an examination which provides valid and reliable assessment of English as a foreign language

英語を母語としない者が英国の大学に入学する際に課せられる英語テストのうちリー ディング・テストに焦点を当て、実際の英文や具体的なテスト項目を用いて、その妥当性と信頼性の観点から、読解力の測定の問題点、項目作成の困難点に触れた発表で あった。用いる英文のトピックの選択と設問の

設定からくる問題点について詳細に報 告された。また、テスト形式やテスト方法の重要性に関しては、クロー ズ・テストや 多肢選択形式の妥当性の低さ、客観的テスト形式が潜在的にもつ問題点に触れられ、 リーディング・テストの項目作成の難しさを再認識させる発表であった。

（報告者：清水裕子　*立命館大学*）

Kyungae Jin (*Korea Institute of Curriculum and Evaluation*, KOREA)
Issues in Developing College Scholastic Ability Test Items

韓国において、大学教育で必要とされる能力を測定するために高校３年生や卒業生が受験する試験である College Scholastic Ability Test (CSAT)の中の英語テストに関 して、その開発概念、構成、過去の得点推移、項目の作成とアイテム・バンクの構築 などが紹介された。また、2003 年に向けての新しい方向性にも触れられたが、CSAT が 日本の大学入試センター試験に近いものであるだけに、下位テストの構成やリスニン グ・テストの実施、テスト結果の分析等、興味深く聞くと同時に、学ぶべきことが多い発表であった。

（報告者：清水裕子　*立命館大学*）

三宅雅明（*武庫川女子大学*）
継続的に良問を提供するための条件

大学入試センターで英語作問部会の会長をされていた経験から、センター入試問題作成の苦労話を紹介された。テスト作成にあたって「公正さ」が重要であることは言うまでもないが、50 問のテストで 50 万人の英語能力を測定しなければいなければならない状況の中でも、受験者の気持ちに配慮する「優しさ」、テストを通して若者を育てる「優しさ」が同様に重要であるという、発表者の信念に、シンポジウム参加者の共感があった。発表者自身の「公正さ」と「優しさ」が感じられた発表であった。

（報告者：木村真治　*関西学院大学*）

＜指定討論＞
清水公男（*東京都立八王子東高等学校*）
池田央（*立教大学名誉教授*）

清水公男先生は、高等学校の現場から見た大学入試問題についての種々の疑問・問題点（例　問題の使い捨て）を指摘された。池田央先生は、清水先生の疑問・問題点を受けて、問題作成準備・作成・実施・分析のプロセスを見直すことにより、テストをより科学的に捉えることの必要性、特に、共通尺度を確立することの重要性を述べられた。

妥当性はテストの重要な概念であり、Item Specification などを通して、妥当性の高いテストを作成する努力は、様々な国でなされている。しかし、それは問題作成者が考えた妥当性であって、実際に問題が何をテストしているのかを知るためには、受験者からのフィードバックも重要である。Curley 先生は SAT-Verbal を外国語に翻訳しない理由を説明されたが、日本語版と英語版のアナロジー・テストを受験し、結果が異なるとすれば、それが翻訳による微妙な問題の意味変化が原因なのか、第一、第二言語でのアナロジー・プロセスの異なりが原因なのかは明らかでない。ETS であれば、この程度の私の疑問に

は既に答えているのであろうが、要するに全ての受験者について、問題作成者が意図した能力測定になっているか否かは、Item Specification を見ただけでは分からない。Barnes 先生が問題視される多肢選択形式の問題は、正しい選択肢を見つける以上の「深い読み」を阻害するだけでなく、読解問題解答過程において実際に受験者がどのような行為が行ったのかについての情報を、まったく提供できないという欠点も持つ。この意味で Yang 先生が報告された、Think Aloud 法によるストラテジー調査は大変興味深いものである。理想的には、テスト全体ではなく、問題項目ごとや大問ごとのストラテジー分析であれば、例えば、文法セクションでは「当て推量」が多いなどの報告も可能であろう。もし、このようなことが判明するのであれば、清水先生が述べられたように、高校生が疑問に感じるような出題を少なくすることができるかもしれない。三宅先生のおっしゃる「優しさ」を持ったテストにもなるだろう。受験生が様々なことを考えて解答しても、正解・不正解の１か０かのデータになってしまうと、統計は一人歩きし出すので注意をしなければならない。池田先生がおっしゃられたように共通尺度の作成は、これからの日本のテストにおいて非常に重要なことである。しかし、それと同時に進めておくべきこともある。池田先生が、問題作成者や分析者、テスト関係者のコミュニケーションの重要性について触れられたこと、大学入試センターの岩坪先生が「数字だけで全てが分かるわけではない」と言われたのを思い出しながら、帰阪の途についた。

（報告者：木村真治　*関西学院大学*）

@@

Information

CALL FOR PAPERS

TEMPLE UNIVERSITY JAPAN 20th ANNIVERSARY - TEMPLE UNIVERSITY APPLIED LINGUISTICS COLLOQUIUM 2003

Sunday, February 16, 2003 TUJ-TOKYO

The organizing committee of the Temple University Applied Linguistics Colloquium invites interested applied linguistic researchers to submit proposals for presentation at the Temple University Applied Linguistics Colloquium 2003 which will be held on Sunday, February 16, 2003 at Temple University Japan in Tokyo as part of the 20th anniversary of Temple University Japan. This colloquium is co-sponsored by the JALT Pragmatics SIG. Two areas of special interest are pragmatics and vocabulary research this year.

Please note that research in any area of Applied Linguistics will be welcome. Two types of proposals will be considered:

1. Reports on completed research
2. Works in progress (including completed

research design and/or data collection)

Presentations will be 30 minutes in length (including 10 minutes for questions and answers).

Please send a 50-word summary for the colloquium program, a 150-word abstract、 and personal information to the following e-mail address with an attached file (Word or Rich Text File) by December 10th.

tuj-linguistics-conf@tuj.ac.jp

For further information, please contact Megumi Kawate-Mierzejewska at mierze@tuj.ac.jp

The attached file should consist of 2 pages:

(1) one cover page with
   (a) name of all authors
   (b) full address
   (c) phone and fax
   (d) e-mail address of all authors
   (e) affiliation
   (f) paper title
   (g) proposed type of presentation (i.e., completed research, work in progress)
   (h) a 50-word summary

(2) an anonymous one-page 150-word abstract with paper title

Please write your name (family, first) in the space for subject (RE) when you e-mail the file.

Notification of acceptance or non-acceptance of your proposal will be sent to you by mid January (around January 10, 2003).

We look forward to receiving your proposals.

Megumi Kawate-Mierzejewska、 Ed.D.
Committee Chair
TUJ Applied Linguistics Colloquium 2003

◆ 転勤、転居等、JLTA の名簿記載事項に変更が必要な場合は、速やかに、事務局までご連絡下さい。また、銀行引き落としによる会費納入を利用している会員で、吸収・合併などにより、銀行名、支店名、口座番号等が変更になった場合は、必ず事務局まで、その旨をお知らせ下さい。銀行引き落としが不可になった場合、その年度のみ、郵便振込による納入をお願いしています。ご理解の上、ご協力をお願いいたします。

◆ 住所、所属、メールアドレスなどの変更があった会員は事務局までご連絡ください。

◆ 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ご連絡ください。

Post script from Editor

"In examinations the foolish ask questions that the wise cannot answer" (Oscar Wilde). I hope this isn't true. (Wt).

日本言語テスト学会事務局
〒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