
JLTA Newsletter

日本語テスト学会

The Japan Language Testing Association

JLTA Newsletter No. 18

発行代表者: 大友 賢二

2003 年 (平成 15 年) 6 月 9 日発行

発行所: 日本語テスト学会 (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>



「改めて考えてみると」

中村優治 (東京経済大学)

Language Testing や英語科指導法の授業を担当したり、ワークショップで testing の話をする、受講者からの質問や意見に改めて考えさせられることがある。その中からここでは 3 つ取り上げて、「改めて考えてみる」ことにしたい。

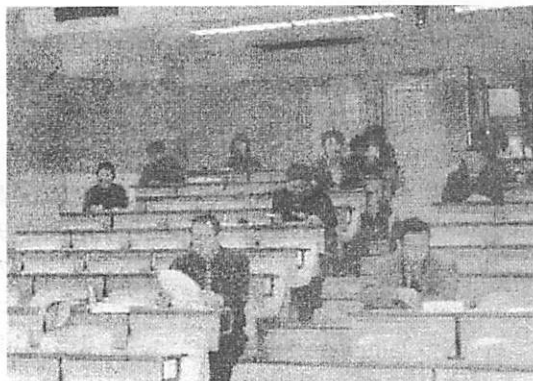
まず、a good test taker (ここではテスト高得点者を意味する) は a good language learner ではないのかということである。手元にある TESOL や SLA の本をみても、a good language learner の定義に a good test taker は含まれていない。私たちはテスト結果に基づいて、高得点者順に合格者を決めたり成績優秀者を発表したりするにもかかわらず、a good test taker を a good language learner の定義にはっきりとは含めていないようである。テストでいい成績をとる人はその言語の勉強のコツを会得した、いい学習者ではないのだろうか。逆に、a good language learner でテスト得点が低い場合、それは何を意味するのだろうか。改めて考えてしまうのである。

次に、a good test item writer (ここではテスト理論上良問とされるものを作成できる a good test maker を意味する) は a testing specialist ではないのかということである。しばらく前になるが、ある機関の、英語を含めた様々なテスト問題の作成検討にかかわる機会があって、改めて実感したことであるが、テスト理論をまったく知らないか、または意識していない人の中にすばらしいテスト問題 (ここでは受験生のレベル、興味などを含めてテスト理論上良問とされるもの) を作成する人が少なからずいるということである。この人たちは a testing specialist ではないのだろうか。一般に、テストの専門家というと、テスト理論を学び、問題分析を行い、統計処理をする能力がある人 (必ずしもこのような人が良問を作成するとは限らないが) などといったイメージが出来上がってしまっているような感じがしないでもない。それぞれの勤務校において、テスト問題作成時にテスト理論とは程遠い分野の同僚の数多くの良問に本来の a testing specialist と称される人が幾度となく助けられているという現実があるという話を聞いた時に、改めて考えてしまうのである。

最後に、a fuzzy test (ここでは学生の能力を大枠で捉えて、大まかな測定をするテストを意味する) は a good test ではないのかということである。「いいテストとは受験者の能力を正確に測定するもの」という判断基準に立てば、この両者は相容れない。しかしながら、テストが fuzzy であるが故に、その結果に精神的に救われ、勉学への意欲を失わずに学生生活・就職活動を送ることのできる学生が少なからずいることもまた事実であると聞く。テストは動機付けになるという観点から考えると a fuzzy test は a good test といえないだろうか。昨今、社会の様々な場面でテストスコアの提示が求められる機会が増え、より正確な数字 (正確な能力測定結果であるかどうかは別) で人を判断する傾向にある。こういうときにこそテストの fuzzy な部分を改めて考えてみたいのである。

第4回 JLTA 言語テストワークショップ 報 告

第4回 JLTA 言語テストワークショップが、平成15年3月8日(土)、沖縄県宜野湾市にある沖縄国際大学で開催された。言語教育関係者、特に中学校・高等学校・大学の英語科教員を対象とした、言語テストの理論と実践に関する基礎的なワークショップであった。数日後に沖縄県の公立高校の入学試験を控えていたが、東京方面からお見えになった先生方も含め約30名の出席者を得、また本学会に新たに入会された先生方も数多くいらっしゃったことは大変喜ばしいことであった。以下、その内容をごく簡単に報告したいと思う。



午前中は、大城賢先生(沖縄国際大学)の司会のもと、本学会の副会長である Randy Thrasher 先生(沖縄キリスト教短期大学)から開会の辞が述べられ、引き続き3人の先生方からご講演を賜った。

1. 「言語テストの基礎」:

大友賢二先生(常磐大学)

1) 言語能力の捕らえ方、2) 言語テストの作成、3) テスト結果の分析、4) 言語テストの新しい動き、について分かりやすくその概要をご説明いただいた。

2. 「テストのデザインと項目作成」:

中村優治先生(東京経済大学)

Test specifications に関して、次の11項目に分けて詳しくご説明いただいた。1) テストの目的、2) 被験者の特性、3) テストのセクション数と試験時間、4) テキストタイプ、5) 測定する言語能力、6) 測定する言語項目、7) タスクの内容、8) 各セクション内のテスト項目数と各テスト

項目の重要度、9) テスト形式、10) 指示、11) 評価基準。

3. 「同上」:

Randy Thrasher 先生(沖縄キリスト教短期大学)高等学校の英語教科書にある Audrey Hepburn に関する短い英文を例に挙げ、内容理解の問題作成上留意すべき点について、参加者全員にその具体案を作成させながら実演された。その中で、読解問題における multiple choice test の長所と限界についてもお話いただいた。また、Carl Lewis に関する英文を用いて、パッセージの正しい理解にはその背景知識が重要である点、あらためて強調された。



午後は、村田典枝先生(沖縄キリスト教短期大学)の司会で、中村洋一先生(常磐大学)から「言語テストデータ分析」と題してご講演を頂戴した。特に、測定と評価を分化し、適切なテストデータ分析を行うことが重要で、それを通して初めて「テストで言語能力は測れるか」という問いに答えを出すことができるという点を中心にご説明いただいた。その中で、「古典的テスト理論」の欠点と限界、それと相互補完的な関係にある「項目応答理論」の特性について、初めての方にも理解できるように、様々な分かりやすい比喩を用いて解説された。

講演終了後、木下正義先生(福岡国際大学)から本学会の使命の重さと今後益々の発展を願って閉会の辞が述べられた。引き続き、場所を移動して懇親会を行ったが、お酒の席でも言語テスト研究の今昔とさらには将来についても活発な議論が交わされた。特に、中村洋一先生の雄弁さはお酒が入るとまさに圧巻であった。最後になるが、本ワークショップの準備から運営に至るまでいろいろとご苦心いただいた

Thrasher 先生、大城先生、そして村田先生に心からお礼を申し述べる次第である。

(報告者 大津敦史 福岡大学)

AAAL Annual Conference 2003 報告

2003年3月22日から25日にわたり AAAL (全米応用言語学学会) が、アメリカのバージニア州アーリントンのペンタゴン (国防総省) に程近いシェラトンナショナルホテルで開催された。イラク戦争が開始された直後であったため、発表のキャンセルやボイコットもあったが、総勢 800 名以上が全米のみならず、世界各国から参加した。

今回の AAAL は Assessment and evaluation 以外にも (1) Analysis of discourse and interaction, (2) Applied linguistics research methodology, (3) Bilingual, immersion, heritage, and language minority education, (4) Language acquisition and language attrition, (5) Language and ideology, (6) Language and learner characteristics, (7) Language and technology, (8) Language, cognition, and the brain, (9) Language, culture, and socialization, (10) Reading, writing, and literacy, (11) Second and foreign language pedagogy, (12) Sociolinguistics, (13) Text analysis, and (14) Translation and interpretation の研究発表を含み、まさしく “The Diversity of Applied Linguistics” という今学会のテーマを体現していた。

さて、この AAAL で最も印象に残った Assessment and evaluation 分野の発表は、Bachman が率いるコロキウム “Validity issues in a web-based language assessment system” であった。このコロキウムでは、UCLA で従来使用している placement test (筆記試験) をコンピューター利用のテスト (CBT) に変更した場合に生じる妥当性の差異の問題を quantitative approach (statistics analysis) と qualitative approach (interview, think-aloud) の両方を使用し分析した結果が報告された。一例を挙げると、Writing では CBT の採点結果が筆記試験より有意に厳しいことが明らかになったが、逆に、CBT で

はスペルチェック機能などのため、CBT による採点は construct validity に影響を与えることが示唆された。また、受験生の英語レベルによって筆記試験と CBT に対する受け止め方が異なることも報告された。

これらの発表は Bachman がチェアを務める中、UCLA の学生によって行われた。おそらくは彼らの thesis または dissertation にかかわる研究であろうと推察されるが、Bachman の指導の下で、quantitative approach と quantitative approach がリサーチの方法論として同等に分析に用いられていたのには、驚きであった。それと同時に、教育者として自分の研究の専門分野にかかわらず学生に対し、統計的アプローチだけではなく幅広い方法論を教示していくことの大切さを改めて強く認識させられた一日となった。

(報告者 吉田弘子 大阪女学院短期大学)

第 13 回 AILA (国際応用言語学会) 報告(補足)

前号で第 13 回 AILA の報告を行いました、さらに酒井志延氏の Problems with Japanese High School Mid-term and Term-end Examinations - A nationwide study を追加いたします。

The JLTA Code of Good Testing Practice (言語テストのスタンダード案)

わが国では毎年実に多くのテストが作成され、実施されている。また TOEFL や TOEIC をはじめとしたいわゆる外部テストを数多の受験者が受け、その結果が何らかの目的に使われている。しかしながら、テストははたして正確に受験者の言語能力を測定しているであろうか、目的にあった使い方がされているであろうか。テストの質を確保するためには何らかの共通の指標が必要である。教育心理の分野では夙に APA (American Psychological Association) のスタンダードが整備されているが、言語テストの分野でも国際言語テスト学会 (ILTA) が中心となって作成されつつある。以下は Thrasher 教授を中心に進められている、日本版「言語テストのスタンダード」の素案である。わが国にはわが

国の特有の問題がある。広く会員のご意見を賜る次第である。thrasher@ocjc.ac.jpまでお寄せいただきたい。(編集委員 渡部良典)

The draft Code of Practice below is a work in progress. It is one attempt to move toward the realization of the JLTA goal of improving language testing in Japan. I have taken the lead in making this draft but, if it is to become what this association believes is good test practice, the code needs the input of as many JLTA members as possible. Each of us should examine the draft code to see if it fits our understanding and experience. Such a code must be as complete as possible but it should also be workable. And the provisions of the code must be clear and understandable. Please study this draft carefully and send any comments or suggested additions or revisions to me at thrasher@ocjc.ac.jp or to the JLTA Secretariat. I will also try to deal with any questions concerning the code that you might have. In the next JLTA Newsletter I will explain in more detail why I believe we need such a code and how it might be applied.

(Randy Thrasher)

Basic Considerations for good testing practice in all situations

1. The test developer's understanding of just what the test, and each sub-part of it, is supposed to measure (its construct) must be clearly stated.
2. All tests, regardless of their purpose or use, must be valid and reliable to the degree necessary to allow the decisions based on their results to be fair to the test takers.

Validity refers to the accuracy of the inferences that are drawn from the test results. If, for example, the test purports to be measuring the ability to use English in business communication, the test is valid to the degree it does in fact measure that ability. However, the ability to use English in business communication is a construct. The test developer must spell out just what that construct is or what it consists of. The test can only be valid if the test construct is a complete and accurate picture of the skill or ability it is supposed to measure. To summarize Messick 1996, construct validity depends on the degree to which positive answers can be given to the following five questions.

1. Does our construct and our implementation of it, include all and only the necessary elements?
2. Do we have these elements correctly weighted?
3. Do these elements interact in the same way in the test task and in real world performance?
4. Does our scoring scheme evaluate the test performance in the same way real-world performance is evaluated?
5. Is there anything about our test that will cause the test takers, or a portion of them, to perform in a less than optimal fashion?

Reliability refers to the consistency of the test results and Messick's sixth question addresses this issue.

1. Are the results generalizable?
 - a. Are the results comparable across time?
 - b. Are the results comparable across settings?

Responsibilities of test designers and test writers

1. A test designer must begin by deciding on the construct to be measured before deciding how that construct is to be operationalized.
2. Once the test tasks have been decided, their specifications should be spelled out in detail.
3. The work of the item writers needs to be edited before the items are pretested. If pretesting is not possible, the items should be analysed after the test has been administered but before the results are reported. Malfunctioning or misfitting items should not be included in the calculation of individual test takers' reported scores.
4. Grading check sheets or rubrics must be prepared for test tasks requiring hand scoring. These check sheets or rubrics must be tried out to demonstrate that they permit reliable evaluation of the test takers' performance.
5. Those doing the grading should be trained for the task and both inter and intra-rater reliability should be calculated and published.
6. Test materials should be kept in a safe place and handled in such a way that no test taker is allowed to gain an unfair advantage over the other test takers.
7. Care must be taken to assure that all test takers are treated in the same way in the administration of the test.
8. Grading procedures must be carefully followed and score processing routines checked to make certain that no mistakes have been made.

9. The test results should be reported in a way that allows the test taker and other stakeholders to draw the correct inferences from it.

Obligations of institutions preparing or administering high stakes exams

Institutions (schools, companies, certification bodies, etc.) developing and administering entrance, certification, or other high stakes examinations must utilize test designers and item writers who are well versed in current language testing theory and practice and have native or near native competence in the language being tested. Items written by non-native speakers of the language being tested must be checked by competent native speakers.

Responsibilities to test takers and related stakeholders:

Before the test is administered.

The institution should provide all potential test takers with adequate information about the nature of the test, the construct (or constructs) the test is attempting to measure (Ideally this should include any evidence and arguments showing that the test tasks are in fact measuring what they are claimed to measure.), the way the test will be graded, and how the results will be reported.

At the time of administration.

The institution shall provide facilities for the administration of the test that do not disadvantage any test taker. Test administration materials should be carefully prepared and proctors trained and supervised so that each administration of the test can be uniform, assuring that all test takers receive the same instructions, time to do the test, and access to any permitted aids. If something occurs that calls into question the uniformity of the administration of the test, the problem should be identified and any remedial action to be taken to offset the negative impact on the effected test takers should be promptly announced.

At the time of scoring

The institution shall take the steps necessary to see that each test taker's exam paper is graded accurately and the result correctly placed in the data-base used in the assessment. There should be on-going quality control checks to assure that the scoring process is working as intended.

Other considerations

If a decision must be made on candidates who did not all take the same test or the same form of a test, care must be taken to assure that the different measures used are in fact comparable. Equivalence must be demonstrated statistically.

If more than one form of the test is used, inter-form reliability estimates should be published as soon as they are available.

Obligations of those preparing and administering commercially available exams

In addition to the obligations placed on any test designer and on those preparing high stakes examinations, developers and sellers of commercially available examinations must:

1. Make a clear statement as to what groups the test is appropriate for and for which groups it is not appropriate.
2. Make a clear statement of the construct the test is designed to measure in terms a layperson can understand.
3. Publish validity and reliability estimates for the test along with sufficient explanation to allow potential users to decide if the test is suitable in their situation.
4. Report the results in a form that will allow the test users to draw correct inferences from them and make them difficult to misinterpret.
5. Refrain from making any false or misleading claims about the test.
6. Produce a test manual available to the public which:
 - a. Explains the relevant measurement concepts so that they can be understood by non-specialists.
 - b. Reports evidence of the reliability and validity of the test
 - c. Describes the scoring procedure and, if multiple forms exist, steps taken to assure consistency of results across forms.
 - d. Explains the proper interpretation of test results and any limitations on their accuracy.

Responsibilities of users of test results

Persons who utilize test results for decision making must:

1. Use results from a test that is sufficiently reliable and valid to allow fair decisions to be made.
2. Make certain that the test construct is relevant to the decision to be made.

3. Clearly understand the limitations of the test results on which they will base their decision.
4. Take into consideration the standard error of measurement (SEM) of the device that provides the data for their decision.
5. Be prepared to explain and provide evidence of the fairness and accuracy of their decision making process.

Special considerations

In norm-referenced testing

1. The characteristics of the population on which the test was normed must be reported so that test users can determine if this group is appropriate as a standard to which their test takers can be compared

In criterion referenced testing

1. The appropriateness of the criterion must be confirmed by experts in the area being tested.
2. Since correlation is not a suitable way of determining the reliability and validity of criterion referenced tests, methods appropriate for such test data must be used.

In computer adaptive testing

1. The sample sizes must be large enough to assure the stability of the IRT estimates.
2. Test takers and other stakeholders must be informed of the rationale of computer adaptive testing and given advice on test taking strategies for such tests.

第5回JLTA言語テストワークショップについて

来る、7月19日(土)、20日(日)に、茨城県水戸市の常磐大学にて、第5回 JLTA 言語テストワークショップを開催いたします(後援:茨城県教育委員会)。英語教育関係者(主に中学校・高等学校・大学の英語科教員を想定)を対象に、言語テストの理論と実践に関する基礎的な研修として、以下の内容を予定しています。最新の、詳細な日程につきましては、後日正式にお知らせいたします。また、随時、ホームページでもお知らせいたします。

7月19日(土)

12:30 - 13:00 受付

13:00 - 13:30 開会行事、オリエンテーション

13:30 - 15:30 講義「言語テストの基礎」

15:30 - 16:00 休憩

16:00 - 17:20 講義「英語教育における測定と評価の課題」

17:30 - 20:00 懇親会

7月20日(日)

08:45 - 09:00 受付

09:00 - 10:20 講義「テスト項目作成の留意点」

10:20 - 10:40 休憩

10:40 - 12:00 テスト項目作成実習

12:00 - 13:00 昼食

13:00 - 15:00 テストデータ分析実習

15:00 - 15:20 閉会行事

The 7th (2003) JLTA Annual Conference

Call for Papers

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

Date: October 25 (Saturday) 2003, 10:00-18:00

Place: Kumamoto University, Research Center for University Education
2-40-1 Kurokami, Kumamoto, 860-8555

The Japan Language Testing Association (JLTA) invites presentation proposals for its annual conference to be held in Kumamoto on October 25th, 2003. Presenters will have 30 minutes to present their paper, followed by 10 minutes for discussion.

Submission Requirements:

Please prepare a 250 word (or 500 Japanese character) abstract, including your name, affiliation, title of your presentation and contact information, and send it to the following address by e-mail (Hiroshi Shimatani) shima@educ.kumamoto-u.ac.jp The file name should be "JLTA 2003 Submission." The deadline for submission of presentation proposals is July 31, 2003. Notifications will be sent by e-mail by August 15, 2003.

Advertisements and Exhibitions:

Deadline : Sep. 20

Advertisement (half page of A4 size paper)

JLTA Members ¥5,000

Non-Members ¥10,000

Deadline : July 20

Advertisement (one page of A4 size paper)

JLTA Members ¥10,000

Non-Members ¥20,000
Deadline : July 20
Exhibition (one long table)
JLTA Members ¥20,000
Non-Members ¥40,000

For further information, please contact the JLTA Secretary General at the following number: TEL 026-275-1964, FAX 026-275-1970 or e-mail: youichi@avis.ne.jp (Youichi Nakamura).

日本語テスト学会 (JLTA) 第 17 回研究例会

日時: 2003 年 6 月 7 日(土)午後 13 時~ 17 時
会場: 北海学園大学経済学部
(札幌市豊平区旭町4-1)

プログラム:

報 告: 「北海学園大学の英語入試問題はどのように変わってきたか」

塩川春彦 (北海学園大学)

発表 I: 「外国語学習者の持つ言語能力の把握とその適切な測定」

浦野 研 (北海学園大学)

発表 II: 「絶対評価でのオーラルコミュニケーションの評価」

斉藤英敏 (北星学園大学)

中川 淳 (札幌稲北高等学校)

福田修平 (札幌稲北高等学校)

参加費: 会員 (JLTA・北海道英語教育学会)

無料、一般 資料代等として 1,000 円

共催: 北海道英語教育学会

問い合わせ先:

北海学園大学経営学部 塩川春彦

e-mail: shiokawa@econ.hokkai-s-u.ac.jp

2003 年度の JLTA 研究助成

2003 年度の JLTA 研究助成は、研究助成委員会、及び理事会の審査の結果、以下の 3 件 (応募順) とすることに決定いたしました。

1. 松野澄江 (名古屋大学・愛知県立大学非常勤講師) Self-assessment, peer-assessment, and teacher-assessment on Japanese university students' writing an English composition
2. Yasuyo Sawaki (Doctoral Candidate in Applied Linguistics, University of California, Los

Angels) A comparison of summary and free recall as reading comprehension tasks in a web-based placement test of Japanese as a foreign language

3. 小泉利恵 (筑波大学大学院博士 課程) 語彙知識によるスピーキング能力の予測可能性

研究助成委員会の審査結果報告は以下の通りです。

1. 1 月 31 日締め切りの研究助成に応募した方は 3 人おりました。規約により 3 件まで候補者を決定することになっていますので、本部から申請書類を研究助成委員会委員長に送付していただいた後に、委員長から各研究助成委員にその書類を郵送し、大変急いでご審査頂き 1 週間以内で、各研究助成委員から審査結果をお送り頂きました。なお、研究助成委員のうち、1 人は米国出張のため審査できず、研究助成委員会にすべて一任されました。
2. 審査の手順は以下の通りです。
(1) 研究助成委員は、研究助成委員会の規約 3. 「選考基準」に基づき、内容は独創性が高いか、内容は学術的意義が高いか、内容は教育的意義が高いか、の 3 つの観点から 5 点法による分析的評価を実施しました。
(2) 研究助成委員から審査結果を電子メールまたはファックスで送信していただいた後、委員長が応募者の研究計画についての評価を計算しました。応募者の総合点の平均点により 1 位から 3 位まで順位をつけました。規約により、3 件まで、候補者を決定することになっていますので、委員長、副委員長で検討し、原案を作成した後、委員の意見を聞き、最終的に研究助成委員会としての結論を出しました。
3. 審査の結果について
1 位、2 位、3 位ともほぼ同じレベルであると判断され、本年度は 3 件とも採択することに決定しました。今回は、研究計画を説明する際に、十分スペースをさいて、参考文献もあげて、説得的であるように、書かれていました。

以上の結果となりました。来年度も、規約にある通りの日程で募集を致しますので、今からご準備ください。

(研究助成委員会 委員長 望月昭彦
副委員長 卯城祐司)

日本語テスト学会第 7 回 (2003 年度)
全国研究大会のご案内と発表者募集

大会テーマ: 「言語学習とテスト: テスティングが言語学習に果たす役割」

