The 21st
Language Testing Research Colloquium

第廿一回言語テスト国際会議

Program and Abstracts

The Social Responsibility of Language Testing in the 21st Century

21世紀における言語テストの社会的責任

July 28-31

LTRC 99

The International Language Testing Association (ILTA)
The Japan Language Testing Association (JLTA)

Tsukuba International Congress Center

Tsukuba, Ibaraki Prefecture, Japan
July 28-31, 1999

LTRC 99 Organizing Committee

Joint-chairs	Kenji Ohtomo	*Tokiwa University* *Professor Emeritus, The University of Tsukuba*
	Randy Thrasher	*International Christian University*
Vice-chair	Yuji Nakamura	*Tokyo Keizai University*
Secretary General	Youichi Nakamura	*Nagano-ken Shinonoi High School*
Members-at-Large	Jeffrey Hubbell	*Hosei University*
	Masayoshi Kinoshita	*Fukuoka International University*
	Hideo Kiyokawa	*Wayo Women's University*
	Katsunosuke Namita	*Hokkaido University*
	Masashi Negishi	*Tokyo University of Foreign Studies*
	Kazuo Otsubo	*Reitaku University*
	Steven Ross	*Kwansei Gakuin University*

The Committee would like to thank the following organizations and individuals for their help:

The abstract evaluators

The officers of the International Language Testing Association (ILTA)

The Japan Language Testing Association (JLTA) for its generous financial support

The Educational Testing Service for its support of the Messick Memorial Lecture

Glenn Fulcher for setting up the web page that allowed electronic submission of abstracts

The members of JLTA and others in Japan who supported LTRC 99 with their donations.

Secretariat for the 21st Language Testing Research Colloquium

758 Shibahara Togura-machi Hanishina-gun
Nagano-ken, 389-0813, JAPAN

TEL +81-26-275-1964, FAX +81-26-275-1970

e-mail address: youichi@avis.ne.jp
URL: http://www.avis.ne.jp/`youichi/JLTA.html

Abstract Evaluators

Alister Cumming	*Ontario Institute for Studies in Education, University of Toronto*
Fred Davidson	*The University of Illinois, Urbana - Champaign*
Alan Davies	*The University of Edinburgh*
Hossein Farhady	*Iran University of Science and Technology*
Glenn Fulcher	*University of Surrey*
Jeffrey Hubbell	*Hosei University*
Antony Kunnan	*California State University*
Yuji Nakamura	*Tokyo Keizai University*
Steven Ross	*Kwansei Gakuin University*
Mary Spaan	*The University of Michigan*
Carolyn Turner	*McGill University*

The Executive Board of the International Language Testing Association

President	Elana Shohamy	*Tel Aviv University*
Vice-President	Alan Davies	*The University of Edinburgh*
Secretary-Treasurer	Caroline Clapham	*Lancaster University*
Immediate Past President	Tim McNamara	*The University of Melbourne*
Members-at-Large	Glenn Fulcher	*The University of Surrey*
	Brian Lynch	*The University of Melbourne*
	Carolyn Turner	*McGill University*
	Mari Wesche	*University of Ottawa*

LTRC 99

and

JLTA Annual Conference

Program

JLTA Annual Conference and LTRC 99 Program

Wednesday July. 28th

The Japan Language Testing Association (JLTA) Annual Conference

12:30-13:00 **JLTA Business Meeting**

Chair: Yuji Nakamura *JLTA Vice-President*

Opening Address	Kenji Ohtomo *JLTA President*
Report on Activities	Youichi Nakamura *JLTA Secretary General*

13:00-16:30 **Plenary Speeches and Invited Paper**

Chairs: Haruo Yanai *National Center for University Entrance Examination*
Jeffrey Hubbell *Hosei University*

13:00-14:30 Hiroshi Ikeda *Professor Emeritus, Rikkyo University*
What We Need for Research on Language Testing in Japan: A Psychometrician's View
(in Japanese)

14:30-15:00 Hossein Farhady *Iran University of Science and Technology*
Coaching and the Consequential Validity of TOEFL vs. Task Based Assessment

15:00-16:30 Nancy Cole *President, Educational Testing Service*
Language Assessment in the Future

The 21st Language Testing Research Colloquium (LTRC 99)

17:00-17:30 **Opening Ceremony**

Chair: Randy Thrasher *ICU* Co-chair: Masayoshi Kinoshita
Fukuoka International University

Welcoming Address	Kenji Ohtomo, Joint-chair, LTRC 99 *Tokiwa University*
Opening Remarks	Elana Shohamy, ILTA President *Tel Aviv University*

17:30-19:00 **Plenary Speech One (Messick Memorial Lecture)**

Chair: Randy Thrasher *ICU* Co-chair: Hiroyoshi Hatori *Bunkyo Woman's College*

Tim McNamara *The University of Melbourne*
Validity in Language Testing: the Challenge of Sam Messick's Legacy

19:00-21:00 **Welcome Party**

Toast Master: Yuji Ushiro
The University of Tsukuba

Toast Mistress: Taiko Tsuchihira
Shinshu Honan Junior College

Thursday July 29th

Student Research Reports

Chair: Gene B. Halleck *Oklahoma State University*

Co-chair: Matsuo Kimura *Aoyama Gakuin University*

8:15-8:45 Yoshihito Sugita *Yamanashi University*
Constructing Rating Scales for EFL Writing Tests

8:45-9:15 Dongil Shin *The University of Illinois at Urbana-Champaign*
Validation of Diagnostic ESL Test for Measuring Lecture Comprehension Ability

9:15-9:45 Amy Yamashiro *Nihon University*
Using Structural Equation Modeling to Validate a Rating Scale

9:45-10:00 Coffee Break

10:00-11:30 **Panel Discussion: Language Testing: Yesterday, Today and Tomorrow**

Chair: Kenji Ohtomo *Tokiwa University*

Co-chair: Katsunosuke Namita *Hokkaido University*

Bernard Spolsky *Bar-Ilan University*
Yesterday: The path taken

Elana Shohamy *Tel Aviv University*
Today: Which paths? What do they mean for us and for others?

Alan Davies *The University of Edinburgh*
Tomorrow: How many paths?

11:30-12:00 **Group Photo**

12:00-13:15 Lunch

Research Reports Session One

Chair: Hossein Farhady *Iran University of Science and Technology* Co-chair: Minoru Wada *Meikai University*

13:15-13:45 Gary Buck, Kikumi Tatsuoka and Irene Kostin *Educational Testing Service*
Developing and Cross-Validating a Set of Cognitive and Linguistic Attributes Applicable to Multiple Test Forms on a Multiple-Choice Listening Test

13:45-14:15 Hisami Saito-Scott *Mission College, Santa Clara, CA*
Analyzing the Dimensionality of a Second Language Reading Test from Cognitive Perspectives

14:15-14:45 J. D. Brown, Thom Hudson and John Norris *The University of Hawaii at Manoa*
Validation of Test-Dependent and Task-Independent Ratings of Performance Assessment

14:45-15:15 Yong-Won Lee *Educational Testing Service*
Examining Passage Related Local Dependence (LD) Using Q3 Statistics in an EFL Reading Comprehension Test

15:15-15:30 Coffee Break

Research Reports Session Two

Chair: Mary C. Spaan *The University of Michigan* Co-chair: Yoshinori Watanabe *ICU*

15:30-16:00 Catherine Burrows *NSW Adult Migrant English Service*
Adopters, Adapters, and Resisters: Did the Assessment of the Certificates in Spoken and Written English Change Teaching in the AMEP?

16:00-16:30 Jane Andrews and Richard Fay *The University of Manchester*
Interculturality and English Language Paired Oral: Exploring Assessors Perceptions

16:30-17:00 Laura MacGregor *Sophia University*
Demystifying the STEP Interview Test

17:00-17:30 Junko Matsui *Meikai University*
Various Interrelated Factors Involved in Listening

17:30-18:00 Tetsuhito Shizuka *Kansai University*
Using Test-takers's Confidence Levels in Scoring Multiple-choice Tests: Rationale, Empirical Findings, and Simulation Results

18:00-19:00 **ILTA Business Meeting**

Friday July 30th

Poster Session One

Chair: Steven Ross *Kwansei Gakuin University*
Co-chair: Ken Norizuki *Shizuoka Sangyo University*

8:15-9:15	Presentation Session (Second Floor Convention Hall 200)
9:15-10:00	Discussion Session (First Floor Conference Room 102)

Sarah Briggs and Barbara Dobson *The University of Michigan*
Using a Spoken Language Corpus in the Development of EAP Listening Tests

Gene B. Halleck *Oklahoma State University* and Daniel J. Reed *The University of Minnesota*
Rating the VOCI: Alternative Methods

Ari Huhta, Anne Pitkanen-Huhta, and Paula Kalaja *The University of Jyvaskyla*
Taking a High Stakes Test--Exploring its Meaning(s) for an Individual

Akihiko Mochizuki *The University of Tsukuba* and Noboru Yamada *Shizuoka Sangyo University*
Measurement and Evaluation of English Communicative Competence--Discrete-point vs. Integrative Tests and Integrative vs. Analytical Evaluation

Youichi Nakamura *Nagano-ken Shinonoi High School*
A Prototype of an Item Pool for Computer-Based, Multimedia Listening Test

Hiroyuki Noguchi *Nagoya University*
IRT Analyses of the Japanese Language Proficiency Test

Reiko Saegusa *Hitotsubashi University,* Mako Aoyama *The Japan Foundation,*
Tsutomu Fukuchi *Chuo University,* Sukero Ito *Tokyo University of Foreign Studies,*
Mikio Kawarazaki *Tokai University,* Keiichi Koide *Gumma Prefectural Women's University,*
Takashi Murakami and Hiroyuki Noguchi *Nagoya University,*
Kazuo Otsubo *Reitaku University,* and Akiko Wada *The Japan Foundation*
Verifying the Validity of the Japanese Language Proficiency Test Using Can-do Statements

Elaine Wylie *Griffith University*
Testing for Justice

Yoshinori Watanabe *International Christian University*
Exploring Washback in Japanese EFL Classrooms

10:00-10:15 Coffee Break

Research Reports Session Three

Chair: Caroline Clapham *Lancaster University*
Co-chair: Mikihiko Sugimori *Ritsumeikan University*

10:15-10:45 Batia Laufer *University of Haifa* and Paul Nation*Victoria University of Wellington*
Vocabulary Recognition Speed Test (VORST)

10:45-12:15 **Plenary Speech Two**

Chair: Charles Alderson *Lancaster University/British Council* Co-chair: Seiki Saito *Kanagawa University*

Ikuo Amano *Professor, Center For National University Finance*
Professor Emeritus, Tokyo University
Admission Policies in Japanese Universities

Simultaneous Interpretation: Yuji Nakamura *Tokyo Keizai University* and Randy Thrasher *ICU*

12:15-13:15 Lunch

13:15-15:15 **Symposium One: National Language Testing Reform: Possibilities and Constraints**

Chair: Antony Kunnan *California State University* Co-chair: Kazuo Otsubo *Reitaku University*

Geoff Brindley *Macquarie University*
Issues and Problems in National Language Testing Reform

Charles Alderson *Lancaster University*
Language Testing Reform in Hungary

Lyle Bachman *The University of California, Los Angeles*
Liying Cheng *The University of Alberta*
Liz Hamp-Lyons *Hong Kong Polytechnic University*
Language Testing Reform in Hong Kong

Bernard Spolsky *Bar-Ilan University*
Language Testing Reform in Israel

Sun Yurong *National Educational Examinations Authority, China*
Michael Milanovic and Lynda Taylor *University of Cambridge Local Examinations Syndicate, UK*
Setting up a dynamic language testing system in national language test reform: the Public English Test System (PETS) in China

Antony Kunnan *California State University*
Beryl Meiron *Cambridge Examinations and IELTS International*
Yutaka Kawamoto *California State University*
A Training Model for Egyptian Language Testers

Research Reports Session Four

Chair: Janna Fox *Carleton University* Co-chair: Masashi *Negishi* *Tokyo University of Foreign Studies*

15:30-16:00 Charles Alderson *Lancaster University/British Council*, Szabo Gabor *Janus Pannonius University, Pecs, Hungary,* and Richard Percsich *Assessment Centre of BMPI (Pedagogical Institute), Pecs, Hungary*
Sequencing as an Item Type

Poster Session Two

Chair: Glenn Fulcher *The University of Surrey*
Co-chair: Tetsuro Chihara *Osaka Jogakuin Junior College*

16:00-17:00 Presentation Session (Second Floor Convention Hall 200)
17:00-17:45 Discussion Session (First Floor Conference Room 102)

Ofra Inbar *Tel-Aviv University*
Language Testing Courses and Language Teaching Initiatives: A Collaborative Process

Akihiro Ito *Aichi Gakuin University*
Testing English Tests: A Concurrent and Internal Construct Validation of the NCUEE-Test in Japan

Masayoshi Kinoshita *Fukuoka International University*, Hiroshi Shimatani *Kumamoto University*, Terry Laskowski *Kumamoto University* , Hiroki Yamamoto *Seinan Jogakuin Junior College*, and Masashi Takemura *Hokkaido Sapporo Kita High School*
The Analysis of English Listening and Reading Comprehension Concerning Japanese and Korean High School Students

Mikiya Koarai *St. Dominique's Institute*
Performance of Japanese Examinees on TOEFL Section II

Naoki Kuramoto *Tohoku University*, Masahiro Kasai *DePaul University*, and Hisami Saito Scott *Mission College, Santa Clara, CA*
Verification of a Japanese Vocabulary Test by the Rule Space Methodology

Jouji Miwa *Iwate University*
The Dictation Test and Its Evaluation for Japanese Speech Using A Portable Computer

Masanori Nakamura and Nunzio Scena *Georgia State University*
Oral Communication Test: Japanese Speakers of English

Daniel L. Robertson *The University of Guam*, and Charles W. Stansfield *Second Language Testing, Inc*
Ten Years Later: The Guam Educators' Test of English Proficiency

Matilde Scaramucci *The State University of Campinas, SP, Brazil*
A Study of Washback in Brazil

Mary C. Spaan *The University of Michigan*
Converting from a General Purpose Writing Task to a Special Purpose Task

19:00 **Banquet**

Toast Master: Shinji Kimura *Kwansei Gakuin University*
Toast Mistress: Yuko Shimizu *Ritsumeikan University*

Saturday July 31st

Research Reports Session Five

Chair: Elaine Wylie *Griffith University* Co-chair: Shien Sakai *Sakurano Seibo Junior College*

8:15-8:45 John Read *Victoria University of Wellington*
The Impact of English Tests for Migrants

8:45-9:15 Jared Bernstein *Ordinate Corporation*, Maxine Lipson *The University of Bologna*, Gene B. Halleck *Oklahoma State University*, and Jane Martinez-Scholze *Defence Language Institute Lackland AFB Texas*
Comparison of Oral Interviews and Automatic Testing of Spoken Language Facility

9:15 -9:30 Coffee Break
9:30-11:30 **Symposium Two: The Rights and Responsibilities of Testers and Test-takers in Language Testing Settings: Ethics, Policy, Practice and Research**

Chair: Bernard Spolsky *Bar-Ilan University* Co-chair: Kazuo Amma *Tamagawa University*

Elana Shohamy, *Tel Aviv University*
Critical Language Testing: Uses and Consequences of Tests, Responsibilities of Testers and Rights of Test-takers

Tim McNamara, *The University of Melbourne*
Policy and Social Considerations in Language Assessment: an Overview of Recent Research

Geoffrey Brindley, *Macquarie University*
Understanding Assessment Cultures

Janna Fox, *Carleton University*
Test Taker and Rater Responses: A Study of Proximal Processes

Cathie Elder, *The University of Melbourne*
Language Testing: Whose Values? Whose Responsibility?

David Nevo, *Tel Aviv University*
Tester and Test-taker: A Two-way Channel of Communication

Alan Davies *The University of Edinburgh*
Professionalism in Language Testing: Do Codes of Ethics help?

11:30-12:00 **Closing Session**

Chair: Randy Thrasher *ICU* Co-chair: Yukie Koyama *Nagaoka University of Technology*

Closing Remarks	Alan Davies, ILTA Vice-President *The University of Edinburgh*
Invitation to LTRC 2000	Lyle Bachman *The University of California, Los Angeles*

10:45-12:15 **Plenary Speech Two**

Chair: Charles Alderson *Lancaster University/British Council*
Co-chair: Seiki Saito *Kanagawa University*

Ikuo Amano *Professor, Center For National University Finance*
Professor Emeritus, Tokyo University
Admission Policies in Japanese Universities

Simultaneous Interpretation: Yuji Nakamura *Tokyo Keizai University* and Randy Thrasher *ICU*

12:15-13:15 Lunch

13:15-15:15 **Symposium One: National Language Testing Reform: Possibilities and Constraints**

Chair: Antony Kunnan *California State University*
Co-chair: Kazuo Otsubo *Reitaku University*

Geoff Brindley *Macquarie University*
Issues and Problems in National Language Testing Reform

Charles Alderson *Lancaster University*
Language Testing Reform in Hungary

Lyle Bachman *The University of California, Los Angeles*
Liying Cheng *The University of Alberta*
Liz Hamp-Lyons *Hong Kong Polytechnic University*
Language Testing Reform in Hong Kong

Bernard Spolsky *Bar-Ilan University*
Language Testing Reform in Israel

Sun Yurong *National Educational Examinations Authority, China*
Michael Milanovic and Lynda Taylor *University of Cambridge Local Examinations Syndicate, UK*
Setting up a dynamic language testing system in national language test reform: the Public English Test System (PETS) in China

Antony Kunnan *California State University*
Beryl Meiron *Cambridge Examinations and IELTS International*
Yutaka Kawamoto *California State University*
A Training Model for Egyptian Language Testers

Research Reports Session Four

Chair: Janna Fox *Carleton University*
Co-chair: Masashi *Negishi* *Tokyo University of Foreign Studies*

15:30-16:00 Charles Alderson *Lancaster University/British Council*, Szabo Gabor *Janus Pannonius University, Pecs, Hungary,* and Richard Percsich *Assessment Centre of BMPI (Pedagogical Institute), Pecs, Hungary*
Sequencing as an Item Type

Poster Session Two

Chair: Glenn Fulcher
The University of Surrey

Co-chair: Tetsuro Chihara
Osaka Jogakuin Junior College

16:00-17:00 Presentation Session (Second Floor Convention Hall 200)
17:00-17:45 Discussion Session (First Floor Conference Room 102)

Ofra Inbar *Tel-Aviv University*
Language Testing Courses and Language Teaching Initiatives: A Collaborative Process

Akihiro Ito *Aichi Gakuin University*
Testing English Tests: A Concurrent and Internal Construct Validation of the NCUEE-Test in Japan

Masayoshi Kinoshita *Fukuoka International University*, Hiroshi Shimatani *Kumamoto University*, Terry Laskowski *Kumamoto University* , Hiroki Yamamoto *Seinan Jogakuin Junior College*, and Masashi Takemura *Hokkaido Sapporo Kita High School*
The Analysis of English Listening and Reading Comprehension Concerning Japanese and Korean High School Students

Mikiya Koarai *St. Dominique's Institute*
Performance of Japanese Examinees on TOEFL Section II

Naoki Kuramoto *Tohoku University*, Masahiro Kasai *DePaul University*, and Hisami Saito Scott *Mission College, Santa Clara, CA*
Verification of a Japanese Vocabulary Test by the Rule Space Methodology

Jouji Miwa *Iwate University*
The Dictation Test and Its Evaluation for Japanese Speech Using A Portable Computer

Masanori Nakamura and Nunzio Scena *Georgia State University*
Oral Communication Test: Japanese Speakers of English

Daniel L. Robertson *The University of Guam*, and Charles W. Stansfield *Second Language Testing, Inc*
Ten Years Later: The Guam Educators' Test of English Proficiency

Matilde Scaramucci *The State University of Campinas, SP, Brazil*
A Study of Washback in Brazil

Mary C. Spaan *The University of Michigan*
Converting from a General Purpose Writing Task to a Special Purpose Task

19:00 **Banquet**

Toast Master: Shinji Kimura
Kwansei Gakuin University

Toast Mistress: Yuko Shimizu
Ritsumeikan University

Saturday July 31st

Research Reports Session Five

Chair: Elaine Wylie *Griffith University* Co-chair: Shien Sakai *Sakurano Seibo Junior College*

8:15-8:45 John Read *Victoria University of Wellington*
The Impact of English Tests for Migrants

8:45-9:15 Jared Bernstein *Ordinate Corporation*, Maxine Lipson *The University of Bologna*, Gene B. Halleck *Oklahoma State University*, and Jane Martinez-Scholze *Defence Language Institute Lackland AFB Texas*
Comparison of Oral Interviews and Automatic Testing of Spoken Language Facility

9:15 -9:30 Coffee Break
9:30-11:30 **Symposium Two: The Rights and Responsibilities of Testers and Test-takers in Language Testing Settings: Ethics, Policy, Practice and Research**

Chair: Bernard Spolsky *Bar-Ilan University* Co-chair: Kazuo Amma *Tamagawa University*

Elana Shohamy, *Tel Aviv University*
Critical Language Testing: Uses and Consequences of Tests, Responsibilities of Testers and Rights of Test-takers

Tim McNamara, *The University of Melbourne*
Policy and Social Considerations in Language Assessment: an Overview of Recent Research

Geoffrey Brindley, *Macquarie University*
Understanding Assessment Cultures

Janna Fox, *Carleton University*
Test Taker and Rater Responses: A Study of Proximal Processes

Cathie Elder, *The University of Melbourne*
Language Testing: Whose Values? Whose Responsibility?

David Nevo, *Tel Aviv University*
Tester and Test-taker: A Two-way Channel of Communication

Alan Davies *The University of Edinburgh*
Professionalism in Language Testing: Do Codes of Ethics help?

11:30-12:00 **Closing Session**

Chair: Randy Thrasher *ICU* Co-chair: Yukie Koyama *Nagaoka University of Technology*

Closing Remarks Alan Davies, ILTA Vice-President *The University of Edinburgh*

Invitation to LTRC 2000 Lyle Bachman *The University of California, Los Angeles*

Abstracts

JLTA Annual Meeting

Hiroshi Ikeda

Hossein Farhady

Nancy Cole

Plenary Speech One

Tim McNamara

Plenary Speech Two

Ikuo Amano

Panel Discussion

Bernard Spolsky

Elana Shohamy

Alan Davies

Symposium One

Geoff Brindley

Charles Alderson

Lyle Bachman, Liying Cheng, and Liz Hamp-Lyons

Bernard Spolsky

Sun Yurong, Mike Milanovic, and Lynda Taylor

Antony Kunnan, Beryl Meiron, and Yutaka Kawamoto

Symposium Two

Elana Shohany

Tim McNamara

Geoff Brindley

Janna Fox

Cathie Elder

Davis Nevo

Alan Davies

Research Reports

Arranged alphabetically by the name of the presenter

Student Research Reports

Arranged in the order of presentation

Poster Session One

Arranged in the order of presentation

Poster Session Two

Arranged in the order of presentation

JLTA Annual Meeting

What We Need for Research on Language Testing in Japan: A Psychometrician's View

Hiroshi Ikeda *Professor Emeritus, Rikkyo University* (Wednesday 13:00 - 14:30)

From my personal view as a pychometrician, what are most needed for research on teaching and testing of foreign languages in the current Japanese education scene are:

1. More use of modern technological developments in language teaching and testing, which include computer-based testing with multimedia devices for listening and speaking, as well as reading and writing. Traditional paper and pencil tests should be no longer the major device for language testing as has been the case up to now.
2. Application of modern test theories to language testing programs which include, for example, item response theory, computerized adaptive testing (CAT) in particular, and generalizability theory for rater's evaluation of performance assessment.
3. Building test-item banks with added-value of test information which is the most important in Japanese testing procedures. As we have no habit of reusing the same test items, absence of accumulated item data banks makes it very difficult to build a CAT system.
4. More validation studies on language testing in terms of scientific research methods which include not only the multitrait-multimethod conception of research design but also reconsideration of what constructs can be measured if we go beyond paper and pencil tests.
5. More network exchange with data or evidence based information among teachers and research workers which will surely accelerate the progress of research on testing in Japan and consequently the revision of teaching methods for efficient learning of English as a second language.

When these problems are resolved, at least in part, I believe a fairly good improvement of language testing research in Japan can be expected and this progress will lead to the discovery of better ways of teaching and learning than we have ever had.

Coaching and the Consequential Validity of TOEFL vs. Task Based Assessment

Hossein Farhady *Iran Univ ersity of Science and Technology* (Wednesday 14:30 - 15:00)

TOEFL is widely used as a certificating device, and is strongly claimed accountable by the people utilizing it. However, there are indications of the vulnarability of TOEFL to the teaching of test taking strategies. The purpose of the present study is to provide empirical evidence that coaching toward TOEFL may invalidate its results as indicators of test takers proficiency level. To test the hypothesis, the scores of 200 subjects on a TOEFL and on a task based language proficiency test were compared. The subjects were selected from among those who attended coaching classes for TOEFL. The findings revealed that coaching leads to spurious scores and consequently to the unaccountability of TOEFL results. Theoretically, it is claimed that teaching towards task based tests would not be harmful to pedagogy and learning, because performance on such tests are simulations of the real life tasks the learners will be expected to perform. This implies that testing practitioners should move toward the use of more authentic and performance based assessment to alleviate some of the problems associated with the consequences of the decisions made on the basis of TOEFL type tests.

Further, the implications of this study, which support the safety of coaching toward a particular task, domain, or subdomain in order to enhance the students' achievement on the one hand, and the cosequential validity of the tests, on the other, are discussed.

Language Assessment in the Future

Nancy Cole *Educational Testing Service* (Wednesday 15:00 – 16:30)

Language assessment will undergo revolutionary change in the next decade. The richer theoretical understanding of language use provides one basis for the change. The new possibilities with modern electronic technologies provide the other. Together the time is ripe for radical change in the assessment of language. In this paper, I will summarize the advances in language theory and in technology that will help to create the changed assessment approaches and to suggest some scenarios of new language assessment that I expect to see by 2010.

Language Assessment of the Past and Today

We set the context by understanding key features of language assessment today. Today assessment of language skills focus largely on recognizing correct structure in written form, on correct or standard usage, on deriving explicit information from written text, and on deriving several types of implicit information from written text. Written text in a test booklet is the focus for assessing reading and for asking the student to judge correctness of text. Writing, speaking and listening are pursued only through supplementary cumbersome and expensive procedures, if at all.

The testing approaches require the same test to be given to many people from different backgrounds and experiences and therefore require that language skills be abstracted from specialized contexts to quite general and, hopefully, contexts familiar to all examinees. The most familiar form of adaptation to such differences is some limited choice the student may have for two or more contexts of reading passages or writing topic, for example.

Measures based within these constraints have been quite powerful for many purposes. However, the possibilities for moving beyond these constraints are now very promising.

Advances in Language Theory

Language theorists have helped us think increasing of language in use, beyond the abstract, structural characteristics on which we focused in the past. Language in use has some critical new features that we have not assessed directly in large scale testing. Various combinations of the reading, writing, listening, speaking skills are integrated in more cases of using language. Uses of language are context-based with many clues to appropriateness and meaning coming from the context as well as language. Language is used in situations in general situations of everyday living as well as in highly specialized settings of professional work.

The integration of language, of context in relation to language, and of use in different life sphere are all features difficult for past testing technologies to manage and not especially well understood by many users of language assessment or, perhaps, teachers of language.

New Assessment Technologies

Electronic technologies bring enormous possibilities for accommodating the new thinking about language. It is now possible to create tests involving reading, actual writing, speaking, and listening in various combinations in language use. Today we can assess speaking skill electronically, evaluate electronically the adequacy of text written by an examinee, and pose listening tasks integrated with other skills or in isolation.

The capability to adapt assessment to individual characteristics provides new possibilities for managing and varying context. The ability to branch to tasks matched in some way to the individual's home country or first language, for example, suggest many, as yet unrealized possibilities for testing language in more context-rich situations familiar to (or, by design, unfamiliar to) the examinee. The ability to branch the testing process allows the use of content in a specialized area familiar to the examinee.

In addition to making it possible to present tasks and record answers electronically in speaking, writing, and listening, advances in electronic scoring add to the practical feasibility of these approaches. Examples of various new forms of assessment based in electronic technology will be illustrated.

Progress To Date

Small steps in many of these new directions are included in the TOEFL now given on computer in many parts of the world. Examples will be used to demonstrate progress. However, we foresee much larger steps as we learn to implement some of the research coming from the TOEFL 2000 project. Examples of such next steps will be used to illustrate further progress.

Plenary Speech One

Wednesday 17:30-19:00

Messick Memorial Lecture

Tim McNamara *The University of Melbourne*

Validity in Language Testing: The Challenge of Sam Messick's Legacy

Samuel Messick's influence on research in language testing has been great. There are two main areas on influence: on our understanding of how inferences made on the basis of test performances can be challenged and supported, and through the specific notion of consequential validity. The former has affected the work of many researchers in our field, most notably that of Bachman and Palmer on validity and the design of language tests. It has also been important in work on performance assessment, including the conceptualization of the validation research required in major test development projects. Messick's writing on consequential validity has informed debate on ethics, impact, accountability and washback in language testing in the work of a number of leading researchers.

But the character of Messick's work challenges us in many further ways. Messick located validity theory in the area of values, and thus entered the debate on the epistemology of research in the behavioural and social sciences. The paper explores the implications of this position, and further outlines a range of other issues that advances in other fields are raising for language testing. Tackling these questions is the real challenge of Messick's legacy.

Plenary Speech Two

Friday 10:45-12:15

Ikuo Amano *Center For National University Finance Professor Emeritus, Tokyo University*

Admission Policies in Japanese Universities

It is internationally known that Japanese universities depend too much on entrance examinations as a method to select entrants and that this has caused severe competition among applying students.

A British sociologist, Ronald Dore, for example, states that the "Entrance Examination Hell" facing university applicants is the most important feature of education in Japan. Moreover, an official report released by the Organization for Economic Co-operation and Development (OECD) claims that the social rank we belong to is determined at the time of the various entrance examinations, and names this phenomenon "the social birth which takes place after the biological birth."

It is true that neither Japanese style entrance examinations nor the same degree competition for university entrance can be seen in Europe or America. However, if we look at East Asian countries such as Korea and China, we cannot help but notice that the entrance examination problem is not unique to Japan but common to all of these countries.

Furthermore, even in western countries where entrance is not determined by a single examination, the selection of university entrants is based on their scholastic abilities. Also, various types of tests are used as means of assessing scholastic ability.

This fact suggests that, in all countries, universities face the problem of how to select students and that at the same time the history, culture, social structure, structure of the educational systems, and development of the education level of each country generates individual differences in the selection process.

This lecture discusses the past history, the present situation and the future direction of the selection of university entrants in Japan, by taking into consideration various international viewpoints.

Panel Discussion

Thursday 10:00-11:30

Language Testing: Yesterday, Today and Tomorrow

Bernard Spolsky *Bar-Ilan University* (former President, ILTA)

Yesterday: The Path Taken

Two critical developments in the nineteenth century helped determine the course of testing in general and language testing in particular for this century. In post-revolutionary France, in pre-imperial Prussia, and in industrializing Britain, tests and examinations were recognized as a powerful means of central control over educational systems and as a method of gatekeeping for jobs and higher education. Only at the end of the century did statisticians point out the fatal flaw, the inevitable uncertainly in human measurement. In the early part of the current century, two further important steps were taken, centered largely in the US though with European acceptance later. The first was to build a science of testing whose major task was seen as reducing the level of uncertainty. The second was a successful propaganda effort, started after the first world war, to sell the new tests as fair and efficient. Language testing too fell under these influences, encouraging the development of central examinations aimed to control curricula and of industrialized tests useful for assisting limits on immigration and school admission. Only towards the end of the century have voices been raised challenging was what was noted as the encroaching power of examinations and suggesting that instead of a fruitless attempt to increase accuracy, one might rather look at the ethical and responsible use of flawed instruments.

Elana Shohamy *Tel Aviv University* (President, ILTA)

Today: Which Paths, and What Do They Mean For Us and For Others?

Current paths in language testing will be analyzed from the perspectives of theory, research and practice. The relevance of these paths to language testing as well as to other disciplines which are tangents to language testing will be examined. These examinations will be based on interviews and observations with a number of leading figures in language testing and of the tangent fields who express their perceptions of our relevance for them. Conclusions as to who we are and how other applied linguists and testers perceive us will lead to recommendations as to new and revised paths that could be pursued in language testing.

Alan Davies *University of Edinburgh* (President-elect, ILTA)

Tomorrow: How Many Paths?

In language testing, as in language teaching and more generally in linguistics, there is a permanent tension between the ideal and the real. In linguistics this tension displays itself in the opposition between formal and functional views of language, between UG context-free theories and interactive context-full theories. In language teaching and in language testing this same opposition is observed: over the last 50 years the definition of language proficiency changes and different kinds of evidence sought as the paradigm has shifted, first to the communicative and then to the postmodern and the ethical. Underlying all such changes is the difficulty, perhaps the impossibility of on the one hand confining language to its abstract nature and on the other determining what is linguistically important in the ongoing flux of multivariable real-life use. Views of language and the practice of language testing will no doubt change over the next 50 years but it seems likely that the underlying tension will remain. How far communication technology will facilitate our ability to take account of these divergent views remains to be seen.

Symposium One

Friday 13:15 – 15:15

National Language Testing Reform: Possibilities and Constraints

Many language testing researchers in the past decade have spent considerable time directing and assisting in national language testing reform efforts in various countries around the world. Generally, these efforts can be classified into two types of projects: projects that have attempted to refocus existing national language testing systems for many levels on a broad scale and projects that have been more modest where the efforts have been to develop or improve specific tests or examinations. The first type involves the testing reform effort to consider many related national educational and societal issues such as developing a national testing policy within changing educational policy and societal needs and priorities, and the impact of testing reform on the curriculum, the textbooks, the teaching practices, and the school students, to name a few. Other matters that need attention include the training of local test developers, administrators and researchers and the establishment of a testing system that will be sustainable after the project funding has been exhausted. Examples of this type of reform effort that have considered some of these matters include the Target Oriented Curriculum Renewal initiative for schools in Hong Kong (see Bachman and Sou, 1998), the Public English Language Testing System reform in China (see Yumin, Quingsi, Milanovic and Taylor, 1998), the Central Board of Secondary Education Curriculum renewal project in India (see Matthew, 1997), and the Sri Lanka Impact Project see Wall and Alderson, 1993). The second type is much more modest in scope and though the contexts may be different, relevant matters to consider include local testing policy, the local impact of the testing reform on the teaching and learning and the teachers and students, and training of local test developers, test administrators, test researchers and test maintenance. Examples of projects of this kind include the Baltic States Year 12 Examination Project (see Wall 1996), the development of the Practical English Test in Poland (see Defty and Kusiak, 1997), and the development of the English Language Placement Test in Namibia (see Campbell, 1998). As these examples are diverse in terms of the scope of the testing reform and the countries represented, this Colloquium will describe some of these and other efforts but will discuss the commonalities among these testing reform efforts and consider some critical questions that will illuminate the reform process. Some of the questions/issues that will be discussed are:
' How and who initiates testing reform?
' What are the factors that help initiate and facilitate testing reform?
' What are the factors that mitigate against or constrain the testing reform?
' What type of training should be given to local test reformers?
' How can success or failure be estimated?
' Where can resistance to change be expected?
' How can sustainability of testing reform be ensured after the effort ends?

1. Geoff Brindley *Macquarie University*

Issues and Problems in National Language Testing Reform

In recent years, educational authorities around the world have put major efforts into the development of new national language testing and assessment systems in school and adult education, often in response to demands for greater accountability and cost-effectiveness. These innovations have ranged from standardized proficiency tests to outcomes-based assessment and reporting systems based on teacher-conducted assessments. However, the introduction of national testing reform has not been without problems, owing to a complex set of political and educational influences which have affected the way in which new testing systems are designed and implemented.

This paper will identify a number of key issues and problems which arise in national language testing reform. The first part of the paper will consider some of the political factors which shape the course of testing innovation. In this context, the potential conflict between political, bureaucratic and educational perspectives on the purposes of testing and assessment will be discussed. Issues of implementation will then be examined in the light of insights from the literature on educational change. It will be argued that testing reform needs to be a collaborative exercise involving genuine consultation between all stakeholders if it is to succeed. Finally, some practical implications arising from the adoption of large-scale performance testing systems will be outlined. In particular, the need to provide adequate infrastructural support in the form of professional development programs, ongoing advice and quality control will be highlighted.

2. Charles Alderson *Lancaster University* ; *British Council Adviser to Hungarian Examinations Reform Project*

Language Testing Reform in Hungary

After radical changes in Hungarian political life in the late 1980s, Hungary has moved towards joining Western institutions. It has recently joined NATO and is one of the first-wave Central European states preparing for accession to the EU in the 21st century. Hungary is concerned to ensure that EU access will bring mobility of students and labour. It therefore wishes to enhance the foreign language competence of the studying and working populations, and to achieve international recognition and comparability for its educational certificates. In order to achieve this recognition for its language certificates, Hungary hopes to align them with the Common European Framework of reference of the Council of Europe.

As part of the educational reform process, Hungary has developed a new National Curriculum for school years 1-10, has declared its intention to create a new set of examinations at the official end of school in Year 10, and to reform Year 12 examinations, by the year 2004. Examination Reform will cover all curricular subjects, and currently involves developing Detailed Requirements for the Year 10 and Year 12 examinations (in the absence of a relevant national curriculum for Years 11 and 12), test specifications for Years 10 and 12 and sample tests.

In the case of English, an initial Baseline Study of the state of English Language Education in Hungary established current policies and examinations, the nature of teaching in schools, and the level of achievement of the English-learning school population. This Baseline Study has not only informed the planning of the exam reform project, but will also be used in an evaluation of the impact of the project in due course. Item writing teams have been created and trained, items at three hopefully different levels of achievement have been written and edited. The first pilot tests have been administered and are currently being analysed. Assessment criteria for the evaluation of performance on writing and speaking tests have been developed and are being refined. In addition, an in-service course has been written and pilotted to familiarise teachers with the principles, methods and content of the proposed new examinations, and to advise teachers on how they might most appropriately prepare their students for these examinations. In addition, a publicity campaign is being planned, and contacts have been established with Ministry officials in order to inform them of developments and concerns and to help to influence possible policy making.

This paper will report on this ongoing project, partially funded by the British Council. In particular it will explore aspects of the examination reform project that might contribute to its success, and others that might endanger it, paying particular attention to the role of individuals, institutions and culture in the politics of examination reform.

3. Lyle Bachman *The University of California, Los Angeles*
Liying Cheng *The University of Alberta*
Liz Hamp-Lyons *Hong Kong Polytechnic University*

Language Testing Reform in Hong Kong

Language testing policy in Hong Kong has been closely associated with language policy in education, and reform in both has historically been initiated and implemented by a central, relatively authoritarian, government, largely to meet the perceived economic and political needs of the country. Although region-wide public language examinations for certification and matriculation have been in place in Hong Kong's schools since the 1940's, reforms in these systems have been relatively recent. One reform, the establishment of the Hong Kong Examinations Authority in 1977, was intended to improve and rationalize the examination system as well as to make more efficient use of limited and specialized human and technical resources. This centralization has resulted in the consolidation of the administration of all school leaving examinations under one agency. More recently, reforms in the curricula for primary and secondary schools and in the training and certification of school teachers have included major changes in school-based language assessment systems. The most recent reform initiative is at the tertiary level, where there is a felt need to assure that university graduates are sufficiently proficient in Chinese and English to meet the personnel needs of the Hong Kong's financial and service industries.

In this presentation we will provide an overview of language testing policy and reform in Hong Kong, focussing on language testing in three areas: 1) in schools, and as part of the benchmarking and training of classroom teachers, 2) at the end of secondary school and 3) in universities.

4. Bernard Spolsky *Bar-Ilan University*

Language Testing Reform in Israel

The writing of this paper is in progress as I receive almost daily battles on the struggle with the Ministry of Education over a reform of the whole Bagrut (school leaving certificate) examination system. The paper I had hoped to be able to give would have reminded you of studies by Elana Shohamy and others of changes in the Bagrut examination in order to encourage greater attention to the teaching of oral English. I would have then described how a few years ago, the English Advisory Committee agreed on a set of proficiency guidelines that became the basis for changes in the form of the existing English Bagrut examinations.

From that, I would have gone on to describe a major effort that culminated about a year ago in the development of a new curriculum for English, set out as a set of standards with benchmarks. A committee was about to start work on specifications for a new set of national examinations based on these standards when the Ministry came out with its own plans to change the whole system. The Bagrut system is complex and powerful; in each subject, there are a number of examinations worth different numbers of points towards the certificate. English for example has four examinations, a four- and a five-point level examination (either of which is sufficient for university admission), and a one- and three-point examination at lower levels. None of the national examinations is calibrated, and all depend on casual monitoring. There is of course no pre-testing of items or item analysis.

The new examination plan started out as part of a major reform of high school education. The initial goal was to reduce the number of subjects in which examinations were given, but this was successfully blocked by teachers who felt that pupils would not take a subject seriously unless there were an examination in it. The next proposal was to offer only one examination in each subject, but the examination was to contain three color coded sections, each at a different level. All pupils were to take all levels, at once if they wished or at three separate sessions. There was no way to fit the existing four levels of English, or the needed oral and aural tests at each level into this pattern, so the English Committee has been locked in a struggle with the Minister and senior officials of the Ministry for several months. By the time of the session, I hope to be able to report the resolution of this struggle, which has shown the complex interplay of political and bureaucratic concerns with attempts as educational reform.

5. Sun Yurong *National Educational Examinations Authority, China*
Michael Milanovic and Lynda Taylor *University of Cambridge Local Examinations Syndicate, UK*

Setting up a dynamic language testing system in national language test reform: the Public English Test System (PETS) in China

The government of the People's Republic of China has always seen proficiency in communicative English as essential to the successful implementation of its Open Door policy to encourage rapid modernisation.

The Public English Test System (PETS) development project was established in January 1997 in response to growing concern within China over the inadequacy of standards of communication in English. The PETS project aims to rationalise publicly available English language tests within a 5-level framework ranging from the level of English required at Junior High School to the level required by Chinese graduates in order to study/work overseas. This presentation will outline aspects of the work accomplished during the three phases of the project, including work related to the development of level criteria, test specifications and sample materials, the training of test writers, and the production of test materials.

This development project is of particular interest because it represents a major attempt on the part of a previously traditionally-oriented national testing system to adopt a more criterion-referenced approach which takes into account the impact that the tests are likely to have not only in the immediate pedagogical context but also in wider Chinese society. The project underlines the importance of test developers consulting with many different stakeholders (students, teachers, administrators, funding agencies, the academic community, etc.) throughout the test development process in order to maximise positive impact and minimise any negative effect. In addition, this project has provided a valuable opportunity to apply and validate a cyclical and iterative model of test development proposed in recent years in which careful consideration is given to the many different features of the test development context.

6. Antony Kunnan *California State University, Los Angeles*
Beryl Meiron *Cambridge Examinations and IELTS International,*
Yutaka Kawamoto *California State University, Los Angeles*

A Training Model for Egyptian Language Testers

Although there may exist an educational testing policy in Egypt for the school system as practiced by the Egyptian National Centre for Educational Examinations and Evaluation, there was until recently no perceivable coherent national language testing policy for English and Arabic.

In the early 1990s with financial assistance from the US Agency for International Development (USAID), the Integrated English Language Program-I (IELP-I) administered by Fulbright was begun with the intention of improving English language education in Egypt. Although some gains were made in area of teaching and curriculum development, no attempt was made at developing an overall multi-faceted national language testing policy. In 1998, with the new IELP-II program also funded by USAID and administered by the Academy for Educational Development (AED) and its partner institutions, a coherent long-term program of English language test reform in Egypt has been planned.

The AED Testing Unit has identified four long-term goals for test reform. They are 1) develop an implement a systematic, sustainable mechanism to improve inter-organizational communication and workplace cooperation; 2) develop quality tests, using item banking practices, for the assessment of Egyptian English language professionals (teachers and professors) English proficiency; 3) design and implement a systematic, trackable, sustainable training program to meet the needs of Egyptian English language professionals in the area of educational assessment; 4) design and implement an information dissemination campaign on language testing that is nationwide and client specific. It should include all major stakeholders in the testing process: students, parents, decision-makers and the general public. The Testing Unit arrived at these goals after a research study titled "English Language testing reform in Egypt: Documentary evidence of the need for reform" (Hozaiyn and Khalifa, 1998).

In order to achieve these four goals, one of the most critical aspects has to be the development of a core team of language testers who are trained in current test development practices. This presentation will focus on this aspect by describing three different training programs: at California State University, Los Angeles; in Cairo for testers trained at CSULA; and in the Egyptian Governorates where training is being given by those trained at the first two sites. The presentation will conclude by raising theoretical and practical questions relevant to testing reform and the role and efficacy of external trainers like us.

Symposium Two

Saturday 9:30 - 11:30

The rights and responsibilities of testers and test-takers in language testing settings: ethics, policy, practice and research.

Chair: Bernard Spolsky *Bar-Ilan University*

Overview

This symposium brings together a group of leading scholars in the field of language testing for the purpose of considering the rights and responsibilities of testers and test-takers in language testing settings, from the perspectives of ethics, policy, practice and research.

The symposium offers a unified focus on the "proximal processes" (Bronfenbrenner 1995) that link the rights and responsibilities of language testers and test takers. Guided by the work of Messick (1993), the papers relate issues of "systemic" and "consequential validity" to language testing policy and practice. Within the framework of critical language testing (Shohamy 1997), tests are viewed as social actions which are embedded in and shaped by the particular political, cultural and historical contexts within which they occur. The papers separately examine the issues of fairness and social responsibility in specific testing settings and demonstrate the complexity of such issues, given the variable ecosocial systems (Lemke 1997) within which language testing occurs. Considering issues of validity in this way allows for both a clarification of the limits of the responsibility of testers and an affirmation of the rights of test takers, as well as their responsibilities.

Each of the six papers approaches the issue of ethical testing practice from a different perspective: by locating the discussion of language testing theory and practice within the current epistemological debate relating to research in the social sciences (Greeno 1998;); by drawing on empirical research to illustrate both conflict and balance in the roles of testers, teachers, institutions and communities; by linking language testing to overt and covert political agendas; by recognizing the uniqueness of testing settings; by using dialogue as a basis for a two-way communication between testers and test takers; and finally, by arguing that only tests which are deemed to be "fair" by those who are affected by them can claim validity for the inferences drawn from their use.

Each 15-minute paper provides a cogent and critical response to social responsibility in language testing. The papers will be followed by formal commentary by two discussants who are renowned in the field of language testing for their consideration of these issues.

1. Elana Shohany *Tel Aviv University*

Critical Language Testing: Uses and Consequences of Tests, Responsibilities of Testers and Rights of Test-takers

This paper introduces a broader view of language testing, one which focuses on the use of tests and situates language tests within social, political and ethical arenas. Using a framework of critical language testing, the paper defines tests as acts which are not neutral but rather are products and agents of cultural, social, political, educational and ideological agendas that shape the lives of individual participants, including teachers and learners. This approach shifts the focus to consequential, systemic, interpretive and ethical criteria as evidence of the validity of tests by collecting data on test ethics, bias, effect on instruction, responsibility of testers, and rights of test-takers. The paper reports on studies investigating aspects of tests' use with regard to intentions, effects and consequences, and the use of tests as disciplinary tools. Given the findings, the paper offers solutions with regard to other models of assessment, the roles and responsibilities of test developers and the rights o test-takers.

2. Tim McNamara *University of Melbourne*

Policy and Social Considerations in Language Assessment: an Overview of Recent Research

In the past few years, the social and political character of language assessment has begun to be intensively discussed. This corrective to the exclusively psychological and individualistic character of most language testing research was long overdue, especially given the essentially bureaucratic purpose of much language assessment. This paper looks at the impetus for this development in debates on the epistemology of research in the social sciences and their impact on language testing theory, on discussions of the nature of communicative competence, on the growing demands on assessment made from developments in public policy, and the failure of traditional psychometrics to meet the needs of teachers and learners in classrooms. The response of the field of language testing research to these developments is traced, including discussions about ethics and the development of standards and codes of practice, the inclusion of the domain of language testing in the burgeoning area of critical applied linguistics, a increasing awareness of washback and the growth of alternative assessment.

3. Geoffrey Brindley *Macquarie University*

Understanding Assessment Cultures

The inclusion of test impact into the procedures for construct validation has a range of important consequences for the role of language testers. Not only do they now need to be technically competent in matters of test design and analysis, but they also have to be able to advise on the likely social consequences of the tests they develop. In order to assume this role, testers require an understanding of the "assessment cultures" of the different audiences involved in language testing. These include policy makers, bureaucrats, employers, parents, teachers and learners.

Using a number of case studies of tests and assessment procedures which have been developed in Australia in recent years, this paper considers ways in which language testers can help to reconcile the potentially conflicting views of these different audiences on the role and purpose of assessment. Directions for collaborative research and test development between testers and key stakeholders are suggested.

4. Janna Fox *Carleton University*

Test Taker and Rater Responses: A Study in Proximal Processes

Researchers from a range of disciplines (Kirshner & Whitson 1997, Lemke 1997, Faigley 1985; Hamp-Lyons 1994) have argued that in order to draw valid inferences from tests the full context of production should be considered. This necessitates shifting the unit of analysis from isolated factors (e.g. test scores, raters or test takers) to the interrelationship of processes and practices embedded in the particular sociocultural setting, what Bronfenbrenner (1995) refers to as "proximal processes". In the present study, a "grounded theory approach" (Strauss & Corbin 1990) is applied to characterize the patterns of test taker (n = 423) and rater (n =12) responses to a test of writing. Subsequently, these patterns are evaluated in a purposeful sample of individual test takers (n = 8) and raters (n = 4). Of particular interest are responses to the fairness of the test, the topic and focal features of the writing. The use of "proximal processes" as the unit of analysis reveals similarities and differences in patterns of response in test takers and raters and has implications for the analysis of sources of test bias.

5. Cathie Elder *University of Melbourne*

Language testing: Whose values? Whose responsibility?

Following Messick (1993) the paper focuses on the essentially value laden nature of the language testing enterprise. A number of scenarios (based on the experience and/or research activities of the author and her colleagues) are presented to illustrate the way personal or institutional values govern the decision to bring a test into being, the way in which a test construct is characterised, and the analysis, interpretation and impact of test results. Particular attention is given to two studies: the first a study of differential group performance on an end of school foreign language examination and the way these differences are interpreted by university selection officers (Elder 1997); the second, research documenting the varying perceptions of a government-instigated educational testing initiative by different stakeholders (Elder & Lynch 1998). These studies confirm what other writers (e.g. Spolsky 1995) have claimed about the limitations of technical approaches to test validation and suggest the need for n expanded conception of validity which takes into account the need for an expanded conception of validity which takes into account the perspectives and value-systems of both the language tester and other stakeholders (including policy makers) and which incorporates both the immediate and more far reaching consequences of test use. It is argued, however, that by broadening the scope of validity in this way, we are inevitably multiplying the kinds and sources of the evidence to be drawn on for test validation purposes, thereby placing what may be an overwhelming burden of responsibility on the language tester.

6. David Nevo *Tel Aviv University*

Tester and test-taker: A two-way channel of communication

This paper will discuss the importance of two-way communication as a basis for a dialogue between testers and test-takers. Using the concept of dialogue evaluation' developed in the context of program evaluation and school-based-evaluation, it will delineate the principles that have to be followed by testers and test-takers to facilitate a dialogue in their mutual interaction.

On the basis of extensive experience in working with students, teachers, testers and administrators, within and outside the school, the benefits of dialogue evaluation will be discussed with an attempt to apply them to the context of language testing. The major claim of this paper is that a dialogue between testers and test-takers can enhance our understanding of the problem for which a test is used and increase the probability that the results of that test will be used in a constructive way to solve the problem. But for this to happen, certain principles have to be followed by testers as well as by test-takers.

7. Alan Davies *University of Edinburgh*

Professionalism in Language Testing: Do Codes of Ethics Help?

Ethics has a clear role in institutional settings where there is concern to declare and to limit institutional duties and responsibilities (the 'within reason' clause). This applies particularly to professions and hence to groups (such as those professionally involved in language testing). When a colleague violates the ethics of his/her profession then the profession's membership sanctions come into play. Because sanctions are difficult to implement without legal backing, codes of ethics are developed to provide guidance to those involved in the professional activity and to all stakeholders as to what is promised and what is not. In spite of their lack of legal bite, codes are necessary even if their only sanction is expulsion from membership. We should never undervalue the importance in professional life of being recognised as a member in good standing by one's peers.

Research Reports

Arranged alphabetically by the name of the presenter.

Charles Alderson *Lancaster University/ British Council*, Szabo Gabor *Janus Pannonius University, Pecs, Hungary*, and Richard Percsich *Assessment Centre of BMPI (Pedagogical Institute), Pecs, Hungary* (Friday 15:30 - 16:00)

Sequencing as an Item Type

In the search for more 'communciative' test methods, some public examinations in Europe have experimented with sequencing, or (re)organisation as an item type for the assessment of reading ability. This test method involves presenting candidates with a number of jumbled phrases, sentences, paragraphs or other text chunks, and requiring them to "put them in the correct order". The task may relate to another text which has to be read before the ordering can take place, or may be self-contained.

Although superficially attractive since they seem to offer the possibility of testing the ability to detect cohesion, overall text organisation or complex grammar, such tasks can be very difficult to construct satisfactorily. Although an original text has only one order, alternative orderings may prove to be acceptable, simply because the author has not contemplated other orders and has not structured the text to make only one order possible. Thus test constructors may be obliged either to accept unexpected orderings, or to rewrite the text in order to make only one order possible.

In this paper, we report on empirical research conducted into this item type as part of the Hungarian Examination Reform. We focus on the development of a number of competing algorithms for scoring which allow partial credit for incomplete orderings, and for correct combinations of elements. Discussion will focus on the reliability of human scoring using such algorithms, practicality and validity issues, and we conclude with a set of recommendations for the development of useful scoring procedures.

Jane Andrews and Richard Fay *The University of Manchester* (Thursday 16:00 - 16:30)

Interculturality and English Language Paired Orals: Exploring Assessors' Perceptions

This research focuses on assessors for English language paired oral interviews (eg Weir 1993). It considers assessors as a potential source of bias through their implementation of rating scales concerned with interaction. To do this, it conceptualises the oral exam as an emergent culture (Holliday 1997) in which interpersonal communication across cultures (Gudykunst, Ting Toomey & Nishida, eds 1996) takes place. The research investigates the current community of practice in preparation for the development of assessor training material addressing this potential for bias.

The research project involves five stages: (1) an initial grounding of the research through interviews with experienced assessors; (2) the use of a think-aloud approach (Eriksson & Simon 1984) in which assessors reflect on their operationalisation of the rating scales with regard to candidate performances accessed through standardisation videos; (3) an analysis of the think-aloud data; (4) the development of a training approach to enhance the effectiveness of standardisation with regard to intercultural and interpersonal issues; (5) the trialling of the training materials. Potential outcomes of the five stage project include an increased understanding of assessor variation in implementing rating scales, and an indication of the likely effectiveness of intercultural training within standardisation processes. Both outcomes contribute to the exploration of language assessment in a borderless world.

Jared Bernstein *Ordinate Corporation*, Maxine Lipson *The University of Bologna*, Gene B. Halleck *Oklahoma State University*, and Jane Martinez-Scholze *Defence Language Institute Lackland AFB Texas* (Saturday 8:45 – 9:15)

Comparison of Oral Proficiency Interviews and Automatic Testing of Spoken Language Facility

Background
Oral proficiency interviews and related procedures typically combine linguistic, cognitive and social aspects of performance in their scoring, and they serve as operational definitions of oral proficiency. A recent automatic spoken language test (PhonePass SET-10) scores performance on speaking and listening skills as an indirect measure of facility in spoken language.

Research Goal
This research seeks to explicate the common and divergent aspects of the two testing methods. An analysis of the materials, procedures, scoring logic, and outcomes of the two test types as administered at several sites leads to a clearer understanding of the two methods.

Design and Method
We report data from four related experiments in which non-native candidates took both a human-scored test and an automatic test; 151 candidates at U. Bologna, 92 at Oklahoma State U., 51 at DLI-Lackland, and 127 at Iowa State U. The interview and scoring procedures were somewhat different in each case, but in all cases the candidates took the two tests in close temporal proximity.

Results
Analysis confirms that OPI-like measures reflect a wider set of cognitive, social, and linguistic skills, while the PhonePass test measures ease and immediacy of comprehension and production. Correlations of scores from the two test types for the populations at the four sites range from 0.68 to 0.77. Reliability, cross classification and related performance measures will be analyzed.

Implications
Depending on the intended use of the human-scored test, an automatic test can be used more or less successfully as a pre-administration screen or used to approximate classification outcomes of the human-scored test.

J. D. Brown, Thom Hudson and John Norris *The University of Hawaii at Manoa*
(Thursday 14:15 – 14:45)

Validation of Task-dependent and Task-independent Ratings of Performance Assessment

Performance tasks were developed based on 64 possible combinations of plus or minus values for linguistic code, cognitive complexity, and communicative demand (sources of task difficulty as discussed in Norris, Brown, Hudson, & Yoshioka, 1998). In addition, rating scales were created based on task-dependent and task-independent categories. The criteria for the task-dependent categories were created in consultation with advanced language learners and language teachers for each individual task. These criteria for success were allowed to differ from task to task depending on the input of our consultants. The task-independent categories were created on the basis of three theoretically motivated components of inherent task difficulty as follows: code (linguistic) accuracy, cognitive adequacy, and communicative appropriacy. Complete performance data were gathered from 90 ESL/EFL students at a wide range of proficiency levels in the United States and Japan. Their performances on thirteen tasks (complex and integrated-skills) were scored by three trained raters using the task-dependent and task-independent criteria. The results were analyzed using descriptive statistics, item difficulty estimates, reliability estimates (interrater, Cronbach alpha, etc.), and dependability estimates (using generalizability theory). The results are interpreted and discussed in terms of: (a) the adequacy of our original task difficulty estimates, (b) similarities and differences between task-dependent and task-independent ratings, (c) test reliability and ways to improve the consistency of performance measurement, (d) test validity and the relationship of our task-based test to the stakeholders and curriculum, and (e) the effectiveness of performance measurement for decision making in educational contexts.

Gary Buck, Kikumi Tatsuoka and Irene Kostin *Educational Testing Service* (Thursday 13:15 – 13:45)

Developing and Cross-Validating a Set of Cognitive and Linguistic Attributes Applicable to Multiple Test Forms on a Multiple-Choice Listening Test

Recently the rule-space procedure has been used to analyze a number of language tests, both first language and second language. This procedure produces a set of cognitive and linguistic attributes which purport to underlie performance on the test analyzed. As the procedure is both complex and new, the question of the generalizability of the results is of paramount importance. Basically, if the set of attributes do not apply to other forms of the test, they are of little use.

We propose to address this issue by attempting to develop one set of attributes which can account for performance on a number of forms of the same test. The procedure is to identify 6 forms of the listening section of the Test of English for International Communication (a rule-space analysis of this test was presented at LTRC 1996). We will use 5 forms of the test to develop a set of generalizable attributes, and then cross-validate the results by applying the set of attributes to a sixth form of the test.

Results indicate that it is possible to develop a set of generalizable attributes, with reasonably good 'fit' to the data. The procedure will be explained, the attributes presented and evaluated in terms of how well they operationalize the listening construct, and implications will be discussed.

Catherine Burrows *NSW Adult Migrant English Service* (Thursday 15:30 – 16:00)

Adopters, Adapters, and Resisters: Did the Assessment of the Certificates in Spoken and Written English Change Teaching in the AMEP?

This research was designed to examine teaching practices in the AMEP to search for evidence of the washback effect resulting from the implementation of a new classroom-based assessment system into Australian adult migrant education. A combination of qualitative and quantitative research methods was used, including structured classroom observations. After analysis of the results, patterns were found in teachers' responses, with teachers behaving as adopters, adapters and resisters. By combining concepts from curriculum implementation research and washback research, a new model for conceptualising washback is proposed.

Batia Laufer *University of Haifa* and Paul Nation *Victoria University of Wellington*
(Friday 10:15 – 10:45)

Vocabulary Recognition Speed Test (VORST)

Vocabulary size and depth tests measure word comprehension, production, or both. They do not test the speed of word recognition, or recall. And yet, successful vocabulary use depends not only upon word knowledge, but also upon quick access to it. Vocabulary Recognition Speed test attempts to measure the speed of recognizing word meanings. It is a computerized version of Nation's (1983) Levels Test with a time measuring device. A result file provides information about separate items (correctness and speed of response) and about each word frequency level (number of correct items, total response time, average response time per correct answer).

The paper examines the relationship between speed of access and learners' vocabulary size; speed of access and word frequency level. 488 learners and native speakers were tested, divided into 4 groups of different vocabulary size and compared on response times (by ANOVAs). Response time was also correlated with vocabulary size. Within-subject comparison was conducted on response times to words of different frequencies (by Repeated Measures). Preliminary results suggest that speed of access is moderately related to vocabulary size and to word frequency. Non-native speakers' increase in speed 'lags behind' increase in vocabulary size. Results of native speakers are more homogeneous across subjects and across vocabulary frequencies. It is argued that speed of access cannot fully be predicted from vocabulary knowledge and therefore speed tests should supplement tests of vocabulary size and depth.

Yong-Won Lee *Educational Testing Service* (Thursday 14:45 - 15:15)

Examining Passage-Related Local Dependence (LD) Using Q3 Statistics in an EFL Reading Comprehension Test

In reading comprehension tests, it is quite common that a set of related items is followed by a reading passage. A testlet approach involves treating a cluster of related items combined with a passage as a single testlet rather than treating each item as an independent unit, for it is suspected that such a cluster of items is very likely to be locally-dependent. However, local dependence is not necessarily a matter of a priori assumption, but more of empirical phenomenon.

In such contexts, Yen's (1984) Q3 statistics is a useful tool for empirically examining local dependence in a test. It is a pairwise index of local item dependence and basically an interitem correlation of residuals between the raw scores and the theta-predicted scores from the 3-parameter logistic IRT model. If two items are to be locally-independent, there should be no correlation between the residuals of the two items after the impact of theta is partialled out from the raw scores.

This paper reports the results of the Q3 statistics analysis of a 40-item EFL reading comprehension test administered to 1857 Korean high school students. Average interitem Q3 within and between passages were computed and compared to the grand mean of Q3 among all item pairs. A sizeable amount of local dependence was found to exist within passages, which provides support for using a testlet approach to the analysis of passage-based item sets in the reading comprehension test used in this study.

Laura MacGregor *Sophia University* (Thursday 16:30 - 17:00)

Demystifying the STEP Interview Test

Since its inception in 1964, STEP (The Society for Testing English Proficiency), the organization responsible for producing the STEP test, has operated largely in secrecy. The test itself is no secret: nearly three million take the STEP test each year (2,858,699 in 1997). However, information about test development and evaluation criteria is not readily available. The STEP test's endorsement by the Japan Ministry of Education (*Monbusho*) has provided protection from public scrutiny

Furthermore, apart from the monthly STEP newsletter and annual research bulletin (both written in Japanese), there is almost no opportunity for the people involved--test-makers, test-givers, test-takers, and teachers--to interact. This is of particular importance to the second-stage STEP test, in which the examinee is evaluated in a private interview.

Recalling that STEP operates in a secret world, the need for feedback from examiners and examinees on their knowledge and impressions about the test is all the more urgent.
This paper addresses this need for information from examiners and examinees. It reports the results of questionnaires and interviews conducted among a group of examiners and examinees who participated in the STEP interview tests in July, 1998. It explores three areas:
1) test preparation
2) test contents (the actual test items)
4) test evaluation
Feedback from examiners and examinees are summarized and a set of recommendations to STEP is presented.

Junko Matsui *Meikai University* (Thursday 17:00 – 17:30)

Various Interrelated Factors Involved in Listening

Various interrelated factors are involved in listening, and a deficiency in any pertinent perception skill can be potentially fatal to comprehension. The present study attempts to demonstrate that it is possible to use statistical information to evaluate the listening skills of Japanese subjects listening to English, and determine what effect set factors have on the listening process. For instance, it is possible to test the relevance of vocabulary versus background information for intermediate and lower intermediate level subjects.

Results of a test correlating TOEFL-ITP scores with improvements in listening skills indicate that vocabulary is more essential for word comprehension than background information, but both vocabulary and background information are necessary to grasp the meaning of an utterance.

The higher a subject's proficiency in listening, the more likely either vocabulary or background would benefit the listener for the dictation exercise. Previous research has revealed that listening is rate dependent - that is the location of phonetic category boundaries changes with rate. (Miller, O'Rourke, & Volatis, 1997). Lower achievers may have not acquired such rate dependent listening skills. Also, Japanese listeners have been found to classify English vowels as similar Japanese correlates (Ingram, J. & Park, S., 1997). Proficient listeners are more likely to have established separate categories for separate vowels, leading to fewer listening errors, whereas less proficient listeners re-assign all foreign sounds into their native phonetic system, thereby increasing the chance of misperceptions. Results show that students who are better at listening have already acquired more of the basic skills such as reduction, linking, assimilation, and consonant/vowel perception needed for comprehension, and therefore any extra information provided leads to instant improvement. Meanwhile, listeners who have not acquired basic skills do not benefit as much from additional vocabulary and background knowledge because their lack of listening skills hinders the phonetic input process.

Likewise, the higher a subject's vocabulary, grammar, and composition proficiency was, the more likely he/she was to initially comprehend the meaning of the target listening segment. However, further information such as additional background was more beneficial to listeners in comprehending meaning, for subjects who have higher TOEFL listening scores. In other words, listening skills were essential in maximizing other pertinent input, such as vocabulary and background information. Unless a subject had the basic listening skills needed, he/she was unable to take full advantage of any other helpful data provided.

The present experiment is one example of how it is possible for educators and researchers to obtain meaningful data through statistical means, such as correlating various listening skills with TOEFL-ITP scores, thereby leading to potential improvements in pedagogical methods.

John Read *Victoria University of Wellington* (Saturday 8:15 - 8:45)

The Impact of English Tests for Migrants

In the last decade, as English speaking countries have sought to attract successful well-educated migrants, language tests have been increasingly used to determine whether these people have adequate proficiency in English to practise their profession or conduct their business. This has raised issues for language testers concerning the validity of the tests and their role as gatekeeping devices within the overall immigration policy of the country.

This report will analyse the particular case of immigration to New Zealand. Since 1995 targeted migrants have been required to obtain a minimum band score on the the General Module of the International English Language Testing System (IELTS) test. The most controversial element of the new policy was a $20,000 bond, intended as an incentive for migrants with limited English to become proficient in the language as soon as possible after arrival. The policy has been widely regarded as discriminating against migrants from non-English speaking backgrounds and it contributed to a dramatic decline in migration applicants, especially from East Asian countries.

The paper will review the rationale for the English language requirement and consider the appropriateness of using the IELTS test for this purpose, for which it was not originally designed. An analysis in terms of current test validity theory will highlight the point that, in a case like this, evidence of the technical quality of a test must be evaluated in conjunction with the social impact of its use in a high-stakes decision-making context.

Hisami Saito-Scott *Mission College, Santa Clara, CA* (Thursday 13:45 - 14:15)

Analyzing the Dimensionality of a Second Language Reading Test from Cognitive Perspectives

The unidimensional Item Response Theory (IRT) model assumes that all the items in a test are measuring one underlying ability (Lord, 1980). However applied linguistics studies raise a question about the adequacy of unidimensional IRT models for analyzing language tests because most language tests measure more than one ability (Buck, 1996).

This study analyzed the reading section of the Test of English as a Foreign Language (TOEFL) using the Rule Space Methodology (RSM) to investigate the condition under which unidimensional IRT can be used for psychologically multidimensional tests.

First, the RSM was applied to the test data in order to capture examinees' cognitive strategies. Using the cognitive strategies identified, various simulated data were analyzed to examine the dependence of dimensional structure with respect to the increase of different types of cognitive paths. Second, psychometric dimensionality of the data sets was analyzed using the full-information factor analysis procedure, Stout's non-parametric procedures, and the Multidimensional IRT Model.

The results suggested: (1) 24 psychological dimensions are involved in the reading test; (2) although the test is psychologically multidimensional, the test can be psychometrically unidimensional if students' solution strategies were interpreted in terms of one cognitive path; and (3) if students' response patterns were categorized into multiple cognitive paths in the rule space, the data exhibit psychometric multidimensionality.

These results will contribute to a better understanding of the relationship between psychometric and psychological dimensionality and the effect of examinees' cognitive strategies on the test validity.

Tetsuhito Shizuka *Kansai University* (Thursday 17:30 – 18:00)

Using Test-takers's Confidence Levels in Scoring Multiple-choice Tests: Rationale, Empirical Findings, and Simulation Results

When a test is scored dichotomously, one thing not reflected in the resultant data matrix is that different degrees of person ability can be involved in producing the same "1" or "0" response; one may respond correctly because one "knows", "thinks" or "guesses" it the correct option. It seems reasonable, then, to presume that getting an item right with high confidence is indicative of greater ability than doing so with low confidence; conversely, getting an item wrong with high confidence may be a sign of lesser ability than getting the same item wrong, but sensing that it may be wrong. This paper proposes a scoring system that represents this supposition and reports on preliminary findings concerning the effects of combining response correctness and confidence level ratings.

EFL Japanese college students taking two multiple-choice reading tests (one paper-and-pencil and the other computer-administered) were instructed to select one of three confidence level ratings for each option chosen. Combining the 1/0 response correctness data and 3/2/1 confidence level rating, 6-point-scale polychotomous data were produced in such a way that a correct response with higher confidence was credited more highly than one with lower confidence.

Findings revealed that the incorporation of confidence level ratings generally improved the overall reliability of the items. Simulation data sets produced by different hypothetical test-takers ("risk-takers" and "risk-avoiders") are also presented as a basis for discussion of possible threats to validity of the resultant scores.

Student Research Reports

Thursday 8:15 – 9:45

Constructing Rating Scales for EFL Writing Tests

Yoshihiro Sugita *Yamanashi University*

In the past decade there has been a major shift in language testing towards the development and use of communicative tests. There is a need for the development of a greater variety of assessment procedures that will elicit discourse, but, more importantly, there is a need to devise rating scales and criterion that will focus specifically on aspects of discourse in the analysis of language samples.

This study examined how we developed assessment procedures that would elicit discourse. Standard rating scales require the matching of examinee performance to a verbal description. Thus, in spite of the fact that language is assessed in discourse, there is no specific focus on discourse features. Criteria for evaluating writing more relevant to the naturalness of language use are needed.

An analytical test of organizational ability was developed in order to construct a rating scale. The purpose of the test is to discriminate briefly what type of rating scales are suitable for assessing students' writing proficiency in a particular classroom. The test consists of anagram solving, word reordering, sentence reordering, and paragraph assembly. The results were discussed in terms of strategic discourse processing. Thus, a feasible rating scale to assess students' use of language is described, and its reliability and validity are examined. Reasons are offered to explain why this type of scale may improve reliability and validity problems associated with standard rating scales.

Validation of Diagnostic ESL Test for Measuring Lecture Comprehension Ability

Dongil Shin *University of Illinois at Urbana-Champaign*

The purpose of this study is to construct and validate a diagnostic lecture comprehension test. The combination of input ranges (wide and limited) and process types (global and local) is theoretically modeled to define lecture comprehension tasks. The model is operationalized to diagnose students' weaknesses and strengths in a lecture comprehension test. An argumentative model by which to judge construct validity evidence is proposed. Construct validity evidence is mainly accumulated to determine the extent to which the diagnostic test represents theorized task types.

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. All evidence types are to support the interpretation of test construct. The first step in the construct validation is to examine content evidence in terms of spec-fit-to-item. When the test specification and items are well linked, the test is administered, and performance data is acquired. The second step is the analysis of predicted and observed item difficulties. The third and fourth steps are to find validity evidence pertaining to correlational study and to internal structure. Correlational study is embedded by multi-trait multi-method and confirmatory factor analysis. Internal structure is examined in terms of dimensionality.

This study will be applied to actual diagnostic assessment from an university ESL service program in the spring semester of 1999. This study offers a principled approach to theory, test construction, and validation study in the ESL listening.

Using Structural Equation Modeling to Validate a Rating Scale

Amy Yamashiro *Nihon University*

Speech communication skills, essential for academic and professional success, require careful planning and research for instruction in EFL. The rating scale developed as the criterion-referenced assessment of the EFL public 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. This paper uses structural equation modeling to analyze the performance assessments to determine the extent to which teacher, peer, and self-assessments may be reliable and valid. A review of the existing literature on performance testing alongside peer and self assessment within second language acquisition indicates need for further research on rating behavior within classroom-based inquiry.

In this study, the data will be collected using the rating scale as part of EFL public speaking courses (N=120, minimum) in Japan. Following the in-class data collection, three trained EFL teachers will rate each speech to determine the inter-rater reliability and generalizability among the teacher ratings. Cronbach alpha, inter-rater reliability, and GENOVA will be calculated. Factor analysis will be conducted to confirm the three traits on the rating scale. Structural equation modeling using EQS 5.0 for the Macintosh will be used to perform the multitrait-multimethod analysis to determine the validity of the three traits and the three rating methods. The presenter will include a discussion of the study's limitations and implications.

Poster Session One

Friday July 30
Presentation Session 8:15 to 9:15 (Second Floor Convention Hall 200)
Discussion Session 9:45 to 10:00 (First Floor Conference Room 102)

Using a spoken language corpus in the development of EAP listening tests

Sarah Briggs and Barbara Dobson *University of Michigan*

Corpora have been used for lexicography and pedagogy, but as Alderson (1996) observes, not yet in language testing. Alderson, however, encourages us to explore applications, and this poster shows how we have begun to explore uses of a new corpus of academic spoken English (MICASE) in our development of EAP listening tests.

A video-delivered test was developed to assess the listening skill of incoming university students. The extended discourse portions of the test were not scripted. Speakers generated the language for the test prompts on the basis of role assignments and notes about general content information they were to convey.

We compared the unscripted language of the test prompts with language segments from the corpus. We analyzed various linguistic and discourse features of the test prompts and of the corpus segments, revealing shared characteristics and salient differences. This analysis using the corpus helped us focus on both the process and the product of the test development project.

Rating the VOCI: Alternative methods

Gene B. Halleck *Oklahoma State University* and Daniel J. Reed *The University of Minnesota*

As the assessment of oral language proficiency has become more widespread, researchers have investigated issues related to rater variability and rater training (Lumley & McNamara 1995; McNamara & Lumley 1997; Moder & Halleck 1997). Halleck and Reed (1998) found that raters could be trained to make determinations of task fulfillment and intelligibility for responses to a Video Oral Communication Instrument (VOCI) without going through the extensive training required to make holistic judgments based on criteria put forth in the ACTFL Guidelines. For this study two ACTFL-certified testers used three different rating procedures to rate 100 ESL VOCIs: 1) holistic ratings that were determined by listening to audio-taped responses and applying the ACTFL Gudielines; 2) prompt-by-prompt ratings determined by listening to the audio-taped responses; and 3) holistic ratings by a second ACTFL-certified tester who was shown the prompt-by-prompt ratings and determined probable holistic ratings based on these more discrete ratings without listening to the audiotaped responses.

Our findings shed light on the relationship of task fulfillment and intelligibility on holistic ratings and reveal interesting patterns of task variability. We look specifically at those VOCIs that were difficult for the second rater to estimate, and discuss the reasons the two ratings (using different methods) did not agree. Examples of the video prompts and the audio-taped responses will be illustrated on the poster.

Taking a high stakes test--exploring its meaning(s) for an individual

Ari Huhta, Anne Pitkanen-Huhta, and Paula Kalaja *University of Jyvaskyla*

This poster reports on a pilot study on testees' experiences of taking a high stakes test. This qualitative study aims at finding out the role and meaning(s) of a school-leaving examination, and especially its test of English, including various forms of washback effects that it might have on language learning.

The pilot study focussed on a 17-year-old female student taking the English language test as par of the school-leaving examination in Finland. She was given a set of topics to talk about freely on tape in a diary form. She kept the tape diary for a few weeks before and after the test. The topics covered her opinions about the test, its importance for her and its effects in the classroom and on her activities outside school. Immediately before the test she was asked to tell about her feelings about the coming test and about her immediate preparation for the test. After the test the prompts encouraged her to reflect on her views about the various parts of the test and to estimate her success in it.

An in-depth description of individual students' experiences will provide a deeper understanding of and a fresh perspective on the role and effects of the final examination and the test of English in this particular educational context.

In the next phase of the study the research procedure will be revised in the light of the pilot study, and the main study will be carried out with more informants at various levels of proficiency in English.

Measurement and Evaluation of English Communicative Competence--Discrete-point vs. Integrative Tests and Integrative vs. Analytical Evaluation

Akihiko Mochizuki *The University of Tsukuba* and Noboru Yamada *Shizuoka Sangyo University*

Background. Since the Ministry of Education in Japan issued the course of study with a great emphasis on the development of communicative competence in 1989, Communicative Language Teaching has been getting popular in Japanese junior high and senior high schools. Lessons at those schools especially at junior high schools have put much stress on the development of communicative competence. However, what has loomed big as a problem in English education in Japan is how to evaluate communicative competence.

Research purpose. This 3-year research project starting in 1996 which is funded by the Ministry of Education seeks validation for Communicative Tests -- listening, speaking, reading, and writing tests---for Japanese junior & senior high school students which are constructed on the basis of Weir's (1993) three-stage framework theory. It aims to investigate, first, which of the two methods, integrative or analytical evaluation, is an accurate measure of communicative competence, second, inter-rater reliability and intra-rater reliability of Communicative Tests, and third, the relationship between discrete-point tests, STEP English test used as a criterion-referenced test, on one hand, and integrative tests such as Dictation, Standard Cloze Test and 4 Communicative Tests on the other hand.

Method. Data are collected from about 480 Japanese junior high school students and about 300 Japanese senior high school students from the central region of Japan. A discrete-point test, STEP test, used as a criterion-referenced test, and 6 integrative tests---Dictation, Cloze Test and 4 Communicative Tests (Listening, Speaking, Reading, and Writing)---are administered at 3 public junior high schools and 4 public senior high schools. Testing time of all the tests is 50 minutes, the same length of one lesson period at Japanese schools. Four Communicative Tests are constructed by adapting Weir's (1993) three-stage framework theory for the use of these tests at Japanese junior and senior high schools. Data analyses aim at estimating validity for integrative vs. analytical evaluation for 4 Communicative Tests, and discrete-point vs. integrative tests. Correlations between integrative and analytical evaluation and between discrete-point and integrative testing over about 480 junior high school students and 300 senior high school students will be reported.

Results. Analyses are now being made.

Implications. A true picture of valid and reliable Communicative Tests (listening, speaking, reading, writing) will be made clear, and the comparison of integrative and discrete-point testing and also integrative and analytical evaluation will provide teachers of EFL with a clue to which type of tests should be administered and which type of evaluation should be conducted depending on the situation.

A prototype of an item pool for Computer-Based, multimedia listening test.

Youichi Nakamura *Nagano-ken Shinonoi High School*

In our era of Computer Based Test (CBT), much focus should be given to construction of item pools, which is a basic component of CBT.

This poster session presents a prototype of an item pool that consists of listening test items, containing multimedia data, such as text, sound, picture and movie files.

Recorded voices of an assistant language teacher (ALT), home made video taped materials, movie files taken from the internet web site, picture files from a CD-ROM on the market and video taped materials released into the market, were used as the materials for the items of the listening test.

The items were given to Japanese senior high school first graders. The response data of the examinees were analyzed, and the item analysis within Classical Test Theory was done. Invariant item difficulty parameters were also estimated using PROX procedure, within Item Response Theory, applying 1 parameter Rasch model.

A prototype of an item pool was constructed, being composed of the digitized sound and movie files, the scripts, the questions and the choices, the testing points, the data of the item analysis and the item difficulty parameters.

Being aimed at the easily accessible system in an ordinary school situation, "HyperCard" is utilized as a database application software, which is attached to all Macintosh computers and whose operation is very easy and straightforward.

IRT Analyses of the Japanese Language Proficiency Test

Hiroyuki Noguchi *Nagoya University*

Since 1984, our group has annually constructed IRT latent trait scales for the Japanese Language Proficiency Test. Three scales have been prepared each year for each subtest: "Writing and Vocabulary," "Listening Comprehension," and "Reading and Grammar."

Procedure:

1. Testing unidimensionality through factor analysis of inter-item tetracholic correlation matrix and scree test;
2. Estimating item parameters through the heuristic method (before 1996) and marginal maximum likelihood method (from 1997);
3. Evaluating test information by plotting test information curve and observing its maximum point as well as the range of scale values above the amount of criterion information;
4. Examining the distribution of subjects' estimated scale values in each subtest and correlation among them.

Findings:

1. All three subtests at each level show a satisfactory degree of unidimensionality with "Listening Comprehension" exhibiting a slightly higher trend;
2. Almost all items in each subtest show relatively higher discrimination power and can be classified at the easy and/or medium difficulty levels;
3. The amount of test information contained in "Listening Comprehension" is smaller than that of "Writing and Vocabulary" and "Reading and Grammar";
4. The distribution of subjects' estimated scale values in each subtest is unimodal, symmetry or slightly skewed to the positive side;
5. "Writing and Vocabulary" and "Reading and Grammar" show relatively higher correlation than other combinations.

Future tasks:

1. Applicability of IRT models other than 2-parameter logistic model;
2. Determining an appropriate equating procedure for each year's IRT scale;
3. Detection of Differential Item Functioning.

Verifying the Validity of the Japanese Language Proficiency Test Using Can-do Statements

Reiko Saegusa *Hitotsubashi University,* Mako Aoyama *The Japan Foundation,* Tsutomu Fukuchi *Chuo University,* Sukero Ito *Tokyo University of Foreign Studies,* Mikio Kawarazaki *Tokai University,* Keiichi Koide *Gumma Prefectural Women's University,* Takashi Murakami and Hiroyuki Noguchi *Nagoya University,* Kazuo Ohtsubo *Reitaku University,* and Akiko Wada *The Japan Foundation*

It is not trivial to interpret the results of a language test which consists of multiple choice questions. This is especially true when the objective of the test is to assess the language abilities of many examinees with a variety of backgrounds. Can-do-statements, which ask examinees how well they can handle the daily situations with their language skills, have been utilized to assess the actual language ability in the real world.

The present study reports the preliminary analysis of correlation between the Japanese Language Proficiency Test and Can-do-statements. Can-do-statements consisting of 42 questions were prepared for 90 examinees from 16 different countries. All of them had sat for the 1st level Japanese-Language Proficiency Test conducted two weeks before. The alpha coefficient of 0.964 for the total remark was obtained, which is considered very high for a self-rating scale. The correlation coefficient between Japanese-Language Proficiency Test and Can-do-statements was, however, 0.42.

These observations indicate that the current Can-do-statements only partially overlap with Japanese-Language Proficiency Test in respect to what specific language abilities are questioned. The current study suggests that Can-do-statements can potentially be used to assess the validity of Japanese Language Proficiency Test. Some of the items in the statements, however, would not be appropriate for the assessment. Further development on this topic will be reported at the presentation with the most current analysis of correlation between the 1998 Japanese-Language Proficiency Test and the improved version of Can-do-statements.

Testing for Justice

Elaine Wylie *Griffith University*

This poster presents a test which has recently (1998) been developed in Australia to determine whether a person with a language background other than English who has been charged with a criminal offence needs an interpreter to communicate effectively in interviews with legal counsel and in court proceedings. The test is designed to be administered by people such as lawyers and field officers in situations where testing experts are not available (e.g. in remote Aboriginal settlement areas). The six sections focus on different aspects of oral-aural communication in legal contexts, testing these directly through face-to-face interaction. The sections increase in difficulty, and the tester abandons the process when the client fails to complete any section to a given standard, thereby demonstrating the need for an interpreter. The tester is warned to interpret the client's behaviour conservatively throughout the test so that the possibility that he or she will fail to identify the need is minimised. The poster will provide examples of prompts that represent distinct types of discourse featured in the test (e.g. explanations of legal terms and cross-examination) and examples of sociocultural factors that testers need to take account of when they are testing different client groups. The poster will also show how the tasks used in the test are being incorporated into a version of the International Second Language Proficiency Ratings (ISLPR) for Legal Purposes.

Exploring Washback in Japanese EFL Classrooms

Yoshinori Watanabe *International Christian University*

It is normal to claim that the whole Japanese educational system and practice are dominated by the university entrance examinations. "Examination hell" is the phrase that is often used to describe the situation. However, since those claims were made without any empirical support, the argument is likely to be sterile. The present paper attempts to cast some light on this issue by reporting a case study on washback effects of the exams on pre-college level education. The research was conducted in 1994, before the exams which would be based on the new curriculum were to be administered in 1997. Thus, the data are expected to be employed as base-line data, against which the new data would be plotted. The data were gathered from classroom observations of a high school and special preparatory school, and interviews with teachers and students. A major focus was placed on teacher's teaching activities in the classroom, but several findings will be reported about learner's motivation as well. Various factors other than the exams will be identified to help innovate EFL through the exams. Analyses of the new exams will also be reported to predict their washback, which will be investigated in the future longitudinal research.

Poster Session Two

Friday July 30
Presentation Session 16:00 to 17:00 (Second Floor Convention Hall 200)
Discussion Session 17:00 to 17:45 (First Floor Conference Room 102)

Language Testing Courses and Language Teaching Initiatives: A Collaborative Process

Ofra Inbar *Tel-Aviv University*

The school practice component in language teachers training programs is acknowledged as being critical for constructing the knowledge base of language teachers. (Freeman and Johnson, 1998). Thus most language teachers training programs follow the theory and practice model, whereby theoretical principals are implemented in the field. Likewise language testing courses in teacher training programs aim to merge theory and practice, enabling the graduating teacher to experience hands-on the planning and carrying out of assessment procedures in the schools. The implementation of the theory and practice model in language testing courses, however, seldom occurs due to mostly technical and bureaucratic difficulties.

This presentation will describe the practice component of a language testing course, in which pre-service teachers of Arabic, who obtained training in testing, assessed the oral language abilities of elementary schools children in Spoken Arabic in schools. The tools for the oral assessment tasks (interview, picture description and role play situations) were developed and examined for their validity and reliability by the students, as was the rating scale designed to assess the oral performance.

The assessment took place as part of an evaluation aimed at providing formative feedback on an initiative to teach Spoken Arabic in Israeli elementary schools. Thus findings which resulted from the students assessments had significance beyond the teacher training level, as they were used as feedback for formative and summative evaluations of the program, and contributed to improving and promoting the language teaching initiative in terms of curriculum, material and assessment considerations.

This poster presentation will describe the background to the project, the assessment process, the hands-on training component and the potential contribution of language testing courses to evaluating language issues in educational systems.

Testing English Tests: A Concurrent and Internal Construct Validation of the NCUEE-Test in Japan

Akihiro Ito *Aichi Gakuin University*

This study examines whether English questions on Japanese university entrance examinations are reliable and valid measures of examinees' language proficiency. The following tests were administered to 100 college freshmen; a English test from the National Center for University Entrance Examination (NCUEE-Test) with the additional paper-pencil pronunciation tests; a carefully constructed cloze test as a criterion measure. Results indicate the NCUEE-Test is a fairly reliable and somewhat valid testing device to measure students' proficiency. The pronunciation Test in the NCUEE-Test is not a reliable measure of the students' listening ability. This study also reports the results of internal construct validation study (Alderson, Clapham, Wall, 1995) on the NCUEE-Test and the author would like to think what kind of procedures should be done to improve the quality of the NCUEE-Test. This study is the integration and synthesis of the author's research projects since 1996.

The Analysis of English Listening and Reading Comprehension Concerning Japanese and Korean High School Students

Masayoshi Kinoshita *Fukuoka International University*, Hiroshi Shimatani *Kumamoto University*, Terry Laskowski *Kumamoto University* , Hiroki Yamamoto *Seinan Jogakuin Junior College*, and Masashi Takemura *Hokkaido Sapporo Kita High School.* All are members of the Japanese and Korean College Entrance Examination Research Group

A recent comparative study, which surveyed college and university students' English abilities and attitudes toward learning English in China, Korea, and Japan (The JACET Kyushu-Okinawa Chapter Project Committee in 1997), found English proficiency levels of the Japanese university students were lower than those of students in of the other two countries. This result seems to correspond to recent TOEFL scores. Although a comprehensive comparison study has been done at the college and university level, there have been very few studies of this nature done at the high school level (Kinoshita et. al, 1997). In addition, changes in foreign language testing in the East Asian region have increased the need to do comparison studies.

In this presentation we will report on two purposes of our research. One is to survey the differences of the entrance examination for universities in both Korea and Japan, and how these differences might influence both high school students' English performance, their attitude toward English itself, and their interests toward learning English. The other is to explore the results of participants' listening comprehension test to determine whether or not test results are higher in one group because, in Korea, a listening comprehension test has been introduced as part of the qualification test for entrance to universities since 1993.

The test design consists of two parts, listening and reading comprehension. The participants for the test were 300 Japanese and 300 Korean high school students, and the test was administered in their respective countries.

An analysis of the results of the test and implications will be presented at the conference.

Performance of Japanese Examinees on TOEFL Section II

Mikaya Koarai *St. Dominique's Institute*

It is often said that Japanese learners of English are good at grammar, but cannot speak, listen, nor read well. Swinton and Powers (ETS,1980) concluded that Japanese test-takers' weak performance on the Reading Comprehension section of TOEFL was due to their weakness in grammar. Researchers have also discovered that the grammar-focused TOEFL Section II (Structure and Written Expression) shows a very high correlation with Reading Comprehension (Section III).

How do Japanese test-takers, in fact, perform on TOEFL Section II? Specifically, two questions arise:

1) What specific items and grammatical problems do Japanese show poorer performance on, compared with non-Japanese applicants?

2) Do Japanese studying in the United States outperform Japanese in their home country?

This paper answers these two questions by reporting the results of a research project which compared the performances of Japanese examinees in Japan, native Japanese examinees residing in America, and non-Japanese examinees on the Structure and Written Expression sections of the TOEFL test. The research was conducted on 2,500 data samples from the August 1996 TOEFL administration.

Background information, including demographic differences regarding age and gender, and reasons for taking the TOEFL will be shared. The presenter will also discuss specific items that the Japanese group showed poor performance on and suggest reasons why.

References

Swinton, S. S., & Powers, D. E. (1980). Factor analysis of the test of English as a foreign language for several language groups (TOEFL Research Report No. 6). Princeton, NJ: Educational Testing Service.

Verification of a Japanese Vocabulary Test by the Rule Space Methodology

Naoki T. Kuramoto *Tohoku University*, Masahiro Kasai *DePaul University*, and Hisami Saito Scott *Mission College, Santa Clara, CA*

The Rule Space Methodology (RSM) has been widely used as a promising method for providing diagnostic information in various testing areas. Taira, Scott & Kasai applied it to a Japanese vocabulary Test developed by Taira et al. (1993). They obtained a satisfactory classification rate of more than 80%, and proved the favorableness of the method in a new field (Taira, Scott & Kasai, 1998).

The purpose of the present study was to utilize RSM for the verification of the test, by using the Network of Knowledge States (NKS) determined from the major knowledge states identified in the previous studies. Seventeen major knowledge states were identified from the responses of 1,500 examinees, using an incidence matrix including 24 attributes. A NKS was drawn on the two dimensional cognitive space with the axes that represent overall ability levels and the unusualness of the response patterns.

The results indicate that if a student mastered the attribute A, he/she would receive a considerable gain in score. While, even if he/she mastered the attribute R, his/her gain would be a little better than nothing. The attribute 'A: Unusual Word' is regarded as one of the most important factor for the purpose of measurement. On the other hand, the attribute 'R: Same Pronunciation' should be considered as a test-wiseness factor, which works for solving questions but not reliable for the purpose of measurement.

This study revealed that the RSM could be used for verifying the content validity of the scale, as well as for the diagnostic purposes.

The Dictation Test and Its Evaluation for Japanese Speech Using A Portable Computer

Jouji Miwa *Iwate University*

I have developed an interactive system of dictation test for Japanese speech using a portable notebook computer. The system is one of CALL (computer assisted language learning) for Japanese speech.

The contents of the test are sentences, words, syllables, special morae and word accent. Contents for consonants are minimal pairs such as [kenki], [genki], [penki] and [benki]. Special morae are tested using long vowels such as [ozi:san] and choked sounds such as [iQto:].

The dictation scores are automatically stored in a disk file. The system is portable so that students can use it in a lecture room, their home and other places.

In an experiment of a dictation test for two Chinese students, scores for words, consonants, long vowels, choked sound and word accent are 83%, 33%, 72%, 59% and 53%, respectively. Specially, the score in consonants is the lowest. /p/, /t/, /k/, /b/, /d/, /g/ are mutually confused, because Chinese students uttered aspirated and un aspirated sounds.

The programming language for the system is Java so that the system is multi-platform for different OS such as Windows, Macintosh and UNIX. The system is also available on the Web server (http://sp.cis.iwate-u.ac.jp/sp/lesson/j/). I will demonstrate the system in the poster session.

Oral Communication Test: Japanese Speakers of English

Masanori Nakamura and Nunzio Scena *Georgia State University*

Since the Ministry of Education in Japan has begun emphasizing communicative aspects of language teaching, more communicative-oriented teaching approaches have been incorporated into language classrooms. However, there seem to be few assessment tools widely available here to measure learners' communicative skill.

In this project, we present a test to be used as a diagnostic tool to assess the linguistic components affecting the comprehensibility of beginner and intermediate second-language learners in Japanese universities. This test is based on a criterion-referenced construct definition, focusing on the language traits that are essential to comprehensibility and communicative competence: pronunciation, grammar, fluency, and vocabulary.

To ensure that the test is primarily of speaking rather than listening or reading ability, written prompts are given in Japanese. This ensures students' understanding of what they are expected to perform in each section. However, the spoken prompts on the master tape are in English, in order to activate their schemata in English. In addition, much of the content of the items draw from the Japanese experience, to ensure that "lack of something to say" does not influencing performance.

Because the large number of students in the average classroom makes administering oral tests problematic, much of the test is administered through booklets and tape recorders, allowing one person to proctor many students at the same time. The rest is accomplished through direct testing, in which spontaneous speech is elicited through group work with other students, allowing more than one to be evaluated at the same time.

Although the test is low-stakes, it has high ambitions: it attempts to assess not just the level of proficiency in each component, but to pinpoint the specific problems areas, such as the production of individual phonemes (rating sheets enumerate the problems common to many Japanese learners). We hope the test will give students a yardstick against which to measure their achievements and failures, and teachers a tool to guide their classroom activities.

Ten Years Later: The Guam Educators' Test of English Proficiency

Daniel L. Robertson *University of Guam*, and Charles W. Stansfield *Second Language Testing, Inc*

Abstract: This report deals with the Guam Educators' Test of English Proficiency (GETEP), produced by the Center for Applied Linguistics in 1989-90 and implemented in the Guam Department of Education (DOE) in 1990.

The report begins with some background information on the situation in Guam which precipitated the imposition of this language test as a condition of employment with DOE and the other attempts at assessing language skills of teachers which preceded the GETEP.

A brief description of the GETEP and its development will be given. Then, the events of the 1990s with reference to education in Guam will be described. Specifically, several factors related to the language proficiency of new DOE teachers will be examined, including the establishment of the Guam Teacher Corps to increase the number of local teachers and the concomitant decrease in the hiring of off-island contract teachers.

Data obtained from DOE regarding the GETEP results from the beginning will be presented and discussed. This discussion will include information regarding test standards-setting and requirements for retesting. It will also include information regarding review classes set up by the GTC to assist local teachers in passing the test or parts of it after their first testing. Information will be provided through interviews of University of Guam faculty members who taught these review classes.

Finally, the presenters will discuss the relationships among the problems in Guam's schools, the GETEP, and other factors during the past decade. These include test maintenance and rater training/retraining as well as the appropriateness of the use of this test given the present situation in Guam.

A Study of Washback in Brazil

Matilde Scaramucci *State University of Campinas, SP, Brazil*

This paper discusses the results of a study aimed at investigating the washback effect of an EFL reading test (which is part of the entrance examinations held by State University of Campinas for over a decade) in two different secondary school contexts in Campinas, SP, Brazil. The methodology triangulates data from classroom observations, interviews with teachers and analysis of test samples and test program. The results show that a deterministic effect does not hold, as the degree of the impact is different in the three contexts. It is greater on what the teachers teach (content) than on how they teach (approach and methodology). Implications for a theory of washback are discussed.

Converting from a General Purpose Writing Task to a Special Purpose Task

Mary C. Spaan *The University of Michigan*

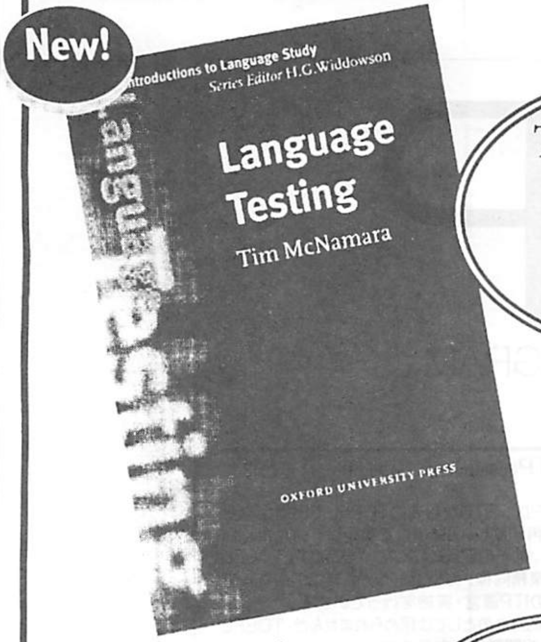
Non-academically oriented ESL students around the world find it beneficial to attain certificates attesting to their general language competence. This is useful to them in obtaining employment, getting job promotions, progressing through the levels in a language school setting, or for personal satisfaction. Similarly, agencies in the private and public sectors welcome language certification from an external source.

This poster briefly describes an international intermediate level English certificate examination, focusing on the development of the writing component. The writing component is a letter or essay written in response to a brief prompt which sets context and provides information to interpret and comment on. At this level, meaningful communication is stressed over linguistic accuracy. In response to requests from examination centers, the test developers decided to create an alternative writing component, one for adults interested in English for Business Purposes. Long-range plans are to give candidates, who may be either adults or young adolescents, the option of either the general or the business writing component. The poster will inform on the development of the business component, and its piloting to approximately 300 Spanish and Arabic speakers who wrote both the general and the business tasks and who completed a questionnaire. Affective variables revealed in questionnaire responses will be compared to writing scores.

Initial scoring of the business writing task has raised some interesting questions about scoring criteria: is there a good fit between the general writing criteria and the business writing responses? Specifically, should content, appropriateness, and register be emphasized more for the business writing?

Oxford University Press

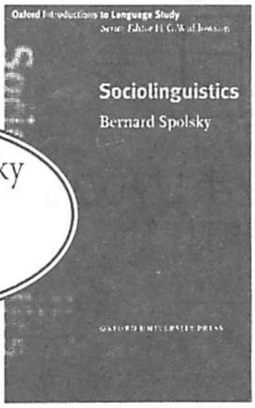
New!



Language Testing

Tim McNamara

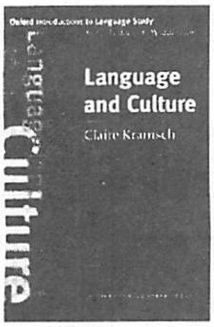
LTRC '99
Plenary
Speaker


Oxford Introductions to Language Study
Series Editor H.G. Widdowson
Sociolinguistics
Bernard Spolsky
OXFORD UNIVERSITY PRESS

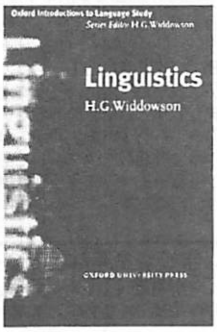
Sociolinguistics

Bernard Spolsky

LTRC '99
Presenter



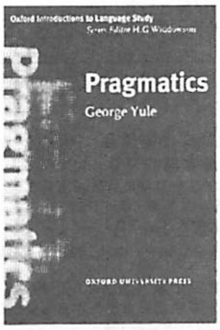
Language and Culture
Claire Kramsch


Oxford Introductions to Language Study
Series Editor H.G. Widdowson
Linguistics
H.G.Widdowson
OXFORD UNIVERSITY PRESS

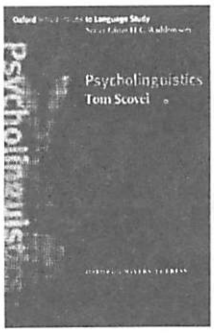
Linguistics
H.G. Widdowson

Oxford Introductions to Language Study
Series Editor H.G. Widdowson
Second Language Acquisition
Rod Ellis
OXFORD UNIVERSITY PRESS

Second Language Acquisition
Rod Ellis


Oxford Introductions to Language Study
Series Editor H.G. Widdowson
Pragmatics
George Yule
OXFORD UNIVERSITY PRESS

Pragmatics
George Yule


Oxford Introductions to Language Study
Series Editor H.G. Widdowson
Psycholinguistics
Tom Scovel
OXFORD UNIVERSITY PRESS

Psycholinguistics
Thomas Scovel

Oxford University Press
2-4-8 Kanamecho, Toshima-ku,
Tokyo 171-8585
TEL: 03-5995-3801 FAX: 03-5995-3919
e-mail: OUP-elt@net21.or.jp
http://www.oup.uk/elt/

Oxford Introductions to Language Study

Series Editor: H.G. Widdowson

Each title in this series of brief introductions to the main areas of language study, contains a state-of-the-art literature survey and a selection of readings from key texts. A bibliography and glossary are also included, along with questions to check understanding. The series is ideal for linguistics students and language teachers.

Please send me more information about the Oxford Introductions to Language Study series:

Name: ______________________

School Name: ______________________

Telephone: ______________________

Fax: ______________________

Address: School ○ Home ○

____

____

〒

____