日本言語テスト学会 (JLTA)

第 19 回（2015 年度）全国研究大会発表要綱

Handbook

of

the 19th Annual Conference

of

the Japan Language Testing Association

Conference Theme:

Toward the Introduction of Productive Skill Tests to University Entrance Examinations

日時： 2015 年 9 月 5 日（土）13:00－17:00 ～ 9 月 6 日（日）9:00－18:00

会場：中央大学 後楽園キャンパス
〒112-8551 東京都文京区春日1-13-27
http://www.chuo-u.ac.jp/access/kourakuen/

主 催：日本言語テスト学会 事務局
〒270-1695 千葉県印西市平賀学園台1-1
順天堂大学さくらキャンパス 小泉利恵研究室
TEL: 0476-98-1001 (代表) FAX: 0476-98-1011(代表)
https://jlta.ac/

The Japan Language Testing Association
1-1 Hiraka Gakuendai, Inzai, Chiba 270-1695, Japan
Telephone: +81-476-98-1001 Fax: +81-476-98-1011
https://jlta.ac/
E-mail: rie-koizumi@mwa.biglobe.ne.jp

Conference Venue:
Chuo University, Korakuen Campus
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
http://global.chuo-u.ac.jp/english/siteinfo/visit/korakuen/

全 国 研 究 大 会 本 部 委 員 (Annual Conference Committee)

渡部 良典 / Yoshinori WATANABE	（上智大学 / Sophia University）
小泉 利恵 / Rie KOIZUMI	（順天堂大学 / Juntendo University）
飯村 英樹 / Hideki IIMURA	（熊本県立大学 / Prefectural University of Kumamoto）
印南 洋 / Yo IN'NAMI	（中央大学 / Chuo University）

全 国 研 究 大 会 運 営 委 員 (Annual Conference Steering Committee)

印南 洋 / Yo IN'NAMI	（中央大学 / Chuo University）
金子 恵美子 / Emiko KANEKO	（会津大学 / University of Aizu）
井上 千尋 / Chihiro INOUE	（University of Bedfordshire, U.K.）
木村 哲夫 / Tetsuo KIMURA	（新潟青陵大学 / Niigata Seiryo University）
澤木 泰代 / Yasuyo SAWAKI	（早稲田大学 / Waseda University）
島谷 浩 / Hiroshi SHIMATANI	（熊本大学 / Kumamoto University）
島田 勝正 / Katsumasa SHIMADA	（桃山学院大学 / Momoyama Gakuin University）
李 洙任 / Soo im LEE	（龍谷大学 / Ryukoku University）
谷 誠司 / Seiji TANI	（常葉大学 / Tokoha University）
藤田 智子 / Tomoko FUJITA	（東海大学 / Tokai University）

全 国 研 究 大 会 実 行 委 員 (Annual Conference Executive Committee)

阿部 真理子 / Mariko ABE	（中央大学 / Chuo University）
印南 洋 / Yo IN'NAMI	（中央大学 / Chuo University）
髙波 幸代 / Sachiyo TAKANAMI	（東洋大学 / Toyo University）

研 究 発 表 審 査 委 員 (Paper Presentation Abstract Reviewer)

印南 洋 / Yo IN'NAMI	（中央大学 / Chuo University）
金子 恵美子 / Emiko KANEKO	（会津大学 / University of Aizu）
澤木 泰代 / Yasuyo SAWAKI	（早稲田大学 / Waseda University）

Table of Contents（目次）

1. Conference Schedule Overview（大会日程表） 2

2. From the JLTA Office: Information for Conference Participants
（学会事務局からのお知らせ） 9

3. Abstracts（発表要旨） 11

4. Workshop Information（ワークショップ情報） 35

5. Access to the Conference Venue（会場へのアクセス） 40

6. Advertisement/Commercial Exhibit Sponsors（広告・展示協賛企業）
39, 46-49

1 . Conference Schedule Overview

September 5, 2015 (Saturday)

13:00－17:00	Workshop 1 (Conducted in English) "You've done your test—now what? First steps in test analysis" (Conducted in English) Myles GROGAN (Kansai University) Atsushi MIZUMOTO (Kansai University) (Chuo University, Korakuen Campus, Building 6, Room 6401)

16:00－18:00	Board Meeting	(Building 6, Room 6730)

September 6, 2015 (Sunday)

8:20－	Registration	(Building 5, Room 5334)
9:00－9:15	Opening Ceremony	(Building 5, Room 5333)
9:15－10:30	Keynote Speech	(Building 5, Room 5333)
10:30－11:00	Invited Lecture	(Building 5, Room 5333)
11:10－11:40	Presentation I	
11:45－12:15	Presentation II	
12:15－13:45	Lunch Break	(JLTA Committee Meetings: Building 5, Room 5334)
13:45－14:15	Presentation III	
14:20－14:50	Presentation IV	
14:55－15:25	Presentation V (Institutional Member Presentations)	
15:25－15:45	Break	(Building 6, Room 6301)
15:45－17:15	Symposium	(Building 5, Room 5333)
17:20－17:40	Closing Ceremony	(Building 5, Room 5333)
17:40－18:00	JLTA General Business Meeting	(Building 5, Room 5333)
18:30－20:30	Banquet	(FOUGAU; 5-minute walk from Building 5)

Commercial Exhibits: Buildings 5 and 6, 3rd floor hallway
Lunch Room for Participants & Participants' Lounges:
Building 6, Room 6301 (Please use only this room for lunch)
(Free refreshments are available.)
Headquarters: Building 5, Room 5334

September 6, 2015 (Sunday)

11:10－15:10	Workshop 2 "How to develop tests that improve students' English proficiency: Focus on validity and reliability" (Conducted in Japanese) Kiwamu KASAHARA (Hokkaido University of Education) Rintaro SATO (Nara University of Education) (Chuo University, Korakuen Campus, Building 6, Room 6401)

Program of the 19th JLTA Annual Conference

September 6, 2015 (Sunday)

8:20－ **Registration** (Building 5, Room 5334)
Conference Attendance Fee: JLTA Members: ¥1,000
Non-members: ¥3,000
(Graduate students: ¥1,000; undergraduate students: ¥0)

9:00－9:15 **Opening Ceremony** (Building 5, Room 5333)
Coordinator: Emiko KANEKO (University of Aizu)
Greetings: Yoshinori WATANABE (JLTA President; Sophia University)

9:15－10:30 **Keynote Speech** (Building 5, Room 5333)
Coordinator: Yoshinori WATANABE (Sophia University)
Title: A Socio-Cognitive Approach to Assessing Speaking and Writing: GEPT Experience
Lecturer: Jessica R. W. WU (The Language Training and Testing Center, Taiwan)

10:30－11:00 **Invited Lecture** (Building 5, Room 5333)
Coordinator: Yoshinori WATANABE (Sophia University)
Title: Building Your Own Validity Argument
Lecturer: Carol A CHAPELLE (Iowa State University)

11:10－12:15 **Presentations I and II**
(Presentation: 20 minutes; Discussion: 10 minutes) (Building 6, Rooms 6302, 6309, 6310, 6317, 6318)

12:20－13:45 **Lunch Break**
Lunch Room for Participants: Building 6, Room 6301
JLTA Committee Meetings: Building 5, Room 5334

13:45－15:25 **Presentations III and IV and Institutional Member Presentations (V)**
(Presentation: 20 minutes; Discussion: 10 minutes) (Building 6, Rooms 6302, 6309, 6310, 6317, 6318)

15:25－15:45 **Break** (Building 6, Room 6301)

15:45－17:15 **Symposium** (Building 5, Room 5333)
Theme: Toward the Introduction of Productive Skill Tests to University Entrance Examinations
Coordinator & Panelist: Masashi NEGISHI (Tokyo University of Foreign Studies)
The Introduction of Four Skills Tests into University Entrance Exams in Japan: Problems to be Solved
Panelist: Neil DRAVE (Hong Kong Examinations and Assessment Authority)
Tests of Productive Skills in Hong Kong's Diploma of Secondary Education (HKDSE)
Panelist: Yan JIN (Shanghai Jiao Tong University)
How Standardized are Performance-Based Assessment of Productive Skills?
Discussant: Jessica R. W. WU (The Language Training and Testing Center, Taiwan)

17:20－17:40 **Closing Ceremony** (Building 5, Room 5333)
Coordinator: Yukie KOYAMA (Nagoya Institute of Technology)

17:40－18:00 **JLTA General Business Meeting** (Building 5, Room 5333)
Selection of the chair
Reporter: Rie KOIZUMI (JLTA Secretary General; Juntendo University)

18:30－20:30 **Banquet** (FOUGAU)
Coordinator: Hidetoshi SAITO (Ibaraki University), Hideki IIMURA (Prefectural University of Kumamoto)

大会日程表

2015 年 9 月 5 日（土）

13:00－17:00	「テストデータ分析入門」（英語で実施） 講師：Myles GROGAN（関西大学）、水本 篤（関西大学） 場所：中央大学後楽園キャンパス（6 号館 6401 教室）

16:00－18: 00	役員会	（6 号館 6730 教室）

2015 年 9 月 6 日（日）

8:20－	受付	（5 号館 5334 階）
9:00－9:15	開会行事	（5 号館 5333 教室）
9:15－10:30	基調講演	（5 号館 5333 教室）
10:30－11:00	招待講演	（5 号館 5333 教室）
11:10－11:40	研究発表 I	
11:45－12:15	研究発表 II	
12:15－13:45	昼食	（JLTA 委員会：5 号館 5334 教室）
13:45－14:15	研究発表 III	
14:20－14:50	研究発表 IV	
14:55－15:25	研究発表 V　(賛助会員発表)	
15:25－15:45	休憩	（6 号館 6301 教室）
15:45－17:15	シンポジウム	（5 号館 5333 教室）
17:20－17:40	閉会行事	（5 号館 5333 教室）
17:40－18:00	JLTA 総会	（5 号館 5333 教室）
18:30－20:30	懇親会	（フーゴー；5 号館より徒歩 5 分）

協賛企業展示：　　5・6 号館 3 階廊下（無料の飲み物等が 6401 にございます。）
一般参加者昼食・休憩室：　6 号館 6301 教室（昼食はこの部屋でお願いいたします。）
大会本部：　　5 号館 5334 教室

2015 年 9 月 6 日（日）

11:10－15:10	ワークショップ　「生徒の力を伸ばす定期テストの作り方―妥当性と信頼性に留意して」（日本語で実施） 講師：笠原 究（北海道教育大学）、佐藤 臨太郎（奈良教育大学） 場所：中央大学後楽園キャンパス（6 号館 6401 教室）

日本言語テスト学会第 19 回全国大会プログラム

2015 年 9 月 6 日（日）

8:20－　　**受　付**（5 号館 3 階）　学会参加費：JLTA 会員 1,000 円、未会員 3,000 円
（ただし大学院生は 1,000 円、学部生は無料）

9:00－9:15　**開会行事**（5 号館 5333 教室）
総合司会　　金子 恵美子（会津大学）
挨　　拶　　渡部 良典　（JLTA 会長・上智大学）

9:15－10:30　**基調講演**（5 号館 5333 教室）
司会　　渡部 良典（JLTA 会長・上智大学）
演題　A Socio-Cognitive Approach to Assessing Speaking and Writing: GEPT Experience
講師　Jessica R. W. WU (The Language Training and Testing Center, Taiwan)

10:30－11:00 **招待講演**（5 号館 5333 教室）
司会　渡部 良典（JLTA 会長・上智大学）
演題　Building Your Own Validity Argument
講師　Carol A CHAPELLE (Iowa State University)

11:10－12:15　**研究発表 I・II**（発表 20 分, 質疑応答 10 分 ; 6 号館 6302, 6309, 6310, 6317, 6318 教室）

12:20－13:45　**昼　食**
（一般参加者昼食控室：6 号館 6301 教室、JLTA 委員会：5 号館 5334 教室）

13:45－15:25　**研究発表 III・IV、賛助会員発表 V**
（発表 20 分, 質疑応答 10 分 ; 6 号館 6302, 6309, 6310, 6317, 6318 教室)

15:25－15:45　**休憩**（6 号館 6301 教室）

15:45－17:15　**シンポジウム**（5 号館 5333 教室）（使用言語：英語）
テーマ　Toward the Introduction of Productive Skill Tests to University Entrance Examinations
（産出技能テストの大学入試導入へ向けて）
コーディネーター・パネリスト　根岸 雅史 (東京外国語大学)
The Introduction of Four Skills Tests into University Entrance Exams in Japan: Problems to be Solved
パネリスト　Neil DRAVE (Hong Kong Examinations and Assessment Authority)
Tests of Productive Skills in Hong Kong's Diploma of Secondary Education (HKDSE)
Yan JIN (Shanghai Jiao Tong University)
How Standardized are Performance-Based Assessment of Productive Skills?
討論者　Jessica R. W. WU (The Language Training and Testing Center, Taiwan)

17:20－17:40　**閉会行事**（5 号館 5333 教室）
司会　小山 由紀江（名古屋工業大学）

17:40－18:00　JLTA **総会**（5 号館 5333 教室）
議長選出
報告　小泉 利恵（JLTA 事務局長・順天堂大学）

18:30－20:30　**懇親会**（フーゴー）
司会　齋藤 英敏（茨城大学）、飯村 英樹（熊本県立大学）

Presentation Overview

Time	Part	Main (5333)	Room 1 (6302)	Room 2 (6309)	Room 3 (6310)	Room 4 (6317)	Room 5 (6318)	Room 6 (6401)
9:15 — 10:30	Keynote speech WU		--	--	--	--	--	--
10:30 — 11:00	Invited lecture CHAPELLE		--	--	--	--	--	--
11:10 — 11:40	I	--	KUMAZAWA & SATO	小野塚	ATEŞ	--	SAITO	Workshop 2 笠原 & 佐藤 (11:10－15:10)
11:45 — 12:15	II	--	HSU	根岸 & 大塚	YIM*	MCLEAN	TAKANAMI	
Lunch break								
13:45 — 14:15	III	--	BOWER	長沼、高野 & 井之川	AHN*	HOLSTER & LAKE	--	
14:20 — 14:50	IV	--	KOIZUMI, IN'NAMI, & FUKAZAWA	藤田	PARK	STUBBE	--	
14:55 — 15:25	V	--	国際教育交換協議会（CIEE）日本代表部: Criterion®	公益財団法人日本英語検定協会：実用英語技能検定（英検）	グローバル・コミュニケーション&テスティング：TOEFL Junior	株式会社アルク：TSST	ピアソン・ジャパン株式会社：Progress	
Break								
15:45 — 17:15	Symposium							

*Invited Paper

Presentation Details

Room 5333

Keynote speech chair　Yoshinori WATANABE (Sophia University)

Keynote speech summary　Naoyuki NAGANUMA (Tokai University)

Symposium summary　Kiwamu KASAHARA (Hokkaido University of Education)

Part	Presenter (Affiliation)	Title (Page)
	Keynote speech Lecturer: Jessica R. W. WU (The Language Training and Testing Center, Taiwan)	A Socio-Cognitive Approach to Assessing Speaking and Writing: GEPT Experience (p. 11)
	Invited lecture Lecturer: Carol A CHAPELLE (Iowa State University)	Building Your Own Validity Argument (p. 12)
I－V	------	------
	Symposium Masashi NEGISHI (Tokyo University of Foreign Studies), Neil DRAVE (Hong Kong Examinations and Assessment Authority), Jin YAN (National College Entrance English Testing Committee), and Jessica R. W. WU (The Language Training and Testing Center, Taiwan)	Toward the Introduction of Productive Skill Tests to University Entrance Examinations (pp. 13-15)

Room 1 (6302)

Chair Part I Emiko KANEKO (University of Aizu)
Part II Emiko KANEKO (University of Aizu)
Part III Yasuyo SAWAKI (Waseda University)
Part IV Yasuyo SAWAKI (Waseda University)

Part	Presenter (Affiliation)	Title (Page)
I	Takaaki KUMAZAWA (Kanto Gakuin University) Takanori SATO (Tamagawa University)	Validity Argument for the EIKEN Can-Do Questionnaire Interpretations (p. 16)
II	Tammy Huei-Lien HSU (Fu-Jen Catholic University, Taipei, Taiwan)	Use of TOEFL Scores to Fulfill College Exit Requirement in Taiwan: A Validation Study (p. 17)
III	Jack BOWER (Hiroshima Bunkyo Women's University & Macquarie University)	Initial Results from an Argument-Based Test Validation of CEFR-Aligned Tests of Reading and Listening, Focusing on the Development Phase of Validation (p. 18)
IV	Rie KOIZUMI (Juntendo University) Yo IN'NAMI (Chuo University) Makoto FUKAZAWA (University of the Ryukyus)	An Argument-Based Validation Framework of a Paired Oral Test in University Classroom Contexts (p. 19)
V	山口学（国際教育交換協議会（CIEE）日本代表部）	ライティング指導ツール Criterion®のご紹介 (p. 33)

Room 2 (6309)

Chair Part I 木村哲夫（新潟青陵大学）
Part II 木村哲夫（新潟青陵大学）
Part III 根岸純子（鶴見大学）
Part IV 高野正恵（上智大学大学院）

Part	Presenter (Affiliation)	Title (Page)
I	小野塚 若菜（筑波大学大学院）	ビジネス日本語テストの読解テスト項目における DIF の推測と検出―経済学分野の専門知識の有無による DIF 分析― (p. 20)
II	根岸 純子（鶴見大学） 大塚 賢一（名古屋短期大学）	テスト形式の相違による被験者発話の評価と特徴 (p. 21)
III	長沼 君主（東海大学） 高野 正恵（上智大学大学院） 井之川 睦美（東京外国語大学）	CEFR 準拠スピーキング評価ルーブリック及び評価タスクの精緻化による改善とその効果の検証 (p. 22)
IV	藤田 智子（東海大学）	第1回言語教育プログラム評価アンケート結果の考察 (p. 23)
V	仲村圭太・森田真之介・鈴木知花・金子雅道・濵野謙太（公益財団法人 日本英語検定協会）	実用技能英語検定試験の文法問題の難易度に関する基礎調査 (p. 33)

Room 3 (6310)

Chair Part I Soo im LEE (Ryukoku University)
Part II Ersin ATEŞ (Anadolu University, Turkey)
Part III Su Yon YIM (Korea Institute for Curriculum and Evaluation)
Part IV Kyungja AHN (Seoul National University of Education, Seoul, South Korea)

Part	Presenter (Affiliation)	Title (Page)
I	Ersin ATEŞ (Anadolu University, Turkey)	Assessing Productive Skills: Challenges Faced by a Novice English Lecturer (p. 24)
II	Su Yon YIM* (Korea Institute for Curriculum and Evaluation)	Primary School Teachers' Intention to Use Peer Assessment in English Writing (p. 25)
III	Kyungja AHN* (Seoul National University of Education, Seoul, South Korea)	The Teaching English in English Certification System for Elementary and Secondary English Teachers in South Korea (p. 26)
IV	Moonyoung_PARK (Iowa State University & University of Aizu)	A Task-Based Needs Analysis for Aviation English Assessment (p. 27)
V	小椋茂生（グローバル・コミュニケーション&テスティング）	TOEFL Junior Comprehensive～TOEFL iBT に直結する　世界共通４技能型コンピュータベーステスト　大学入試でも活用～ (p. 33)

*Invited Paper

Room 4 (6317)

Chair Part II Soo im LEE (Ryukoku University)
Part III Katsumasa SHIMADA (Momoyama Gakuin University)
Part IV Katsumasa SHIMADA (Momoyama Gakuin University)

Part	Presenter (Affiliation)	Title (Page)
I	------	------
II	Stuart MCLEAN (Kansai University)	The Appropriateness of Utilizing the Word Family Unit With Japanese University Students (p. 28)
III	Trevor A. HOLSTER (Fukuoka University) J. W. LAKE (Fukuoka Jogakuin University)	Guessing and the Rasch Model (p. 29)
IV	Raymond STUBBE (Kyushu Sangyo University)	Are Residuals Important When Comparing Yes-No Test Scoring Formulas? (p. 30)
V	福江友樹・平野琢也（株式会社アルク）	英語スピーキング能力テスト TSST のご紹介 (p. 34)

Room 5 (6318)

Chair Part I Tomoko FUJITA (Tokai University)
Part II Tomoko FUJITA (Tokai University)

Part	Presenter (Affiliation)	Title (Page)
I	Hidetoshi SAITO (Ibaraki University)	Initial Evidence on the Validity of the Assessment of the Junior High School Three-Person Discussion. (p. 31)
II	Sachiyo TAKANAMI (Toyo University)	*With* versus *Without*: Comparing the Effects of L1 Translation in Multiple-Choice Spelling Test Formats (p. 32)
III	------	------
IV	------	------
V	上村武司（ピアソン・ジャパン株式会社）	オンラインテスト『Progress』の実践報告 (p. 34)

2. From the JLTA Office: Information for Conference Participants

To All Participants
- Please use public transportation to come to the venue.
- No smoking is permitted on campus.

Registration

1. The conference registration site is located at the lobby on the third floor of Building 5.
2. The conference attendance fee is ¥1,000 for members (including institutional members) and ¥3,000 for non-members (¥1,000 for non-member graduate students and ¥0 for non-member undergraduate students). If non-members apply for membership at the registration desk, the conference attendance fee will be ¥1,000. The JLTA annual membership fee is ¥8,000 for a general member and ¥5,000 for a student member; the admission fee for the JLTA membership is ¥1,000.
3. Please wear your conference name card throughout the conference.
4. The banquet fee is ¥3,000. The banquet registration is conducted at the registration desk. The banquet will be held at FOUGAU. (See the map on p. 43).
5. The conference handbook is available at the registration desk on the day of the conference and is not sent by post in advance.

Lunch and Participants' Lounge etc.

1. Please use Room 6301 on the 3rd floor of Building 6 for eating lunch as the Participants' Lounges. Complimentary refreshments are available in Room 6301.
2. Lunch can be purchased at a nearby Lawson convenience store (50 meters west of the campus gate) or restaurants around Korakuen station (see the map on p. 43). On-campus cafeterias are closed on Sunday.

Accommodation

We are afraid that we provide no accommodation services through our association. Please make arrangements by yourself.

***Emergency Contact E-Mail Address*:** rie-koizumi@mwa.biglobe.ne.jp (Rie KOIZUMI)
Received e-mail messages will be automatically forwarded to her mobile phone.

To Presenters

1. Presenters will have 20 minutes to present their paper, followed by 10 minutes for discussion.
2. Please register at the registration desk first. Please go to the designated room 10 minutes prior to the starting time of the presentation.
3. If you are not a JLTA member, please pay the ¥3,000 "Presentation fee" (different from "Attendance fee") at the registration desk. This rule applies to every presenter on the program.
4. You are expected to connect your computer to the projector and operate it yourself. The projector and connector cable are in the room. There is sound system and you can play sounds from your computer (There will be an audio terminal connector [for PC connection through a stereo mini plug] and a D-sub 15-pin cable [for audio] in the presentation room. Mac users should bring their own Mini DisplayPort to VGA Adapter.
5. Wi-Fi Internet access is available.
6. Please bring handouts in case your PC or the projector does not work.
7. If you need a letter of invitation, contact Rie KOIZUM at rie-koizumi@mwa.biglobe.ne.jp

To Chairs

1. One chair is assigned to each presentation.
2. Please make sure that the presentation does not exceed the allotted time.
3. Please start the presentation at the time designated in the program. Please do not change the starting time or the order of the presentations.

学会事務局からのお知らせ

大会参加者へのご案内

・学会用に駐車場は準備しておりません。公共の交通手段をお使いください。
・大学内は禁煙です

■受付

1. 5 号館 3 階 5334 教室で行います。
2. 学会参加費は、会員 1,000 円（個人・賛助会員を含む）、未会員 3,000 円（ただし大学院生は 1,000 円、学部生は無料）です。未会員の方でも、受付で入会手続きを行えば学会参加費は 1,000 円となります。JLTA 年会費は、一般会員は 8,000 円、学生会員は 5,000 円、入会費は 1,000 円です。
3. 学会中は、名札をお付けください。
4. 懇親会費は 3,000 円です。受付でお支払いください。懇親会は「フーゴー」にて開かれます（マップ p. 43 参照）。
5. 参加者の方には、『JLTA 第 19 回（2015 年度）全国研究大会発表要綱』を受付で配布します。『要綱』は事前に郵送しませんので、ご注意ください。

■昼食・休憩室等

1. 昼食・休憩室として、6 号館 3 階 6301 をご利用ください。無料の飲み物は 6 号館 3 階 6301 にございます。
2. 昼食は受付では販売いたしません。近隣のローソンおよび後楽園近辺で昼食が購入できます。

■宿泊

宿泊の斡旋はいたしておりません。

■緊急連絡先の E メールアドレス　rie-koizumi@mwa.biglobe.ne.jp（小泉利恵）
携帯電話の E メールアドレスに転送されます。

発表者へのご案内

1. 20 分の発表と 10 分の質疑応答の時間があります。
2. 受付を済まされ、発表開始 10 分前には、発表会場にお越しください。
3. 未会員の方は、「研究発表費」（参加費とは別）の 3,000 円を、受付時に支払いをお願いいたします。これは、プログラムに掲載する共同研究者すべてに適用されます。
4. ご発表にプロジェクタを使われる方は、PC との接続作業は各自で行ってください。発表会場にはプロジェクタとともに接続ケーブルもございます。PC からの音声を会場のスピーカーから出すことができます (音声ケーブル [ステレオミニプラグによる PC 接続用] および RGB ケーブル [D-sub15 ピン] は会場にあります)。Mac 用のケーブルはご持参ください。
5. Wi-Fi インターネットへの接続ができます。
6. 予測できない不具合に備え、ハンドアウトのご持参をお勧めします。
7. 出張依頼状などが必要な方は、rie-koizumi@mwa.biglobe.ne.jp (小泉利恵) までご連絡ください。

司会の方へのご案内

1. 1 発表につき 1 人の司会をお願いしています。
2. 時間オーバーにならないようにお気をつけください。
3. 決められた時間に発表を始めてください。発表時間や順番を変えないでください。

3. Abstracts (発表要旨)

Keynote Speech (Room 5333) 9:15－10:30

A Socio-Cognitive Approach to Assessing Speaking and Writing: GEPT Experience

Jessica R. W. WU (The Language Training and Testing Center, Taiwan)
jw@lttc.ntu.edu.tw

The General English Proficiency Test (GEPT) is a 5-level, criterion-referenced EFL testing system implemented in Taiwan to assess the general English proficiency of EFL learners at all levels. The GEPT was designed as a skills-based test battery assessing both receptive (listening and reading) and productive (speaking and writing) skills. Since its first administration in 2000, the GEPT has been taken by more than 6.5 million learners, and has become the largest-scale standardized English test in Taiwan. GEPT scores are widely used for education (e.g., university entrance and graduation benchmark) and employment purposes (e.g., qualifications for civil servants and English language teachers). In the wake of the introduction of productive skills to the university entrance examination system in Japan, this talk aims to share the GEPT experience. Several key issues about speaking and writing tests will be presented in relation to the socio-cognitive framework for validation (Weir, 2005). A number of examples about GEPT validation will be illustrated to demonstrate that both a priori and a posteriori validity evidence are required to establish test quality. The talk also emphasizes the importance of facilitating communication between test developers and teachers when introducing new assessment. Also, raising teachers' assessment literacy is an effective step to allow stakeholders to cope collaboratively with the challenges of developing students' speaking and writing skills.

Bio

Jessica WU holds a PhD in Language Testing. She is currently Program Director of the R&D Office at the Language Training and Testing Center (LTTC). She also teaches language testing courses in the PhD program of the National Taiwan Normal University. She has been deeply involved in the research and development of the General English Proficiency Test (GEPT), which targets English learners at all levels in Taiwan. She also serves as an adviser to the government at a national level on the development of L1 tests. She has published numerous articles and book chapters in the field of language testing and has presented her work at conferences around the world. She is currently a member of the Executive Board of the Asian Association for Language Assessment (AALA). Her current projects include promoting assessment literacy to test stakeholders, constructing language assessment for specific academic purposes, and supporting the internationalization of secondary and high education in collaboration with universities, educational organizations, and examination boards at both local and international levels.

Invited Lecture (Room 5333) 10:30－11:00

Building Your Own Validity Argument

Carol A CHAPELLE (Iowa State University)
carolc@iastate.edu

Argument-based validity has become important as a means of conceptualizing validation research and presenting theoretical and empirical rationales for language test interpretation and use (Chapelle, Enright & Jamieson, 2008; Chapelle, Chung, Hegelheimer, Pendar & Xu, 2010). Despite its demonstrated utility, language testers can find argument-based validity somewhat difficult to apply to their own testing contexts. Much of what is written on this topic is less accessible than one might hope, and examples of how to build a validity argument are few. Moreover, the terms used in validity argumentation (e.g., generalization and explanation) are grounded in traditions of educational measurement and used inconsistently across different sources relied upon in language testing (Bachman, 2005; Bachman & Palmer, 2010; Kane, 2006). My presentation will reveal how the modern tools of argument-based validity can be used to create novel validity arguments for specific test interpretations and uses. To do so, I will introduce the fundamental concepts (e.g., inference, claim and steps in an argument) and provide examples of how these can vary depending on the intended interpretation and use of a language test. I will outline the process of validity argument development beginning with an interpretation/use argument consisting of what the tester would like to be able to claim about test interpretation and use (i.e., what the scores mean and how they should be used). The interpretation/use argument provides the plan for undertaking particular types of research. I will show how results of the research can serve as backing, or support, for claims in the validity argument. Overall, my presentation will demonstrate validity argument as a means for language testers to conceptualize the justification of test score interpretation and use rather than a template of research requirements created by others.

References

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*(1), 1-34.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language™*. New York, NY: Routledge.

Chapelle, C. A., Chung, Y-R., Hegelheimer, V., Pendar, N., & Xu, J. (2010). Towards a computer-delivered test of productive grammatical ability. *Language Testing*, *27*(4), 443-469.

Kane, M. (2006). Validation. In R. Brennen, (Ed.), *Educational Measurement* (4th Edition), (pp 17-64). Westport, CT: Greenwood Publishing.

Bio

Carol A. CHAPELLE is Distinguished Professor of Liberal Arts and Sciences at Iowa State University, where she teaches primarily in the doctoral program in applied linguistics and technology. Her research has resulted in the publication of ten books and over one hundred papers in journals and books. Professor Chapelle's research into technology for language assessment and learning has led her into many facets of applied linguistics including the study of grammar and lexis, multimodal communication, and language ideology—an exploration that fostered the conception of the field's first comprehensive, ten-volume *Encyclopedia of Applied Linguistics* (Wiley-Blackwell, 2013). In addition to editing the *Encyclopedia*, she is co-editor of the Cambridge Applied Linguistics Series, she has served as editor of *TESOL Quarterly* (1999-2004), and she has served as president of the American Association for Applied Linguistics (2006-2007). Her recent awards include the 2012 Lifetime Achievement Award in Language Testing given by the University of Cambridge and the International Language Testing Association, the 2012 Samuel J. Messick Memorial Lecture Award given by Educational Testing Service in Princeton, New Jersey, as well as the 2015 Distinguished Scholarship and Service Award given by the American Association for Applied Linguistics.

Symposium (Room 5333)　15:45－17:15

Toward the Introduction of Productive Skill Tests to University Entrance Examinations
(産出技能テストの大学入試導入へ向けて)

Coordinator	**Masashi NEGISHI (Tokyo University of Foreign Studies)**
Panelists	**Masashi NEGISHI (Tokyo University of Foreign Studies)**
	Neil DRAVE (Hong Kong Examinations and Assessment Authority)
	Yan JIN (Shanghai Jiao Tong University)
Discussant	**Jessica R. W. WU (The Language Training and Testing Center, Taiwan)**

Introduction

Coordinator: Masashi NEGISHI (Tokyo University of Foreign Studies)
negishi@tufs.ac.jp

The Japanese university entrance exam system is undergoing an unprecedented scale of reform. In the most significant development since 1979, when the First-stage Joint Achievement Test was introduced, the Central Council for Education has proposed to replace the current system, known as the National Center Test for University Admissions, with a completely new test.

The reform is expected to have a great impact on English education, because the Ministry of Education, Culture, Sports, Science and Technology (MEXT) is planning to assess four skills using private-sector English proficiency tests. Productive skills, especially speaking, have long been the neglected aspect of English proficiency in Japan. If they are to be included in high-stakes tests, such as university entrance exams, it would exert a beneficial impact not only on language classrooms but also on society at large. However, considering the sheer number of the candidates for the National Center Test, the introduction of direct testing of productive skills will present significant challenges in many respects. However, as Hughes (2002) argues, "we cannot afford not to introduce a test with a powerful beneficial backwash effect." The question is how best to implement the new system.

This symposium thus focuses on the challenges we are expected to encounter. Hopefully we find some solutions by sharing the experiences of forerunners overseas. Negishi outlines the current English university entrance exam reform in Japan, and discusses its foreseeable impact and difficulties we would have to overcome. Drave will share experiences of Hong Kong's school-leaving qualification called the HKEAA's Diploma of Secondary Education (DSE), which includes speaking and writing components. Yan, who, in the past two and half decades, has been working for the CET, a large-scale English language assessment for university students in China, will elaborate on the assessment of productive skills (speaking, writing and, translation) in the CET.

Symposium Paper 1: The Introduction of Four Skills Tests into University Entrance Exams in Japan: Problems to be Solved

Masashi NEGISHI (Tokyo University of Foreign Studies)
negishi@tufs.ac.jp

The Central Council for Education announced plans to reform Japanese entrance examinations. As part of these reforms, MEXT is planning to assess four skills using private-sector English proficiency tests, a development expected to have a great impact on English Education in Japan. Previous research shows that there has been a linguistic gap between entrance exams and the Course of Study or authorized textbooks. However, the real issue is that we have not put due weight on the assessment of productive skills. Speaking in particular has long been neglected in Japan, and it is believed that this lack of emphasis has resulted in poor productive skills among the Japanese people.

The introduction of the four skills tests into university entrance exams might revolutionise the teaching and learning of English in Japan. However, it should be noted that, under the current system, more than 500,000 candidates sit the English exam of the National Center Test for University Admissions. Considering this sheer number of the candidates, the introduction of direct testing of productive skills will present many significant challenges pertaining to administration and scoring. Although one might argue that machine delivery and scoring will solve most of the problems, we cannot tell how people may react to these new technologies. In addition, it will be necessary to develop a system to resolve inevitable technical issues. A related issue concerns the challenge of constructing a conversion table for numerous English proficiency tests, given that diverse tests are developed on different models.

Despite these challenges, it is the author's sincere wish to find a way to carry through this unprecedented reform.

Bio

Masashi NEGISHI is Professor at Tokyo University of Foreign Studies (TUFS), Japan. He received his PhD from the University of Reading, UK. He has participated in a number of research projects, including national education surveys. His current interests focus on the application of the Common European Framework of Reference for languages (CEFR) to English language teaching in Japan and the analysis of criterial features across the CEFR levels. He was on the Commission on the Development of Foreign Language Proficiency and is at present a member of various committees on Foreign Language Education Policy at national and local level.

Symposium Paper 2: Tests of Productive Skills in Hong Kong's Diploma of Secondary Education (HKDSE)

Neil DRAVE (Hong Kong Examinations and Assessment Authority)
ndrave@hkeaa.edu.hk

In 2012, Hong Kong introduced a new qualification for school leavers, the Diploma of Secondary Education (HKDSE). The HKDSE employs standards-referenced reporting (SRR) of results, which means that candidates' levels of performance are reported with reference to explicit standards. English Language is compulsory for all candidates. This brief presentation summarises the tests of productive skills in HKDSE English Language, focusing on Speaking.

Bio

Neil DRAVE is the Senior Manager responsible for English subjects at the Hong Kong Examinations and Assessment Authority (HKEAA), which runs the city's public examinations. Before joining the HKEAA, Neil spent many years teaching English, Literature and American Studies at various Hong Kong universities.

Symposium Paper 3: How Standardized are Performance-Based Assessment of Productive Skills?

Yan JIN (Shanghai Jiao Tong University)
yjin@sjtu.edu.cn

Comparable treatment of all test takers is an essential requirement of test fairness, especially in the case of large-scale testing. To meet the requirement, standardized procedures are established for test construction, administration and scoring. Standards for Educational and Psychological Testing (American Educational Research Association et al., 1999; 2014), for example, outlines a set of standardization procedures for test development and use. Test standardization, however, poses great challenges for performance-based assessment of productive skills, which is increasingly employed in language assessment to improve construct representation and test washback. This presentation will briefly examine the level of standardization in the assessment of productive skills, including writing, speaking and translation, in the College English Test (CET). The CET is a test for learners of English at the tertiary level in China and has a candidature of about 18 million each year. First, I will discuss the level of standardization in the implementation of the computer-based CET Spoken English Test. Second, I will address the issue of task difficulty which is of particular relevance to the construction of performance-based assessment tasks. Third, I will discuss the technical measures to ensure rater consistency in rating performances on writing and translation. Finally, I will argue for the view that standardization of large-scale performance-based assessment of productive skills is a necessary prerequisite of test validity.

Bio

Yan JIN is Professor at Shanghai Jiao Tong University, P. R. China. She is Chair of the National College English Testing Committee of China. She has been involved in many research projects related to test design and development and has published many articles in academic journals. She is co-editor-in-chief of *Language Testing in Asia* and is also on the editorial board of international journals such as *Language Testing*, *Classroom Discourse*, *Asia TEFL* and a number of Chinese journals such as *Foreign Language Testing and Teaching*, *Foreign Languages in China*, *Foreign Language World*, *Foreign Language Education in China*, *Contemporary Foreign Languages Studies*.

Paper Session

Room 1 (6302)　Part I (11:10－11:40)

Validity Argument for the EIKEN Can-do Questionnaire Interpretations

Takaaki KUMAZAWA (Kanto Gakuin University)
ktakaaki@kanto-gakuin.ac.jp
Takanori SATO (Tamagawa University)

In recent years, test-takers' scores on proficiency tests are reported with statements about real-life tasks that they may be able to accomplish at their English levels. The EIKEN can-do list (STEP, 2008) was developed to show the link between test results and what learners are probably capable of doing in English with their proficiency. The EIKEN test consists of seven grades, and test-takers select a test with an appropriate level and receive a certificate if they pass the test. STEP (2008) suggests that the list can be used as a self-assessment material for learners and a reference resource for test score users (e.g., companies) when interpreting the test result. Since the EIKEN test is one of the popular proficiency tests in Japan and a strong driving force to learn English, this study investigates its validity as a self-assessment material.

This study uses the argument-based validity framework to investigate the following research questions: (a) to what extent do the can-do items work, (b) how well do the four-point rating scales function, and (c) to what extent is the questionnaire reliable and dependable, and (d) to what extent is the questionnaire unidimensional. The study examines data elicited from a large-scale survey in which 2,590 junior high school students indicated the level of their confidence of 16 EIKEN can-do items (Sato, 2010). These items consisted of four statements in Grades 5, 4, 3, and Pre-2, respectively, and the statements of each EIKEN grade were composed of the four skill areas. They indicated their confidence levels using the four-point rating scale (1 = cannot at all, 2 = can barely do, 3 = can do, and 4 = can do well).

This study analyzes the data of 1,740 students who provided responses to all the 16 items. Factor analysis is conducted to examine the factor structure of the items. Additionally, the Rasch rating scale model (Andrich, 1978) is used to investigate how the items and the four-point rating scale function as well as the reliability and dimensionality. To estimate the uses of the questionnaire for making relative and absolute decisions, decision studies are conducted.

The results show that the items are working and the rating scales are optimally functioning. The questionnaire is highly reliable but not as dependable. The Rasch model reveals that it is unidimensional, and two factors are extracted in the factor analysis. Ways to use the questionnaire for educational purposes are discussed.

Room 1 (6302)　Part II (11:45－12:15)

Use of TOEFL Scores to Fulfill College Exit Requirement in Taiwan: A Validation Study

Tammy Huei-Lien HSU (Fu-Jen Catholic University, Taipei, Taiwan)
hhsu9uiuc@gmail.com/ 089975@mail.fju.edu.tw

In view of the decline of English language proficiency amongst Taiwanese college students, the Ministry of Education (MOE) of Taiwan has established English exit requirements which call for students to pass any English language test with scores equivalent to at a minimum of B1 level of the CEFR. This study aims to collect concurrent validity evidence of TOEFL score and to investigate the extent to which it can serve as an indicator of students' graduation readiness in Taiwan. The collected data included students' final grades for the required General English course and post-instruction TOEFL scores. The textbook for the General English course was chosen intentionally with content and difficulty level mapped to the B1 level of the CEFR. It is hypothesized that students who pass the General English course have a greater chance of obtaining a TOEFL score equivalent to at least the CEFR B1 level.

Data was collected in a large northern university in Taiwan in spring 2014.　Fifty-seven students who enrolled in the General English class focusing on speaking, listening and reading skills, took a pre- and post-instruction TOEFL test. Both quantitative (i.e. correlation analysis, paired t-test) and qualitative (i.e. interview) methods were included for data collection and analysis. Given several speaking tests were un-scorable due to poor voice quality, the analysis only included scores of reading, listening and writing section to maximize the usable data. Two sets of scores for each participant were created: 1) partial holistic score (i.e. excluding speaking scores), and 2) each section score and the partial holistic score converted to the CEFR proficiency level. Both sets of data suggest post-instruction TOEFL scores are not appropriate indicators of students' readiness for graduation. There was little association between students' TOEFL scores and their final grades for the General English course(r = .295). This was further substantiated by the percentage of students reaching the graduation threshold (i.e. B1 of the CEFR): while 94.74% (N = 54) of the students passed the general English course, only 63.24% (N = 35) of students reached the graduation threshold as measured by the TOEFL. In other words, by relying on the TOEFL score alone, more than one third of participants in this study failed to satisfy the graduation requirement. Interview data also indicated low face validity of the TOEFL test. Despite the constructs measured in the TOEFL and General English course different, implications drawn from this study suggest a need for a test preparation course to facilitate students fulfilling the graduation requirement, if test results will be used as a sole tool for evaluation. Furthermore, implication for the broader impact of the MOE's policy and needs for collecting context validity evidence will be discussed.

Room 1 (6302)　Part III (13:45－14:15)

Initial Results from an Argument-Based Test Validation of CEFR-Aligned Tests of Reading and Listening, Focusing on the Development Phase of Validation

Jack BOWER (Hiroshima Bunkyo Women's University & Macquarie University)
jbower@h-bunkyo.ac.jp/ jack.bower@students.mq.edu.au

In-house tests of reading and listening ability intended to place students at levels A1-B1 of the Common European Framework of Reference for Languages (CEFR) are used at the Bunkyo English Communication Center at Hiroshima Bunkyo Women's University. These criterion-referenced tests, known as the Bunkyo English Tests (BETs) are used for course and class placement. They are also intended as one of the measures for student progress against CEFR-based curricula.

A validation study of the development and implementation of the BETs is being conducted using an argument-based test validation framework. The argument based framework adopted draws on aspects of existing frameworks, particularly Kane's (2013) Interpretation and Use Argument (IUA), the expanded version of Kane's (2004) earlier argument-based framework applied by Chapelle, Enright and Jamieson (2008) for validation of the TOEFL, and Bachman and Palmer's (2010) Assessment Use Argument. This validation study focuses on what Kane calls the development phase of test validation in which the IUA and the test are developed together, and as evidence is gathered and analyzed, modifications are made to the test and to the IUA until a satisfactory evidence-based alignment is achieved.

This presentation will firstly give a brief outline the components of the BET IUA. The IUA consists of the six inferences elaborated in Chapelle, Enright and Jamieson's framework, namely: domain description, evaluation, generalization, explanation, extrapolation and utilization.

The main focus of this presentation will be an explanation of preliminary backing for several of the inferences in the BET IUA. This backing includes evidence from Rasch analysis of the 2015 BETs, results of student and teacher surveys including factor analysis, correlations with the Oxford Online Placement Test and the TOEIC, and also correlations with student grades, and speaking test scores. In addition changes planned to test structure and content of the 2016 iterations of the BETs based on an assessment of the preliminary backing evidence will be summarized.

References

Bachman, L. & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world.* Oxford: Oxford University Press.

Chapelle, C. A., Enright, M. K., & Jamieson, J. (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2*(3), 135–170.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Room 1 (6302)　Part IV　(14:20－14:50)

An Argument-Based Validation Framework of a Paired Oral test in University Classroom Contexts

Rie KOIZUMI (Juntendo University)
rkoizumi@juntendo.ac.jp
Yo IN'NAMI (Chuo University)
Makoto FUKAZAWA (University of the Ryukyus)

In this presentation, we propose an argument-based validation framework to be adopted for a paired oral test in university classroom contexts. An argument-based approach to validation is being increasingly employed in the field of second language assessment (Chapelle & Voss, 2013). We use the validation framework proposed by Chapelle, Enright, and Jamieson (2008), which specifies the Domain definition, Evaluation, Generalization, Explanation, Extrapolation, and Utilization inferences to argue for the validity of interpretations and uses based on test scores. Each inference has a warrant and some assumptions to be supported by evidence and questioned by rebuttals. The types and the number of inferences, warrants, and assumptions vary across test contexts; thus, test developers/users need to specify their own framework.

Our paired oral test requires two students to interact with each other following a task card that specifies a discussion topic or roles to play. Previous research suggests that paired oral tests can elicit a variety of oral functions and natural, co-constructed discourses (e.g., Galaczi & ffrench, 2011). It is also expected, but as yet untested, that they can generate positive washback on teaching and learning since they can be associated with pair activities in classrooms and are relatively easy to administer. We have conducted two studies, and obtained generally positive evidence. Our first study involved examining rater and task reliability, rating scale properties, a factor structure of the test, and a relationship between paired oral scores and English proficiency. Our second study involved increasing the number of paired oral tasks and examining their quality to make a task pool scaled on a single dimension in order to improve the task relevancy and representativeness of the intended construct. We relate the results—both positive and negative ones—in the argument-based framework and clarify future directions to strengthen the validity argument. For example, whether the test facilitates students' learning process is an important assumption in the Utilization inference, which needs to be examined empirically using, for instance, questionnaires and students' behavioral records. Another assumption to be tested could be whether the test can actually elicit a wide range of intended speech functions, using checklists developed by O'Sullivan, Weir, and Saville (2002), which can be located in the Explanation inference. The strengthening of validity argument by systematically collecting required evidence would help clarify the features of paired orals, thereby facilitating their use by providing a viable option in an in-class speaking assessment.

Room 2 (6309)　Part II　(11:10－11:40)

ビジネス日本語テストの読解テスト項目における DIF の推測と検出―経済学分野の専門知識の有無による DIF 分析―

小野塚 若菜（筑波大学大学院生）
wakanaonozuka@gmail.com

ビジネス日本語能力を測定するテストにおいては, 経済学的な専門知識がある受験者のほうが結果に有利であろうか。

ある特性を測定するテストにおいて, 同じ能力水準を有するが異なる下位集団に属する受験者同士を比較したときその受験者の属する集団の違いによって特定のテスト項目の難易度に違いが生じている場合, その項目は「特異項目機能（Differential Item Functioning; DIF）」を示すと言う（Angoff, 1993）。本研究は，現在開発中でビジネス場面におけるコミュニケーション能力を測定するテスト「BJ-CAT ビジネス日本語テスト」のうち読解テスト項目を資料とし，経済学分野を専門とする受験者群とそれ以外の受験者群による統計的 DIF 分析を行った（Mantel-Haenszel 検定法を使用）。

また，孫ほか（1995）は，統計的な DIF 分析の循環的欠陥を補う上で重要な役割を果たすのが，「教科分野やテスト作成の専門家の主観」であるとしているが，一方で，専門家が推測する DIF 項目は，統計的な DIF 分析で見出される DIF 項目とあまり一致しないという報告もある（Engelhard, 1990; Ross & Okabe, 2004）。そこで本研究では，経済学分野の専門家がどのようなテスト項目で DIF があると推測するかを探るため，経済学を専門とする専門家 5 名による分析が行われた（主観的 DIF 分析）。インタビュー調査の結果，経済の専門家が推測する DIF として 5 つのビリーフが確認できた。統計的 DIF 分析の結果については 2 項目が検出され，項目の内容の検討を試みたが，そこから原因を推察できるような一貫した傾向は見出せなかった。また，統計的 DIF 分析と主観的 DIF 分析の結果は一致せず，先行研究と同様の結果が得られた。

参考文献

Angoff, W. H. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3-23). Hillsdale, NJ: Erlbaum Associates.

Engelhard Jr., G., Hansche, L., & Rutledge, K. E. (1990). Accuracy of bias review judges in identifying differential item functioning on teacher certification tests. *Applied Measurement in Education, 3,* 347-360.

Ross, S. J., & Okabe, J. (2006). The subjective and objective interface of bias detection on language tests. *International Journal of Testing, 6,* 229-253.

孫媛 & 井上俊哉. （1995）. 「アメリカにおける差異項目機能(DIF)研究」. 学術情報センター紀要, 7, 193-216. 国立情報学研究所.

Room 2 (6309)　Part II　(11:45－12:15)

テスト形式の相違による被験者発話の評価と特徴

根岸 純子（鶴見大学）
negishi-j@tsurumi-u.ac.jp
大塚 賢一（名古屋短期大学）

スピーキング・テストにおいて、単独被験者のみを対象とした口頭試験ではなく、より自然なインタラクション能力を測定するため、ペアあるいは少人数グループといった複数被験者に対する口頭試験が各国で実施されている。例えばケンブリッジ英検では、単独と複数被験者のテストを組み合わせて実施している。複数被験者方式が日本であまり普及していないのは、通常の面接や絵の描写といった単独被験者対象の口頭試験と同等の能力測定が可能であるかが未知数であること、また単独および複数被験者の発話の相違点があまり明らかになっていないことも一因であると思われる。そこで、単独と複数被験者の発話に対する評価の違いおよび発話の特徴について分析した。

本研究では、英語力の異なる 24 名の大学生被験者が、①単独被験者による絵の描写、②ペアによる対話、③3 名グループによる対話、の 3 形式の異なるスピーキング活動に参加し、5 名の日本人英語教員がその評価を行った。その後、文字起こしをした産出データを用いて評価者による評価との関連性を分析するとともに、被験者の発話の特徴を比較した。その結果、①の単独被験者の場合は、発話量（単語数＜シラブル数）と評価には高い相関があったが、複数被験者の場合は、発話量との相関はさほど高くなかった。しかし、話者交替当たりの発話量の多寡が評価との関連性を示すことが多くみられた。①の形式においては、絵の描写という性質上、語彙レベルおよび描写内容において制限が起こりがちであるが、②や③では生き生きと自分の考えを述べることにより高得点を獲得する中位レベル（TOEIC で 500～700 点）の学生が多くみられた。また、会話分析から、Galaczi (2008) のいう対話のインタラクション形式（協同的・並列的・非対称的）をみてみると、単独ではみられない、被験者同士の協同的やりとりが多くみられた。これらのことから、複数被験者形式は単独形式と同等の結果を得ることは難しいが、逆に複数でないと測定できない重要な能力を捉えることができると考えられた。

Room 2 (6309)　Part III　(13:45－14:15)

CEFR 準拠スピーキング評価ルーブリック及び評価タスクの精緻化による改善とその効果の検証

長沼君主（東海大学）
n.naganuma@tokai-u.jp
高野正恵（上智大学大学院博士後期課程）
井之川睦美（東京外国語大学）

東京外国語大学英語学習支援センターでは、2009 年度よりヨーロッパ言語共通参照枠（CEFR）に準拠したライティング及びスピーキング評価を行っている。ライティング評価では、English Profile Programme の評価タスクをベースとした、E メール、物語、論説の 3 つのタスク別の評価を行ってきたが、スピーキング評価では、ACTFL-OPI に基づいた Standard Speaking Test の評価タスクを参照して開発を行ったこともあり、ロール・プレイ、シークエンス・ピクチャー、クラスター・ピクチャーのそれぞれのタスクの個別評価は、総合評価の参考としかしてこなかった。

そこで長沼ら（2014）では、スピーキング評価におけるタスク別評価を可能とすべく、ライティング評価と合わせたタスクのジャンル別ルーブリックを開発した。CEFR の Written Assessment Criteria Grid では、言語面の評価観点（Range、Coherence、Accuracy）に加えて、描写文や物語文といったジャンルごとの内容面（Description、Argument）の評価観点が設けられている一方で、Oral Assessment Criteria Grid では、言語面に音声言語としての特徴に応じた 2 観点（Fluency、Interaction）が追加されるもののジャンル別観点は設けられていない。長沼ら（2014）では、CEFR の Illustrative Scales 及び EQUALS-ALTE の開発した Bank of Descriptors（As Scales）の段階的レベル記述を参照し、スピーキング評価における内容面でのジャンル別観点記述を行った。

本研究では、2014 年度後期のタスク別評価のパイロット的導入及び 2015 年度前期の正式導入データをベースとして、2014 年度以前のデータをタスク別に再評価を行った結果と比較をし、内容面での評価観点の導入による総合評価への影響を検証する。また、タスク別評価を実現するにあたって、それぞれのスピーキング評価タスクにおける Follow-up Questions をタスクのジャンルと合わせて絞り込み、精緻化を行ったが、改善された Follow-up Questions の効果についても、その有効性の確認を行う。こうして得られた知見をもとに、現在、ウェブベースの評価・学習タスクの開発も計画しており、言語ポートフォリオとしての活用についても議論する。

Room 2 (6309)　Part IV　(14:20－14:50)

第 1 回言語教育プログラム評価アンケート結果の考察

藤田 智子（東海大学）
tfujita.tomtom@gmail.com

本論は、2014 年 9 月に実施した「大学言語教育プログラムにおけるプログラム評価アンケート調査」をもとにしている。本アンケートは、日本の大学言語教育プログラムに中心的に携わっていらっしゃる先生方を中心にご協力をお願いし、そのうち 40 人の方から回答をいただいた。本アンケートには「プログラム評価を実施された方（1）」と「実施していない方（2）」という 2 種類の異なる用紙があり、回答者は該当する用紙を選ぶ方式であった。全部で 10 問あり、それぞれの最初の質問は、用紙（1）「プログラム評価をした目的は何か」と用紙（2）「プログラム評価をしない理由は何か」となっている。その他の質問では、プログラム評価を実施していない方用の（2）には、用紙（1）とほぼ同じ内容の質問が書かれているが、それぞれ「プログラム評価を実施するとしたら」という条件がついている。他に、プログラムのゴールや目的を設定しているかどうか、また、プログラム評価を実施したのは誰で結果を誰または何に対して報告したか等がある。さらに、プログラム評価のためのデータで重要なものは何であるか、そして、プログラム評価の問題点についても尋ねている。

結果を集計したところ、40 大学のうち、22 大学の先生方からプログラム評価は実施されているという回答が得られた。また、実施しない理由は、1）プログラム評価を担当する教員の負担増、プログラム評価に対する 2）専門知識の不足、3）教員の評価されることへの反発、4）必要性を感じていない、などであった。プログラムのゴールや目標は全体の 75%で設定されており、プログラム評価を実施するのはそのプログラムの教員という回答が全体の 80%を占めた。

今回は限られた人数のアンケートではあったが、大学言語教育プログラム評価に関する何らかの傾向は得られたように思われる。しかし、何をもってプログラム評価を実施したことになるかという定義を提示しなかったため、幾つかの質問に対して答えが限定しにくいという問題点が生じたことが今後の課題である。

Room 3 (6310)　Part I　(11:10－11:40)

Assessing Productive Skills: Challenges Faced by a Novice English Lecturer

Ersin ATEŞ (Anadolu University, Turkey)
er.ates26@gmail.com

Testing and assessment are the core components that help monitor learner's progress in any educational setting, on which both teachers and students place great importance as well as parents because it determines failure and success in many cases. In English language learning / teaching process, testing and assessment are basically realized via a number of tools such as quizzes, assignments, midterm and final exams. Therefore, English Language Teacher Education programs at universities offer some testing and assessment courses to teacher candidates to provide them a background in terms of testing theories, assessment tools and item types. However, it is a matter of question whether these courses really make a teacher candidate ready to prepare tests. The aim of this study is to share the experiences of a novice teacher as a testing unit member in the School of Foreign Languages in a Turkish state university. Having a novice teacher and a novice testing unit member participate in a number of workshops, which are aimed to provide information and hands on experience about how to write items, the qualities of a good item, writing open-ended questions and so on may not be enough to fully accommodate him with all the qualities and experience a testing unit member should have. What's more, expecting him to prepare exams to test productive skills, i.e. writing and speaking may be far too demanding. This study will help you see the test preparation process, especially tests for productive skills, and the challenges from a novice teacher's perspective, which will eventually provide insights into the hardship of the job as well as a platform to discuss hot to overcome the difficulties.

Room 3 (6310)　Part II　(11:45－12:15)

Invited Paper

Primary School Teachers' Intention to Use Peer Assessment in English Writing

Su Yon YIM (Korea Institute for Curriculum and Evaluation)
edu1syy@kice.re.kr

The relationship between assessment and learning has been key concern among researchers. Traditionally, the relation between the two was rather linear; Assessment has been used mainly to measure what students had learned at the end of the learning unit (assessment of learning). In situations where assessment of learning is practiced, students are rather passive in the assessment process. However, the current view on assessment highlights the supportive role of assessment for learning (assessment for learning). In this view, assessment occurs during the learning process and students take active role in the assessment process. The Ministry of Education in Korea encourages teachers to adopt the concept of assessment for learning in classrooms so that students become more autonomous in their learning. Peer assessment is one of the methods which embody the concept of assessment for learning since it provides students with opportunities to be active in the assessment process and receive feedback while they are learning. In fact, the examination of national textbooks reveals that assessment is viewed as part of learning to facilitate the learning process.

The current study investigates primary school teachers' intention to use peer assessment for low-achieving students. A total of 63 primary school teachers (62 females and 1 male) in Korea responded to the questionnaire which was constructed based on the theory of planned behaviour (TPB). Seven–point Likert scale was used from 1 (strongly disagree) to 7 (strongly agree). Results show that Korean primary school teachers in average intended to use peer assessment for teaching writing in English class (M = 4.02, SD = 1.33). The analysis of open-ended questions shows that many teachers consider students' level of English as the main constraint to prevent implementation of peer assessment in English writing. The result suggests that teachers' belief about fixed ability needs to be addressed so as to implement peer assessment effectively.

Room 3 (6310)　Part III　(13:45－14:15)

Invited Paper

The Teaching English in English Certification System for Elementary and Secondary English Teachers in South Korea

Kyungja AHN (Seoul National University of Education, Seoul, South Korea)
kjahn@snue.ac.kr

This presentation aims to provide an overview of the Teaching English in English (TEE) certification system for elementary and secondary English teachers in South Korea. The TEE certification system was adopted in 2009 with a focus on evaluating and developing English teachers' skills in using English for classroom communication. The system is comprised of two levels: TEE-Ace and TEE-Master specifically in Seoul Metropolitan Office of Education, whose names are different in other Metropolitan and Provincial Offices of Education.

This talk first reviews English educational reform policies and other complex educational, social, and cultural contexts that impact English education and teacher education in South Korea. Since the Korean Ministry of Education announced the Communicative Language Teaching-based curricular reforms in the mid-1990s, there has been a consistent call for improving English teachers' linguistic and pedagogical competence. The TEE certification system was adopted to set publicly-available standards for teachers' skills in TEE and to improve students' English communicative competence through English teachers' development in TEE.

The second topic that this presentation focuses on is the qualifications of the TEE-Ace and Master candidates as well as the assessment process they go through to obtain each level. While becoming TEE-Ace certified includes an evaluation of the applicant's English lesson, becoming TEE-Master certified includes evaluation of the applicant's portfolio and an interview.

Third, this presentation reviews the rubrics for assessing the two levels. The Ace rubric is used for evaluating English lessons taught in English focusing on each teacher's instructional skills (70%) and English communication skills (30%). The first domain is comprised of opening, lesson objectives, activities and tasks, and closing of the lesson as well as overall lesson procedures. The second domain consists of the teacher's accurate and fluent English use and appropriate English use for the students' proficiency level. For the TEE-Master level, the rubric examines the teacher's instructional skills and English proficiency as well as English education experiences and qualifications as a TEE-Master teacher.

Lastly, this talk discusses the results of implementing the TEE certification system such as the pass rate of the evaluation and the number of certified TEE-Ace and TEE-Master teachers. Furthermore, English teachers' perceptions of the system are discussed based on related studies and research reports. The effects of the system are also examined from the perspectives of English teachers and students. Suggestions and prospects regarding the TEE certification system are further examined.

Room 3 (6310)　Part IV　(14:20－14:50)

A Task-Based Needs Analysis for Aviation English Assessment

Moonyoung PARK (Iowa State University & University of Aizu)
mypark@u-aizu.ac.jp

In response to growing concerns over aviation safety due to the limited command of aviation English by non-native English speaking practitioners, this study investigated air traffic controllers' real-world tasks in radiotelephony communication with pilots to inform the future development of pedagogic tasks and task typology for aviation English teaching and learning. For this research, a task-based needs analysis (Long & Norris, 2000; Norris, Brown, Hudson, & Yoshioka, 1998) was utilized to explore the target aviation English tasks, task topics, and task typology in the target language use (TLU) situation, including perceived importance and analyzed difficulty of the tasks. The data included initial semi-structured interviews with two expert military air traffic controllers followed by an online task-based needs analysis conducted with 81 non-native English speaking air traffic controllers in a military aviation context in the Republic of Korea. Through the interviews and surveys, the researcher identified authentic aviation English tasks according to the four language skills (reading, listening, speaking, and writing) and task topics in the TLU situation. Among 275 target tasks revealed in the survey, a large number of target tasks overlapped and could be integrated into 40 target tasks over the four language skills. The identified target tasks were further investigated with the numeric value of perceived importance and difficulty according to the four language skills. The analysis of 275 target tasks could also result in the identification of 14 task topics and specific language skills involved with each topic. Based on the identified tasks, the researcher created an aviation English task typology to provide an initial authentic and sound resource for the future development of aviation English tasks, lesson plans, and curricula.

Room 4 (6317)　Part II　(11:45－12:15)

The Appropriateness of Utilizing the Word Family Unit With Japanese University Students

Stuart MCLEAN (Kansai University)
stuart93@me.com

The choice of the lexical unit, that is, what inflectional and derivational forms are included with a word's base form, has major ramifications on the validity of corpora derived wordlists, wordlist based vocabulary tests, and research that utilizes either of these. The research paradigms of incidental learning, and the explicit vocabulary learning of high frequency vocabulary are based on the assumption that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort. Gardner (2007), however, states that if this assumption is incorrect it bears upon the validity of these two research paradigms, and countless numbers of vocabulary studies. Thus, the validity of different lexical units to the different constructs of productive and receptive lexical knowledge were directly and comprehensively investigated. Participants (N = 286) completed a bilingual vocabulary levels instrument and two 102-item instruments measuring students' productive and then receptive knowledge of inflectional and derivational forms of 12 target words. Uniquely, this study includes forms of target words which include not only one (e.g. development) affix but multiple affixes (e.g. standardization, decentralize) from levels 2 to 6 of Nation and Bauer's (1993) affix classification. This research utilizes form and meaning recall, thus avoiding limitations of previous studies which have utilized monolingual and or multiple choice instruments. Data from the inflected and derived form items were included in analysis only if participants first demonstrated knowledge of the base form of the word. Reliability of data collected from each of the three instruments was > .90. In line with previous research, participants' receptive knowledge of affixes was greater than productive knowledge, and mean affix knowledge positively correlated with vocabulary knowledge. Participants' productive and receptive knowledge of inflectional forms, and -er and re- derived forms was almost complete. However, participants' receptive and productive knowledge of all word family derivational forms was incomplete, the combination of affixes was particularly problematic for participants, and the majority of participants failed to demonstrate receptive knowledge of derivational forms including -ized, -ization, and -ibility, and limited knowledge of derivational forms including -able, -ance and -ment. These findings question the appropriateness of the word family with Japanese university students. This questions the validity of inferences made following investigations which directly or indirectly utilized the word family unit with Japanese participants, and it is argued, native speakers of other languages also.

Room 4 (6317)　Part III (13:45－14:15)

Guessing and the Rasch Model

Trevor A. HOLSTER (Fukuoka University)
trevholster@gmail.com
J. W. LAKE (Fukuoka Jogakuin University)

The invariant measurement provided by the Rasch model allows the mapping of person ability to item difficulty, such as in Beglar's (2010) investigation of the Vocabulary Size Test (VST). However, Stewart (2014, p. 279) criticized Beglar's use of the Rasch model because "guessing does interfere with direct inferences made between the raw scores on such tests and true L2 vocabulary size", noted that Rasch analysis of random numbers showed good fit statistics, and suggested the use of 3-parameter IRT (3PLIRT) was needed to investigate guessing.

This study investigated two research questions concerning multiple-choice vocabulary tests: whether Rasch analysis can identify problematic patterns of random guessing, and whether lucky guessing interferes with the direct mapping of raw scores to logit measures of person ability and item difficulty. RQ1 was investigated by comparing the fit statistics, reliability coefficients, and dimensionality of a random dataset, a real dataset, and a combined dataset. RQ2 was investigated by comparing patterns of changed responses between a pre-test and a post-test. Study 1 sampled 250 female Japanese university students and 250 randomly generated response sets. Study 2 sampled 323 Japanese university students.

Study 1 found that random response sets caused serious data-model misfit, making large-scale random guessing easily detectable. Study 2 showed that correction for lucky guessing resulted in a linear rescaling of both raw scores and logit measures, posing no threat to the direct mapping of raw scores to logit measures. Random guessing did not pose a major threat to interpretation of scores under the Rasch model, but criterion referencing is problematic under 3PLIRT (Chiu & Camilli, 2013). This study empirically demonstrated that random guessing does not invalidate Rasch analysis of thoughtfully developed selected response (i.e. multiple-choice) tests. Rasch fit statistics allowed diagnosis of unexpected response patterns such as random guessing.

From the perspective of measurement, serious misfit is problematic because it indicates distortion of the measurement scale, the Rasch model allows us to identify which responses are misfitting the overall pattern in the data, providing diagnostic analysis such as Engelhard's (2009) investigation of students with disabilities. For classroom level diagnosis, Winsteps provides "Kidmap" results for individual students. One possibility for future studies would be to explore the diagnostic uses of Kidmaps.

This research was supported by grant-in-aid for scientific research from the Japan Society for the Promotion of Science and the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) (Kakenhi grant #25370643).

Room 4 (6317)　Part IV (14:20－14:50)

Are Residuals Important When Comparing Yes-No Test Scoring Formulas?

Raymond STUBBE (Kyushu Sangyo University)
raymondstubbe@gmail.com

Yes-no tests are notorious for containing lexical over-estimation, where test-takers signal knowledge of items which they do not know a correct meaning for (Anderson & Freebody, 1983; Meara & Buxton, 1987; Eyckmans, 2004; and others). To account for this over-estimation weakness a number of yes-no test scoring formulas have been developed. A number of studies have been performed comparing these various scoring formulae (Huibregtse, Admiraal & Meara, 2002; Eyckmans, 2004; Mochida & Harrington, 2006; Pellicer-Sánchez & Schmitt, 2012; Stubbe, 2012). Huibregtse, Admiraal & Meara (2002), comparing h-f, cfg *m and Isdt, found that Isdt was the best formula. Mochida & Harrington (2006) and Stubbe (2012) compared the same four formulae and reported that Isdt was the best scoring formula, based on correlations with the multiple-choice (MC) criterion measure also used in those studies as well as the closeness of the scoring formula mean with the MC mean. Eyckmans (2004), however, found that cfg was the best scoring formula, when a translation test (L2 to L1) was used as the criterion measure.

Although closeness of the means and correlations with the criterion measure can be informative, a better approach may be to analyze the residuals, which are the differences between each adjusted yes-no score and the matching criterion measure score, on a student by student basis. The purpose of the present study is to compare the four above scoring formulae to determine whether or not residuals are also important when evaluating yes-no scoring formulae.

Low-level Japanese university students enrolled in compulsory English classes (n = 455) took a yes-no test of 96 real words and 32 pseudowords, followed by an L2 to L1 translation test of the same 96 real words. The four adjusting formulae, h-f, cfg, *m and Isdt, were first calculated and then compared. Looking solely at closeness of the means to the translation test, it appears as if cfg is the best formula as it had the smallest difference with the translation test mean (6.35 words), whereas the next closest difference was h-f (15.23 words). However, the correlations with the translation test were much lower for cfg compared to h-f (.739 versus .833, respectively). From these two analyses it is difficult to determine the better formula. Once residuals were calculated (20.38 and 17.90 words, respectively), it became obvious that h-f was the superior scoring formula. Thus it was concluded that residuals are important when comparing yes-no test scoring formulas.

Room 5 (6318)　Part I (11:10－11:40)

Initial Evidence on the Validity of the Assessment of the Junior High School Three-Person Discussion

Hidetoshi SAITO (Ibaraki University)
cldwtr@mx.ibaraki.ac.jp

This paper examines evidence concerning the validity of the new assessment of a three-person discussion that is used in the Interactive English Forum (IEF) in Ibaraki prefecture—a discussion contest among junior high school students.

The initial driving force behind this contest—and its assessment—was to transform what were then grammar centered classrooms into more communicative ones. Teachers were encouraged to help students, with sufficient English ability, to enter the contest.

The original rating scale, which had been used over a decade, was intuitively constructed and lacked descriptors. Previous studies identified several problems in the original IEF assessment such as disordering of scale, nonuse of categories, and user dissatisfaction.

This paper discusses four pieces of evidence about the validity of the newly developed IEF assessment:

- content evidence,
- internal evidence,
- criterion-related evidence, and
- consequential evidence.

Central in the content evidence were the identification of the level of description by primary users (i.e., the teachers), and category borders were determined by the combination of examining descriptor content and Rasch measures. Trained raters used the new assessment for the 2014 contest. The data obtained during this contest provided internal evidence that no items or raters were misfits. The three items in the assessment, however, were still found to be inadequate to cover the full range of person abilities, resulting in low Rasch reliabilities for Second (n = 36), and Third Year (n = 36), Students' Sections but not for the B Section (n = 44) — the free entry section with participants of the widest ability range. For criterion-related evidence, 26 students in the Second Year Section volunteered to participate in the TSST (a monologue-based telephone speaking test), and the correlation between TSST scores and the new IEF assessment, r = .58, further supports the validity of the new assessment. For consequential evidence, questionnaire responses from trained raters of the 2014 prefectural contest were compared with those from raters of the 2012 prefectural contest. A positive shift in the responses from the two rater groups indicates the increased usefulness of the new version. Junior high school teachers also responded to another questionnaire of the utility of the new assessment, which also generated positive reactions. Developing and using the new assessment thus benefits both the teachers and students across the prefecture.

This paper discusses directions for gathering further evidence to support or to prompt a revision of the new assessment.

Room 5 (6318)　Part II (11:45－12:15)

With versus *Without*: Comparing the Effects of L1 Translation in Multiple-Choice Spelling Test Formats

Sachiyo TAKANAMI (Toyo University)
sachiyotakanami@gmail.com

Spelling knowledge is not just the ability to write correctly spelled words. In English speaking countries, writing with correct spelling is a reflection of good education. However, in countries where English is a foreign language (EFL), expectations regarding correct spelling of English words may be quite different.

The purpose of this study is to examine the effects of L1 (Japanese) translation in a multiple-choice (MC) spelling test. Multiple choice spelling tests are one of ten types of spelling tests that include six recall tasks and four recognition tasks. Among test formats, those using recall tasks were found to be effective in evaluating learners' spelling ability. However, some test formats using recognition tasks seemed too easy for the learners. Only MC spelling tests were difficult enough to discriminate between learners according to their spelling ability. In previous studies, MC spelling tests did not include L1 translation (or definitions of the target words), because the studies targeted native speakers of English. However, the effects of adding L1 (Japanese) translation in an MC spelling test needed to be examined. Thus, this study is a follow-up study of MC spelling tests. The distractors were derived from the results of previous spelling tests. The following are the examples of the two test formats: (a) without L1 translation (e.g., *alredy, already, arleady, allready*), and (b) with L1 translation (e.g., 〔すでに〕 *alredy, already, arleady, allready*).

A total of 33 words were included in the two test formats. The participants were non-English majors who had taken various types of spelling tests before the study (N = 20). The participants were divided into two groups based on their estimated vocabulary size (i.e., upper and lower). In order to compare the test scores from the two test formats (*With* or *Without* L1 translation), two-way mixed ANOVA was conducted (Group x Test Formats). The result showed that there was no significant interaction, nor main effect of each factor. Added to this, the correlation between the two tests was quite low. However, the results of the item analysis (using Classical Item and Test Analysis Spreadsheet or CITAS) showed some intriguing results. Although, the reliabilities of two tests were not so high, the *With* format of the MC spelling test showed slightly more sufficient results compared to the *Without* test format. Moreover, seeing the results, not all the items in the MC spelling test needed to be deleted. However, due to the restricted number of participants and test items, the results of this study only succeeded in showing some examples. Further study is needed to develop more appropriate test formats for MC spelling tests.

Part V: 賛助会員発表 (Institutional Member Presentations) (14:55－15:25)

ライティング指導ツール Criterion®のご紹介　（教室 1・6302）

山口 学（国際教育交換協議会（CIEE）日本代表部）
myamaguchi@cieej.or.jp

Criterion®は学生が提出したエッセイを分析し、わずか数秒でスコアとフィードバックを提供することで時間のかかるライティング指導を効率化するオンラインツールです。Criterion の採点エンジンである E-rater®は TOEFL iBT®テストでも利用されています。本発表では Criterion のデモおよび国内での導入状況・事例をご紹介します。

実用技能英語検定試験の文法問題の難易度に関する基礎調査
（教室 2・6309）

仲村圭太（公益財団法人　日本英語検定協会）
ke-nakamura@eiken.or.jp
森田真之介（公益財団法人　日本英語検定協会）
鈴木知花（公益財団法人　日本英語検定協会）
金子雅道（公益財団法人　日本英語検定協会）
濵野謙太（公益財団法人　日本英語検定協会）

実用英語技能検定（英検）においては、2 級以下の短文の語句・空所補充セクションにおいて、語い・文法力を測定する項目を設けている。本発表においては 2014 年度に出題された文法力を測定する項目に焦点を当てて、ラッシュモデルにより垂直等化された文法項目の難易度を集計し、各文法項目間の難易度の差の調査結果を共有する。

TOEFL Junior Comprehensive～TOEFL iBT に直結する　世界共通４技能型コンピュータベーステスト　大学入試でも活用～
（教室 3・6310）

小椋茂生（グローバル・コミュニケーション&テスティング）
ogura.sh@gc-t.jp

2014 年秋にスタートした TOEFL Junior Comprehensive は英語４技能およびこれらの技能の統合型技能を多面的に測定。新指導要領でも求められている統合型スキルや多様な言語活動におけるアウトプットを測定します。測定領域は日本の高校生の英語レベルに最適な CEFR の A2～B2。大学入試での活用も始まっています。

TOEFL iBT に直結する世界共通４技能 CBT の内容を詳しくご紹介します。

英語スピーキング能力テスト TSST のご紹介

（教室 4・6317）

福江友樹（株式会社アルク）
tfukue@alc.co.jp
平野琢也（株式会社アルク）

アルクは、電話を使った英語スピーキング能力試験 Telephone Standard Speaking Test（TSST）を開発し、2004 年から運用しています。本発表では TSST の概要（受験方法、質問項目、評価官の品質管理、結果公開等）のほか、主に TSST の４つの評価基準についてご説明いたします。TSST の質問はデータベースからランダムに抽出された 10 問で構成され、１問の回答時間は 45 秒です。発話の評価は訓練を受けた 3 人が独立してあたります。様々な評価基準がある中、なぜ TSST は特定の評価基準を採用しているか、何をもって英語が話せると判断しうるか、訓練を受けた人が評価することの意味などについてお話いたします。
（2014 年までの実績データも紹介致します）

オンラインテスト『Progress』の実践報告

（教室 5・6318）

上村 武司（ピアソン・ジャパン株式会社）
takeshi.kamimura@pearson.com

本発表では、本年度よりProgressを導入した大学の一例として、摂南大学での実践を報告する。

Progressは、ピアソン社が開発した英語の上達度を測定するオンラインテストであり、４技能ならびに文法、語彙の計６分野を測定できる。また、本テストは CEFR＜A１～C１の範囲で全６レベルに分かれており、学習者に適したレベルのテストを受験させることができる。さらに、Progress は１セットに３回のテストを含み、事前、事後テストに加えて中間テストを実施することが可能である。そのため、学習者の英語が着実に上達しているか、またカリキュラム内容に改善点がないか等をより詳しいデータに基づいて確認することもできる。また、テストは通常５分以内に自動採点される。結果は、総合力（Overall）に加えてスキル別（G・S・V・R・W・L）の６分野で、ピアソン社が独自に開発した指標であるGSE（Global Scale of English）および Can-doリストによって表示される。

摂南大学では今年度より、外国語学部の１、２年次生を対象に年２回程度（６月および１月）の実施（事前・中間・事後の計３回）を予定しており、そのうち事前テストが実施された。参加者は１年次生１７６名で、CALL 教室において１時限３５～４０名ずつ、計５時限分で実施された。

テスト結果の分析から、他の外部テストよりも（1）（６つの観点から英語能力を測定しているため）より詳しく能力を測定できること、そして（2）これまでは細かく測定できなかった低学力層のスキル差を、Progress ではより明確に測定することができたことがわかった。また、（3）参加者たちに対して教育的な波及効果があることも明らかとなった。

改善点としては、（1）ICT リテラシーの低い参加者に対して十分な事前練習の機会が必要であること、また （2）テスト時の動作安定性をより担保する必要性があること、などが確認された。

4. Workshop Information（ワークショップ情報）

Workshop 1（ワークショップ 1）

Title: “You’ve done your test—now what? First steps in test analysis”

(Conducted in English)

Lecturer: Myles GROGAN (Kansai University)
Atsushi MIZUMOTO (Kansai University)
Chair: Akiyo HIRAI (University of Tsukuba)

- Date: September 5, 2015 (Saturday), 13:00－17:00
- Venue: Chuo University, Korakuen Campus (Room 6401 in Building 6)
- Attendance Fee: 1,000 yen
- Max Capacity: 48 (first-come, first-served basis)
- Prerequisite: Familiarity with MS Excel or a similar spreadsheet is helpful, but not required. Participants will receive data sets taken from tests with single-point test items (discrete items), and will work through examples using Excel and some online tools. Participants may wish to bring data sets of their own in a spreadsheet format, and this may be beneficial for participants. The goal of the course is to look at norm-referenced tests in general.

- Aim

1.To understand basics of Classical Test Theory, and introduce some alternatives such as IRT.
2.To get hands-on experience calculating reliability indices such as Cronbach Alpha or the Standard Error of Measurement.
3.To gain a foundation in testing approaches, enabling the discussion of tests and test design for teaching and for syllabus design.

- Procedure

1. Lecture 1: Getting data - Basic tips and tricks
2. Hands-on workshop 1: Reliability
3. Hands-on workshop 2: Looking at test items (Item Facility and Discrimination)
4. Hands-on workshop 3: Reliability measures and other calculations
5. Lecture 2: Alternatives to Classical Test Theory (particularly IRT)

- How to register

1. The deadline of the registration is Saturday, August 29. (Note: If the workshop does not reach the maximum capacity, the registration on the day of the workshop conducted is allowed.)
2. When you register, please provide the information below and email it to Mr. Yuichiro YOKOUCHI (University of Aizu) at u16yoko@gmail.com [Note: If you write your questions in (5) below, the lecturers may be able to answer them during the workshop.]

Let us know the following information when you register the workshop.

(1) Your name, affiliation, and email address.
(2) What kind of tests do you usually use (classroom tests, coursebook test, placement tests)?
(3) What kind of reflection do you usually do with your grades or test results? (Check individual student performance, discuss with colleagues, check items? etc.)
(4) How do you hope to use your knowledge of testing in the future?
(5) Questions to the lectures, if you have. (Optional)
(6) Request to this workshop, or JLTA workshops in general. (Optional)

題目：「テストデータ分析入門」（英語で実施）

講　師　Myles GROGAN（関西大学）
水本 篤（関西大学）
司　会　平井 明代（筑波大学）

日時：　2015 年 9 月 5 日（土）13:00－17:00
場所：　中央大学後楽園キャンパス（6 号館 6401 教室）
参加費：　1,000 円
定員：　48 名（申し込み順）
参加条件：　Microsoft Excel（やその他の表計算ソフト）を使ったことがある（必須ではありません）。参加者には演習形式で分析を行うデータが当日配布され、Excel やオンライン分析ツールで実際に分析を行っていただきます。自分のデータを持ち込んでいただいても結構です。今回のワークショップでは集団準拠型テスト（norm-referenced tests）を対象とします。

目的：
1. 古典的テスト理論の基礎を理解し、項目応答理論について概観する。
2. クロンバック・アルファや測定の標準誤差などを実際に計算してみる。
3. 言語テスティングについての理解を深め、指導やシラバス作成にその知見を反映する。

手順：
1. 講義 1：データ収集と整理
2. 演習 1: 信頼性
3. 演習 2: 項目分析
4. 演習 3: 信頼性係数とその他の指標の算出
5. 講義 2: 古典的テスト理論以外の方法（項目応答理論など）

申し込み方法：
1. 申し込みの締切は、8 月 29 日（土）です。（定員に達していない場合には当日参加も可能です。）
2. 申し込み時に、下記の情報を横内 裕一郎（会津大学）u16yoko@gmail.com まで e-mail でご連絡ください。

(1) 氏名・所属・e メールアドレス
(2) 普段、どのようなテストを実施していますか。（授業内でのテスト、プレースメントテストなど）
(3) 成績をつけたり、テストの結果を見て、どのような振り返りを行っていますか。（学習者個人レベルでの確認、同僚との話し合い、項目の分析など）
(4) テスティングの知識を将来どのように活用しようと考えていますか。
(5) 講師へのご質問（希望者のみ）
(6) その他、ワークショップまたは JLTA ワークショップ全体に対して何かご要望がありましたらお書きください。（希望者のみ）

Workshop 2（ワークショップ 2）

題目：「生徒の力を伸ばす定期テストの作り方―妥当性と信頼性に留意して」（日本語で実施）

講　師　笠原 究（北海道教育大学）
佐藤 臨太郎（奈良教育大学）
司　会　平井 明代（筑波大学）

日時：　2015 年 9 月 6 日（日）11:10－15:10（12:00－13:00 のお昼休憩 1 時間を含みます）
場所：　中央大学後楽園キャンパス（6 号館 6401 教室）
参加費：　1,000 円
定員：　48 名（申し込み順）
参加条件：　中高の英語教員で、定期考査問題をより良いコミュニケーション能力測定のツールとしたいとお考えの方。定期考査によって生徒の英語力を向上させたいとお考えの方。テスティングの知識がある・なしを問いません。勤務校でご自身が作成された定期考査問題か、あるいは、勤務校で使用された定期考査問題を持参してください。

目的：
4. テスト問題作成において、知っておくべき基本的な原理原則を理解する。
5. 英語によるコミュニケーション能力を測定するためにはどうすべきかを考える。
6. 様々なテストの問題形式を概観し分析する。
7. 実際にテスト問題作成を行い、議論討議をする。

手順：
6. 講義（テスティングの基本的理論：定期テストを作成する上で押さえておくべきポイントを確認しましょう。）
7. 意見交換等（テスト問題の分析等：大学の学生が作成した問題を検討します。）
8. テスト問題の製作実習（上記 1, 2 を踏まえて、あらためて問題を作成してみましょう。）

申し込み方法：
1. 申し込みの締切は、8 月 29 日（土）です。（定員に達していない場合には当日参加も可能です。）
2. 申し込み時に、下記の情報を横内 裕一郎（会津大学）u16yoko@gmail.com まで e-mail でご連絡ください。

(1) 氏名・所属・e メールアドレス
(2) テスティングの理論に関して過去に学ばれたことはありますか。
　　全くない　　少しある　　ある　　相当学んだ
(3) 定期考査作成において日頃、困難に感じていることは何ですか。
(4) 講師への質問。（希望者のみ）
(5) その他、ワークショップまたは JLTA ワークショップ全体に対して何かご要望がありましたらお書きください。（希望者のみ）

Workshop 2

Title: "How to develop tests that improve students' English proficiency: Focus on validity and reliability" (Conducted in Japanese)

Lecturers: Kiwamu KASAHARA (Hokkaido University of Education)
Rintaro SATO (Nara University of Education)
Chair: Akiyo HIRAI (University of Tsukuba)

- Date: September 6, 2015 (Saturday), 11:10－15:10 (including lunchtime from 12:00-13:00)
- Venue: Chuo University, Korakuen Campus (Room 6401 in Building 6)
- Attendance Fee: 1,000 yen
- Max Capacity: 48 (first-come, first-served basis)

Prerequisite: Junior or senior high school teachers who want to improve the quality of the term tests to measure students' commutative abilities and those who want to improve students' English proficiencies through the term tests. No knowledge on testing theories is required. Participants are required to bring the term test they made or used in their schools.

- Aim

1. To understand basic principles in creating tests.
2. To consider how teachers should measure students' communicative abilities in English.
3. To survey and analyse a various types of question formats.
4. To make tests and have a discussion on them.

- Procedure

1. Lecture (e.g., theories of testing)
2. Discussion (e.g., analysis of the tests made by university students)
3. Practice (e.g., creating a test)

- How to register

1. The deadline of the registration is Saturday, August 29. (Note: If the workshop does not reach the maximum capacity, the registration on the day of the workshop conducted is allowed.)
2. When you register, please provide the information below and email it to Mr. Yuichiro YOKOUCHI (University of Aizu) at u16yoko@gmail.com [Note: If you write your questions in (5) below, the lecturers may be able to answer them during the workshop.]

Let us know the following information when you register the workshop.

(1) Your name, affiliation, and email address.

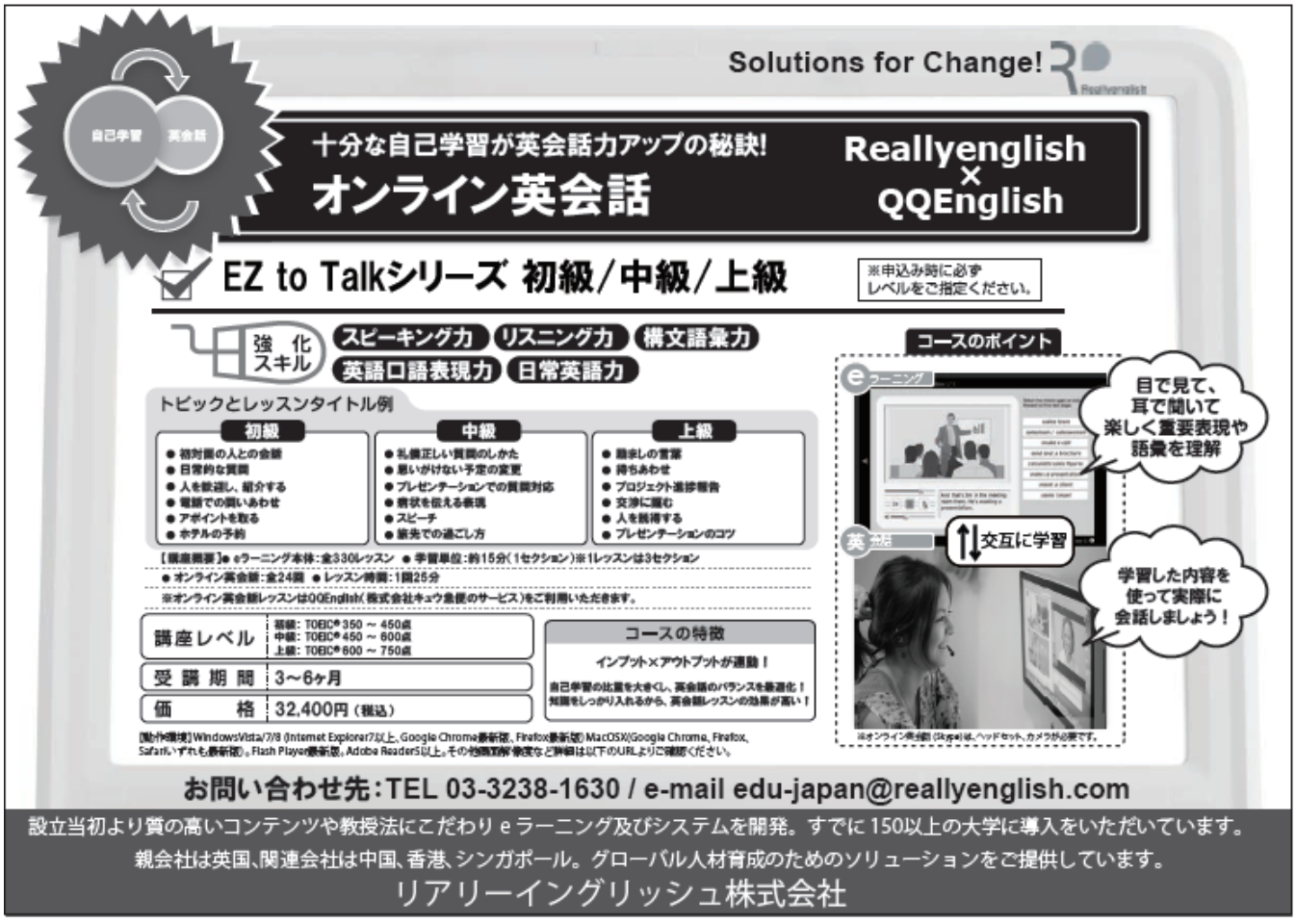
(2) Have you ever learned about the theories in testing?

Never a little some a lot

(3) What difficulty do you find in making the term test?

(4) Questions to the lectures, if you have. (Optional)

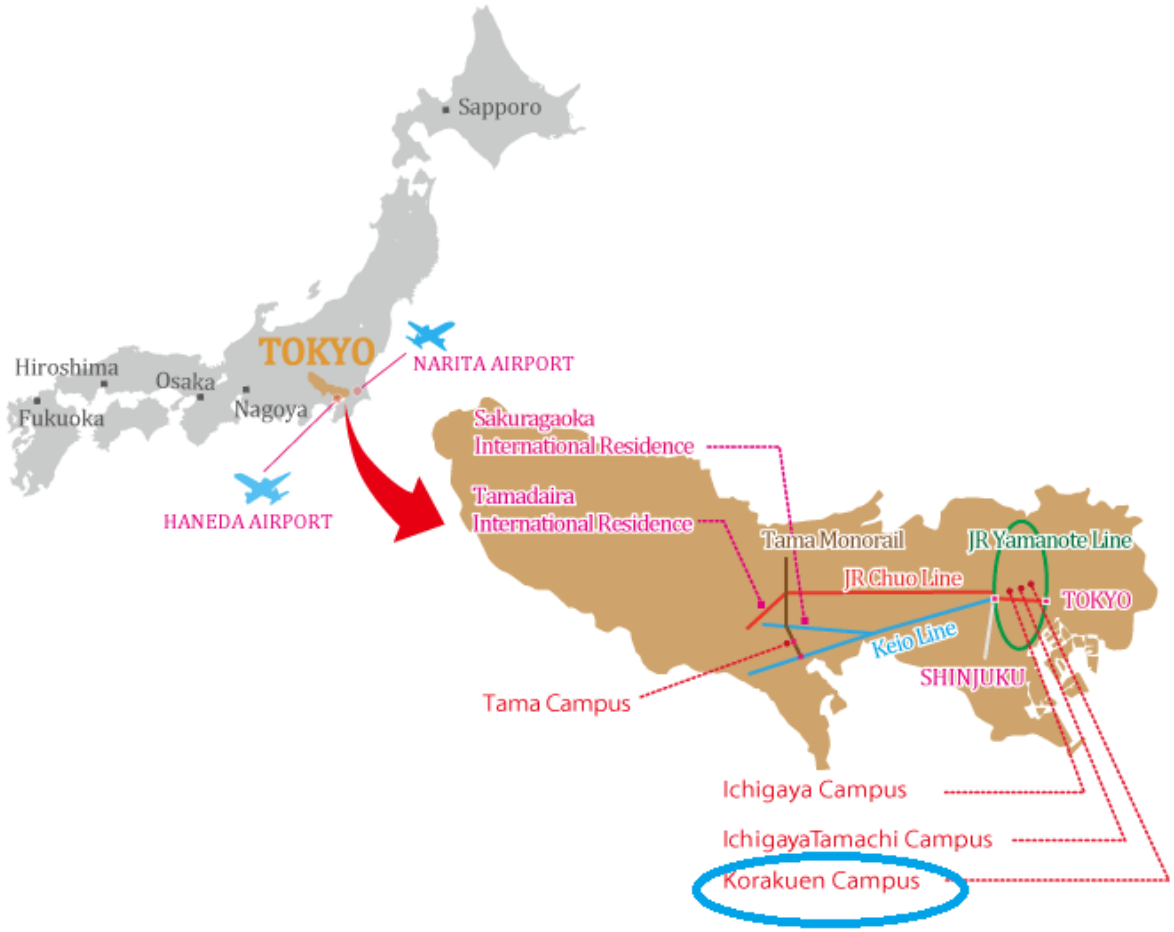
(5) Request to this workshop, or JLTA workshops in general. (Optional)



5. Access to the Conference Venue (Chuo University Korakuen Campus)

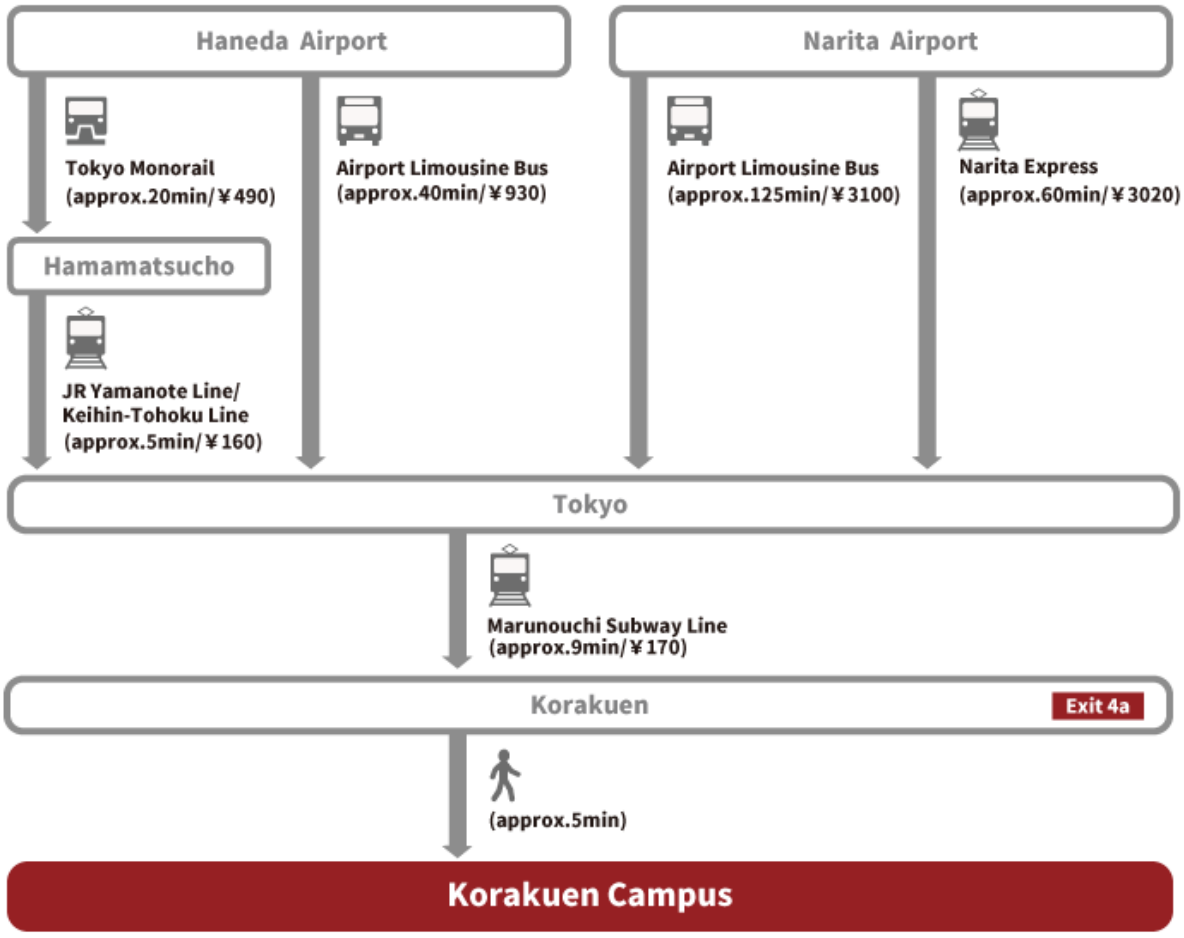
Address (住所)：1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan （〒112-8551 東京都文京区春日1-13-27）

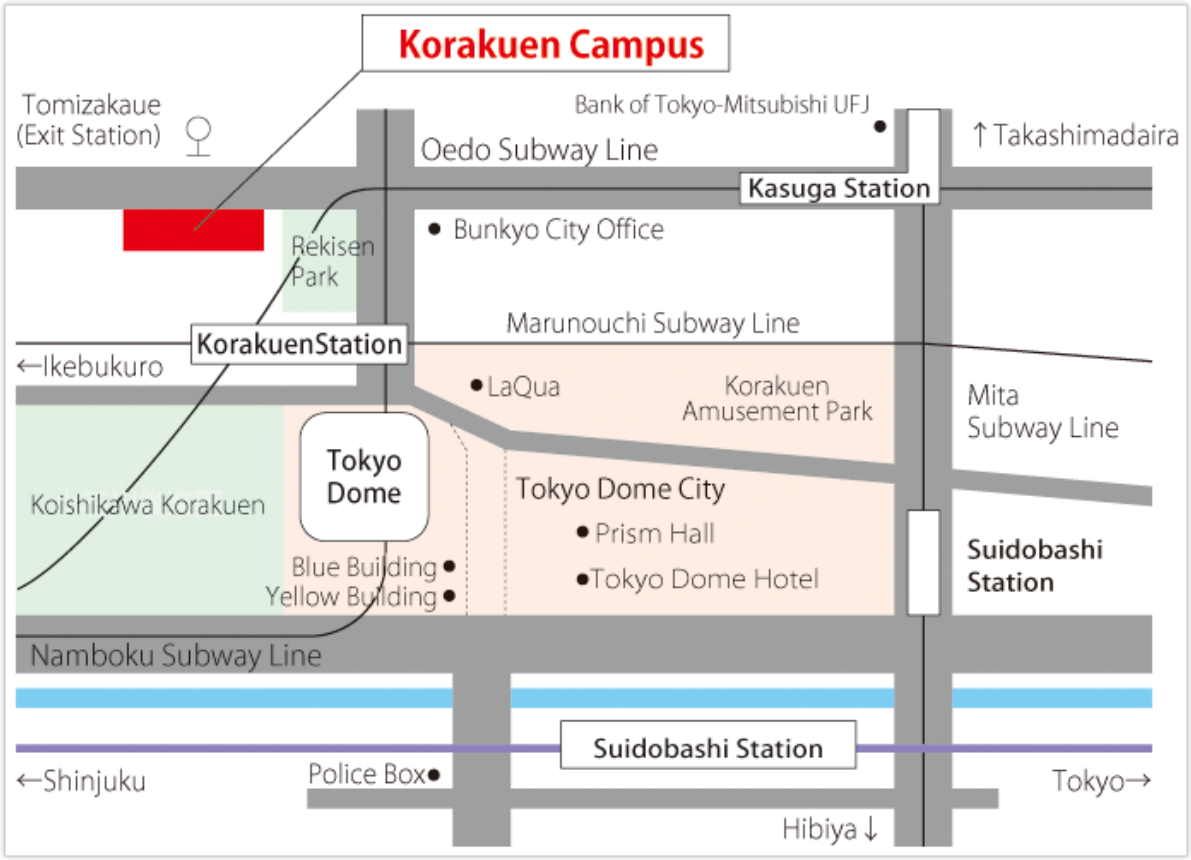
For visitors from overseas



Access from International Airports

(fares as of June 2015)





Closest Stations to the conference venue
5-minute walk from Korakuen Station on the Marunouchi and Namboku Subway Lines
7-minute walk from Kasuga Station on the Oedo and Mita Subway Lines
12-minute walk from Suidobashi Station on the JR Sobu Line
17-minute walk from Iidabashi Station on the JR Sobu Line

For visitors from within Japan (国内からの参加者の方々)
Please see the above map from Korakuen station or the one below in Japanese. (上記の地図もしくは以下をご覧ください。)

東京メトロ丸ノ内線・南北線『後楽園駅』から徒歩 5 分
都営三田線・大江戸線『春日駅』から徒歩 6 分
JR 中央・総武線『水道橋駅』から徒歩 12 分
JR 中央・総武線『飯田橋駅』から徒歩 17 分

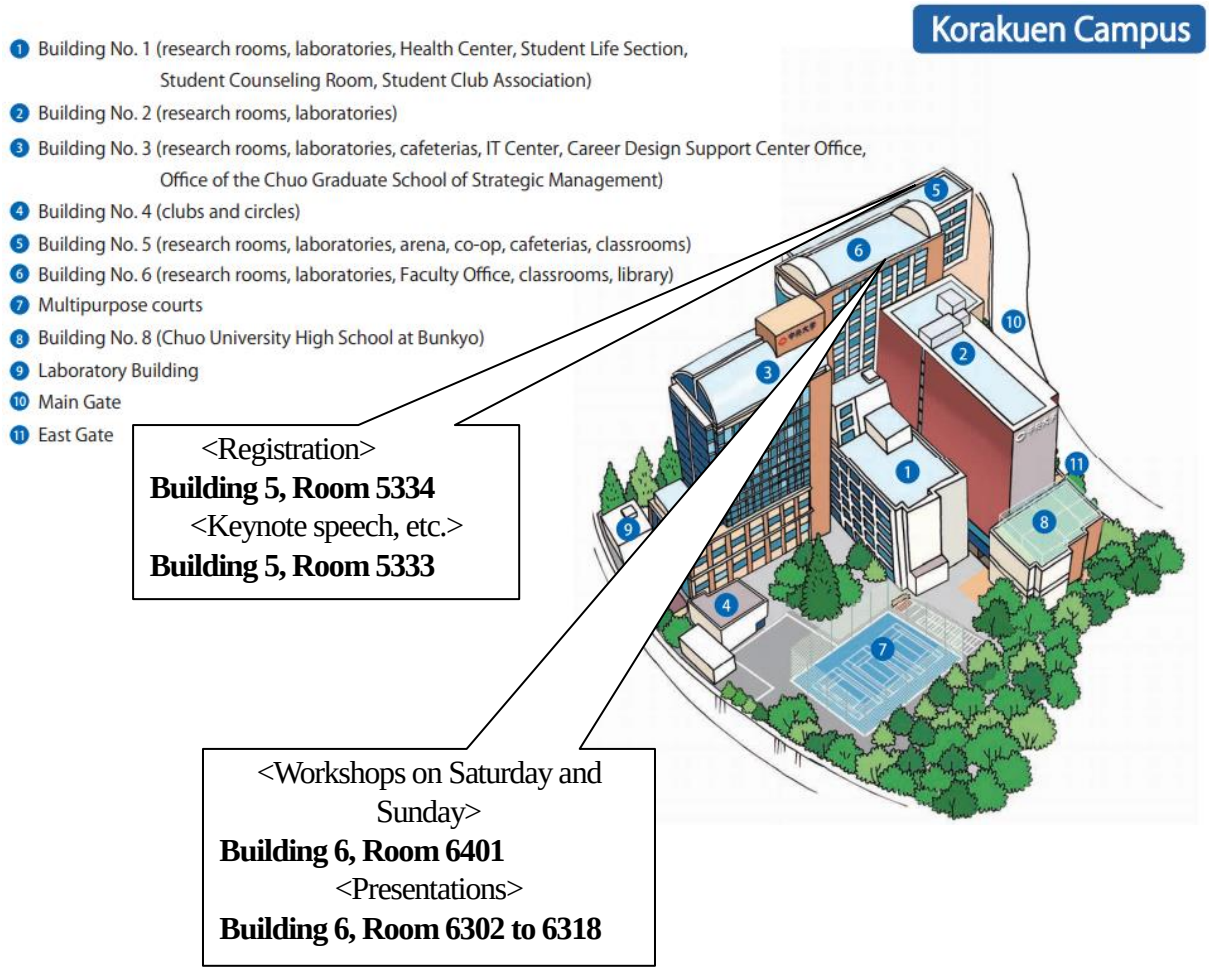
Convenience store and banquet place

Lawson convenience store

You are here.

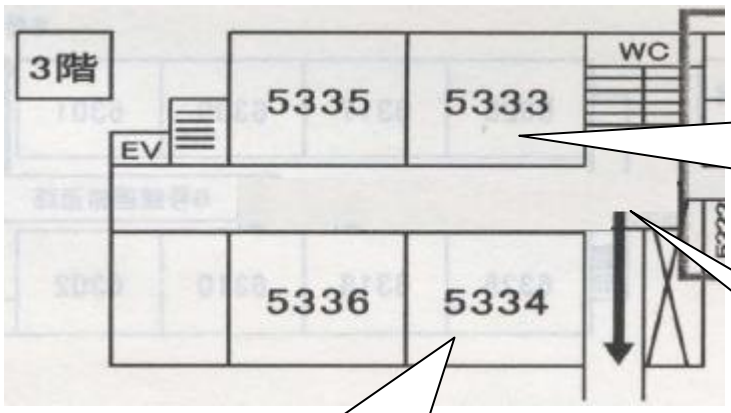
Restaurants around here. FOUGOU—our banquet place—is here.

Campus map



Floor Plan

Building 5, 3rd floor (5 号館 3F)



5333
Opening Ceremony
Plenary Speech
Invited Lecture
Symposium
Closing Ceremony
JLTA Business Meeting

5334
Registration Committee Meeting
Headquarters

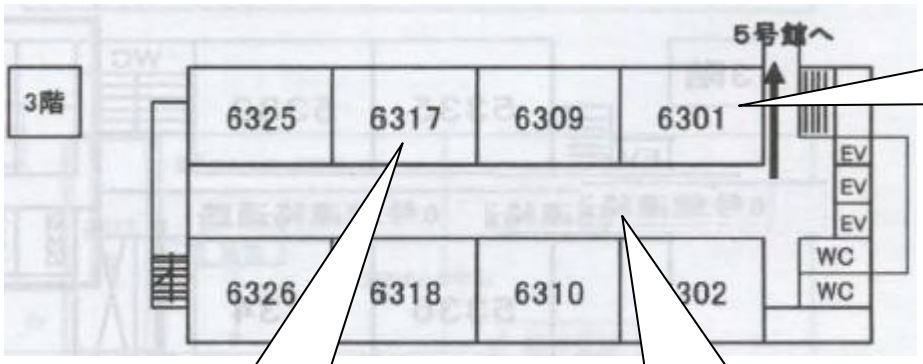
To the Building 6

Hallway (3 階廊下)

Exhibits/協賛企業 展示

Building 6, 3rd floor (6 号館 3F)

To Building 5



6301
Participant
Lounge
Refreshment

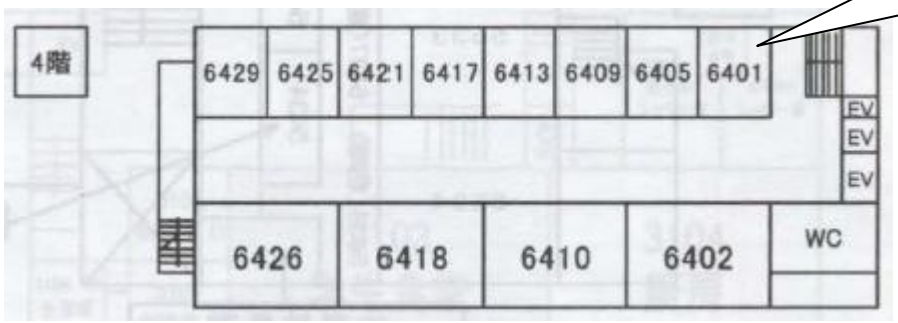
6302, 6309, 6310, 6317, 6318
Presentation

Hallway (3 階廊下)

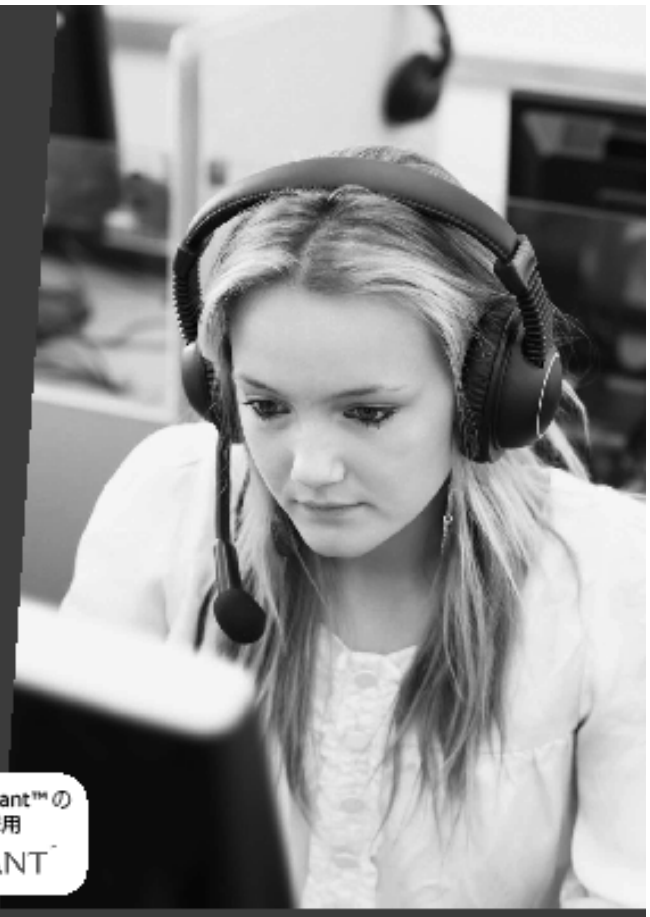
Exhibits/協賛企業 展示

6401
Workshop 1 on Saturday &
Workshop 2 on Sunday

Building 6, 4th floor (6 号館 4F)



Progress



学生の熟達度が明確に!
4技能+文法力+語彙力が計測できる
オンライン英語運用能力テスト。

スピーキングにはVersant™のテクノロジーを採用
VERSANT

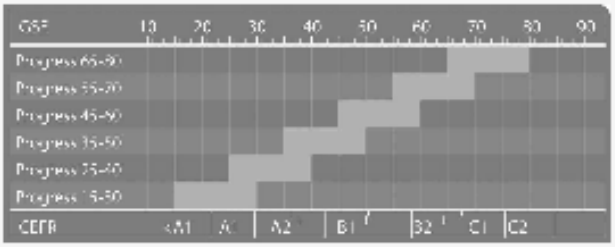
1. スピーキング・ライティングも計測
4技能と文法や語彙力など、総合的な英語運用能力スキルを測定。
技能毎にスコアが測定されるため、日々の学習課題が明確に把握できる。

2. 3回のテストが1セットに
1セットに3回分のテストが含まれているので、
例えばセメスター毎の熟達度が測定可能。

3. スピーディに結果を算出
約60分の試験終了後、自動採点システムにより通常15分以内で
結果を算出。ウェブ上での結果確認が可能。

4. アダプティブ方式を採用
受験者の解答により出題内容が変化。
より正確な英語力が測定できる個別対応型。

［どのテストを受けるか、6つの英語力レベルから選択可能］



*Progress 65-90は2015年秋にリリース予定

［幅広い出題形態］

1. 選択問題
2. 穴埋め問題
3. 並べ替え問題
4. 読解問題
5. 音読問題
6. ディクテーション
7. リピーティング
8. 描写問題
9. 音声認識問題
10. 要約問題
11. 作文・小論文

www.pearsonelt.com/progress

Join the conversation
#testingprogress

ピアソン・ジャパン株式会社
www.pearson.co.jp

〒106-6021
東京都港区六本木1-6-1 泉ガーデンタワー21階
Tel 03-5549-8630　Email: elt.jp@pearson.com

Progressのスコアは、
グローバル・スケール・オブ・イングリッシュに
準拠しています。

詳しくはこちらをご確認ください。
www.pearson.co.jp/company/gse

グローバル・スケール・オブ・イングリッシュ（GSE）とは？
グローバルに比較可能な英語能力の客観的指標で、10から90
のスコアで英語能力を評価するもの。

ALWAYS LEARNING

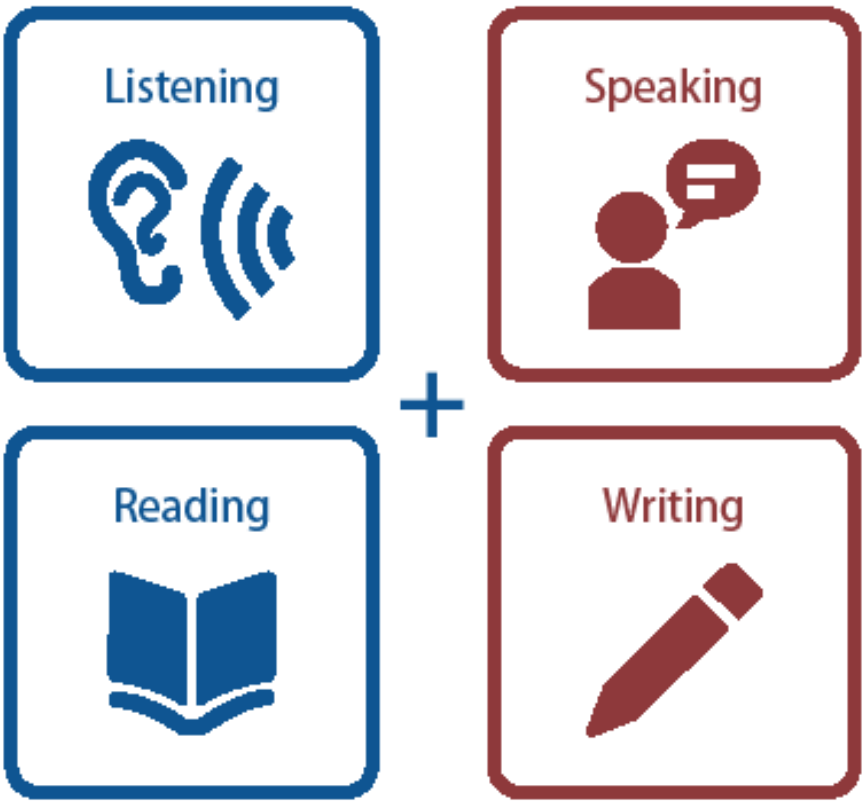
PEARSON

TOEIC® プログラムで英語4技能のコミュニケーション能力を測定

TOEIC®プログラムでは、TOEIC®テストとTOEIC®Speaking&Writing（TOEIC®S&W）により、「聞く」「読む」「話す」「書く」英語4技能を測定できます。

「聞く」「読む」英語力を測るなら
TOEIC®テスト

「話す」「書く」英語力を測るなら
TOEIC® S&W



TOEIC® S&Wは、パソコンを使って音声を吹き込んだり、文章を入力して解答するテストです。TOEIC®テスト（リスニング、リーディング）とは別に実施されます。

活用事例や詳細情報は、TOEIC®Speaking&Writing公式サイトへ ›› http://www.toeic.or.jp/sw/

一般財団法人 国際ビジネスコミュニケーション協会
〒100-0014 東京都千代田区永田町2-14-2 山王グランドビル
TEL:03-5521-5901
FAX:03-5521-5915
最新情報・テスト詳細は公式サイトへ
http://www.toeic.or.jp

ETS, the ETS logo, PROPELL, TOEIC, TOEIC Bridge, TOEIC BRIDGE are registered trademarks of Educational Testing Service in the United States, Japan and other countries and used under license.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Cambridge English は
世界が認める国際資格。

- 国際児童英検から IELTS まで英語コミュニケーション力を測定
- 英国ケンブリッジ大学の非営利機関、ケンブリッジ大学英語検定機構が開発
- 4技能を世界基準（ヨーロッパ言語参照枠：CEFR）で評価
- スピーキングテストはケンブリッジ認定の試験官が対面式で評価
- 世界 20,000 機関以上が英語力の証明として認定
- 毎年 500 万人以上が受験

✶ IELTS は、ケンブリッジ大学英語検定機構、ブリティッシュ・カウンシル、IDP: IELTS Australia のパートナーシップにより共同運営されています。

Cambridge English

www.cambridgeenglish.org/jp

❇ infoJapan@cambridgeenglishreps.org
❇ TEL: 080-5545-8969

Follow us on Facebook
www.facebook.com/
CambridgeEnglishEastAsia

英語4技能評価の理論と実践

望月昭彦、深澤 真、
印南 洋、小泉利恵◉編著

◉A5判·320頁
定価＝本体2,400円＋税

CAN-DO・観点別評価から技能統合的活動の評価まで

よりよい英語指導のための評価の指針と実践例

CAN-DOリスト導入や観点別評価を含め、学習指導要領に沿った指導に対応する評価が求められています。また、4技能型外部試験の大学入試への活用などを控え、日常の評価だけでなく、幅広く英語評価一般に関する知識も今後ますます必要とされます。

本書は、小学校から大学まで学校教育の実態に即した評価法を、理論・実践両面から取り上げ詳しく解説。

これからの英語評価を考えるための最新のガイドです。

大修館書店　〒113-8541 東京都文京区湯島2-1-1 http://www.taishukan.co.jp　販売部電話 03-3868-2651

Commercial Exhibits（展示協賛企業　50 音順）

一般財団法人国際ビジネスコミュニケーション協会
The Institute for International Business Communication (IIBC)
http://www.toeic.or.jp/iibc.html

株式会社 教育測定研究所
The Japan Institute for Educational Measurement, Inc.
http://www.jiem.co.jp/

ケンブリッジ大学出版局
Cambridge English
http://www.cambridge.org

ピアソン・ジャパン株式会社
Pearson Japan
http://www.pearson.co.jp/

Commercial exhibits are located in the hallway on the 3rd floor of Building 5 and 6. Please ask staff at the registration desk for details.
展示は会場 3 階の廊下で実施いたしております。詳しくは受付でお尋ねください。

We would like to acknowledge the Eiken Foundation of Japan and JSPS KAKENHI (Grant Number 26370737, principal investigator: Rie KOIZUMI) for helping to make Dr. Jessica R. W. WU's Keynote speech and Dr. Carol A CHAPELLE's invited lecture possible at the annual conference of the Japan Language Testing Association. Our special gratitude also goes to Chuo University for making its Korakuen Campus available as the venue for the 19th Annual Conference of the Japan Language Testing Association.
Jessica R. W. WU 先生の基調講演の実現にあたりましては、日本英語検定協会様より、Carol A CHAPELLE 先生の招待講演の実現にあたりましては、小泉利恵先生の科学研究費補助金, 基盤研究(C) (No. 26370737) よりご支援・ご協力を賜りました。また、第 19 回日本言語テスト学会全国研究大会の実現にあたり、中央大学より、会場の提供を初め、ご支援・ご協力を賜りました。
誠にありがとうございました。

The next year's annual conference will be held in autumn 2016 at Tokai University in Tokyo. The conference schedule will be announced via the JLTA website as soon as the details become available. We look forward to seeing you there.
2016 年度の日本言語テスト学会全国研究大会は、2016 (平成 28) 年秋に東海大学で行われます。詳細が決まり次第、JLTA のホームページでお知らせいたします。ご参加のほどよろしくお願いいたします。

日本言語テスト学会 (JLTA) 第 19 回 (2015 年度) 全国研究大会発表要綱
Handbook of the 19th Annual Conference of the Japan Language Testing Association
発行日：2015 年 8 月 1 日
発行：日本言語テスト学会 (JLTA)　　会　長：渡部良典 (上智大学)
事務局：〒270-1695　千葉県印西市平賀学園台1-1　順天堂大学
さくらキャンパス　小泉利恵研究室　　TEL: 0476-98-1001 (代表)
FAX: 0476-98-1011 (代表)　　E-mail:　rie-koizumi@mwa.biglobe.ne.jp
編集：全国研究大会実行委員会
印刷：株式会社国際文献社　　〒162-0801 東京都新宿区山吹町358-5
TEL: 03-5937-0249　　FAX: 03-3368-2822