

外国語教育評価学会 (JLTA)
第4回 全国研究大会 (2000年度) プログラム

The Fourth Annual Conference
of
The Japan Language Testing Association

大会テーマ : 「コンピュータ使用による言語テストの理論と実践」
Computerized Language Testing: Theory and Practice

日時 : 2000年10月8日 (日) 8:30~18:00

会場 : 東京経済大学6号館

(〒185-8502 東京都国分寺市南町1-7-34)

外国語教育評価学会

The Japan Language Testing Association (JLTA)

事務局

〒389-0813 長野県埴科郡戸倉町芝原758

TEL 026-275-1964 FAX 026-275-1970

e-mail: youichi@avis.ne.jp

URL: <http://www.avis.ne.jp/~youichi/JLTA.html>

全 国 研 究 大 会 運 營 委 員

大 友 賢 二 (常 磐 大 学)

Randolph Thrasher (国 際 基 督 教 大 学)

中 村 洋 一 (常 磐 大 学)

中 村 優 治 (東 京 經 済 大 学)

第 4 回 大 会 プ ロ グ ラ ム

8:30～ 受 付 (6号館1階)

8:50～9:00 開会の挨拶 (7階大会議室) 総合司会 中村 優治 (東京経済大学)
会長 大友 賢二 (常磐大学)

9:00～10:20 研究発表 (発表30分, 質疑10分) I 9:00～9:40

II 9:40～10:20

第1室 (7階中会議室1) 司会 Steven Ross (関西学院大学)

発表 I Adapting the Levels Test into a Japanese Translation Format

Yoshimitsu Kudo (Rikkyo University)

発表 II What a Computerized Reading Test Can, but a Pencil-and-Paper Reading
Test Cannot, Do

Shizuka Tetsuhito (Institute of Foreign Language
and Education Research, Kansai University)

第2室 (7階中会議室2) 司会 望月 昭彦 (筑波大学)

発表 I Modifying an Existing Test to Allow it to be Delivered by Computer

Randolph Thrasher (国際基督教大学)

発表 II Peer Ratings in EFL Writing Course: User Acceptance and Characteristics

Hidetoshi Saito (Hokusei Gakuen University)

Tomoko Fujita (Rikkyo University)

第3室 (7階中会議室3) 司会 山崎 朝子 (武蔵工業大学)

発表 I 「英語能力簡易測定のための語彙テスト— 開発されたテストの妥当性・信頼性の検証」

片桐 一彦 (東京学芸大学 博士課程)

発表 II 「パブリックスピーキングテストの分析: ラッシュモデルが教えてくれるもの」

中村 優治 (東京経済大学)

10:20～10:30 休 憩

10:30～12:30 基調講演 (7階大会議室) 司会 浪田克之介 (北海道大学)

紹介 会長 大友 賢二 (常磐大学)

「コンピュータテストの理論的背景について」

講師 前川 真一 (大学入試センター)

12:30~13:30 昼 食 (役員会)

13:30~14:10 研究発表 (発表 30 分, 質疑 10 分)

第 1 室 (7 階中会議室 1)

司会 法月 健 (静岡産業大学)

「Aパラメーターから見た言語テストの弁別力」

湯浅 克己 (ILC BETAテストセンター)

第 2 室 (7 階中会議室 2)

司会 飛渡 洋 (国士舘大学非常勤)

「Test Data Analysis Program: TDAP Ver. 1.0E について」

中村 洋一 (常磐大学)

第 3 室 (7 階中会議室 3)

司会 渡部 良典 (国際基督教大学)

Classical and IRT Approaches to System Feedback on Language Testing for University Entrance

Steven Ross (関西学院大学)

14:10~14:20 休憩

14:20~16:00 パネルディスカッション (7階大会議室)

司会 木下 正義 (福岡国際大学)

「コンピュータ使用による言語テスト (CAT/CBT) が外国語教育に与えるもの」

発表者: 峯石 緑 (広島国際大学)

伊藤 彰浩 (愛知学院大学)

智原 哲郎 (大阪女学院短期大学)

根岸 雅史 (東京外国語大学)

16:00~16:30 総 会 (7階大会議室)

司会 中村 優治 (東京経済大学)

議長選出

報告 中村 洋一 (常磐大学)

16:30~16:40 閉会の挨拶 (7階大会議室)

副会長 Randolph Thrasher (国際基督教大学)

16:40~18:00 懇 親 会 (7階ホール)

司会 谷 憲治 (早稲田大学非常勤)

発表要旨 (ABSTRACTS)

基調講演

「コンピュータテストの理論的背景について」

前川 真一 (大学入試センター)

最近、我が国においてもコンピュータを用いたテストが話題となってきた。コンピュータを用いたテストの研究は、項目反応理論の応用として 1970 年代から行われてきたが、実用的な段階に入ったのはごく最近のことである。本講演では、コンピュータを用いてテストを行う際に基本となる項目反応理論の解説と、コンピュータテストの基本的技術の紹介を行う。

パネルディスカッション

「コンピュータ使用による言語テスト (CAT/CBT) が外国語教育に与えるもの」

峯石 緑 (広島国際大学)

伊藤 彰浩 (愛知学院大学)

智原 哲郎 (大阪女学院短期大学)

根岸 雅史 (東京外国語大学)

峯石 緑 (広島国際大学)

①大きく改善されると思われる点 (一般論)

通常、受験のために遠方のテスト開催地に行かなくてはならないことから、地方在住者が被りがちなデメリットがかなり解消され、より多くの受験者が、より整った同様の環境下で受験できるようになる可能性が増すこと。同じ理由から、受験者数が増え、より多くの英語学習者に精度の高い試験受験の機会がもたらされるようになると考えられること。

② 実際上の問題点 (一般論)

コンピュータ・リテラシーが低い (あるいは無い) 受験者は、従来の形式で受験せざるを得ないこと。試験を実施する側の立場からは、コンピュータによる試験と筆記試験を並行して管理していくうえでの困難さ。

③ 私の学校で・私の授業で・私自身に起こりうる変化 (戸惑いと希望)

現在の勤務校が情報・経営系のプログラムを多数有する大学であるため、コンピュータを使用して精度の高い標準テストを行うことについての学生の情意面での反応は、極めて好意的であることが予想され、導入について学生への不安はない。しかし、語学専修の大学ではないせいもあり、CALLの設備が無く、実施上の困難点として、授業の枠の中でテストに備えさせるための機会が持てないことが挙げられる。コンピュータ使用による個別学習が可能であるとしても、元々それほど高い英語力を持っていない学習者には、せめて最初のうちは教師による支援が是非とも必要であると考えられるため、何とか授業の中でも指導できるようにしたいものである。

伊藤 彰浩 (愛知学院大学)

本発表では、発表者のこれまでの研究内容と経験を踏まえ次の3点について論じる。具体例は発表時に示す。

(1) コンピュータを利用した言語テストを用いれば、項目応答理論を用いた英語学力の測定と評価がより容易に実施できるようになる。(2) しかし、コンピュータの整備がなされても、教師と学習者の測定や評価に関する意識がかわらなくてはコンピュータ化されたテストを利用することは混乱を招くことになりかねない。(3) 近年の第2言語習得研究や認知心理学研究でいわれる領域固有性を考えると、言語を用いて具体的な活動が達成される点を重視した評価を遂行することも重要である。コンピュータを用いたテストは、画面を見たり音声を聴いて解答する形式が一般的だと思われるが、言語産出をどの程度測定できるか疑問がのこる。よって、コンピュータ化されたテストとアナログ型の測定・評価法を折衷して利用することが必要なのではないか。

智原 哲郎 (大阪女学院短期大学)

コンピュータによる TOFFL 受験が始まったが、Paper-and-Pencil Test から Computer-Based Test へのシフトは、評価という点から察すれば、新たな問題が出てくるものの、従来抱えていた問題を解決するものが多くなると思われる。

テストは教授法の効果や学習の成果を見るためのものであるが、テストが授業や学習に与える影響も大である。CBT の導入によっても、テストのプラスとマイナスの波及効果が考えられる。CBT では、より authentic なテスト材料が使用され、より正確な言語能力の測定が可能になることから、CBT が授業のツールとして機能することによって目指す言語能力の習得につながっていくであろう。授業は、いわゆる言語についての知識の習得だけに終わらず、4 技能をバランスよく訓練しながらより現実に即した言語活動を中心とした方向に向かう。更に、スキーマの増長を目指した内容重視型の授業に切り替わっていく。一方、カリキュラムやその運営において、CBT への対応が多大となってくる。学生/英語教員へのコンピュータリタラシーの必要性、より多くのクラスへのコンピュータの導入、ソフトの開発、など人的資源の調達や設備の整備が必要になる。勤務校におけるカリキュラムを基に、実際の授業や教員への影響などを通してこれらの問題を論じたい。

根岸 雅史 (東京外国語大学)

改善点

1. テストの信頼性の向上: 「無駄な項目」を削減することによる信頼性の向上
2. 結果のフィードバックの即時性: 技術的にはテスト終了直後に結果をフィードバックすることが可能

問題点

1. コンピュータ使用による言語テストを3次元の視点から見た場合の問題
 - ・ 言語材料やトピックなどの組み合わせが「偶然に」任されることの危惧
 - ・ サブスキルの組み合わせが「偶然に」任されることの危惧
2. 受験者の心理的な問題
 - ・ CATの原理を知る受験者の心理

今後の可能性

1. 発表技能の測定の可能性

- ・ 音声認識技術との関連：コンピュータの音声認識技術が高まることにより、スピーキング・テストがネットワーク上で行われる可能性がでてくる。
 - ・ 発表技能のコンピュータによる自動採点の可能性：現状のTWEはネットワークでつながった人間が採点しているが、コンピュータによる自動採点は原理的に可能なのか。また、可能だとすれば、いつ頃実現されるのか。
2. 診断的機能の高いフィードバック実現の可能性
- ・ いわゆる言語能力を判定するだけのテスト・プログラムから、「診断」そして「学習」への有機的につながったプログラムが可能となる？

研究発表

Adapting the Levels Test into a Japanese Translation Format

Yoshimitsu Kudo (Rikkyo University)

Whereas Nation's vocabulary test was found reliable and valid by Read (1988), a number of researchers such as Schmitt and Meara (1997) and Beglar and Hunt (1999) have examined the test's validity. The latter went so far as to revise the 2000 Word Level and the University Word Level of Nation's test.

However, no researcher has adapted Nation's test taking relatively low-level learners into account. Therefore, a new translation definition-matching format test was developed for criterion-referenced purposes that includes five lower levels, namely, the 1000-, 2000-, 3000-, 4000-, and 5000-word levels. 325 secondary-level Japanese students participated in the pilot study. To examine dependability, squared-error loss agreement – phi (lambda) dependability indexes – were computed. To analyze validity, Guttman scaling and Item-Level correlations were employed.

Although each level was found highly dependable and the confidence intervals were small, validity was not strongly claimed because the mean of the 5000-word level was higher than that of the 3000- and 4000-word levels. The problem was identified as loanwords included in the 5000-word level which led to the higher mean. In the main study, those loanwords were replaced by non-loanwords. The tests were administered to 504 secondary-level students. This time, the means of the five levels steadily declined and were highly dependable and valid. Again, the confidence intervals were very small. This new, dependable, valid, and practical test can be used for both diagnostic and achievement purposes, and has a considerable possibility of being developed to a computer adaptive test by making an item bank.

What a Computerized Reading Test Can, but a Pencil-and-Paper Reading Test Cannot, Do Shizuka Tetsuhito (Institute of Foreign Language and Education Research, Kansai University)

The paper reports on the development and the evaluation of a computerized reading test. The first

feature of the test is that its items are sentence-based, not discourse-based. The advantage of sentence-based item type is that it is more readily handled in computerized adaptive testing. The second feature of the test is that it records the examinee's reading speed. Reading speed, when reliably measured, is predicted to contain proficiency information not revealed by number correct information. The third feature of the test is that it records confidence levels the examinee has in each response he/she makes. The reported confidence rating for a response is combined with its correctness using a method termed the Clustered Objective Probability Scoring (COPS), which ensures that the resultant score is free from possible influence of individual differences in personality.

Some 210 Japanese EFL college students took this test as well as the reading section of Cambridge First Certificate in English (FCE), which served as the criterion measure of the subjects' reading proficiency. The results clearly indicated that incorporating confidence and speed resulted in substantially higher reliability and validity than when scoring was done based on 1/0 correctness alone. Even more remarkable was the fact that reliability and validity levels attained by combining the three measures at the point at which only the first four items were attempted exceeded those achieved by scoring exclusively on 1/0 data at the point at which all 40 items were tried. Implications for computerized reading testing will be discussed.

Modifying an Existing Test to Allow it to be Delivered by Computer

Randy Thrasher (国際基督教大学)

In 1976 Jack Upshur and I were asked to develop a test that could be used by Japanese trading companies, international banks and other companies here to help them decide which of their employees had sufficient proficiency in English to be posted abroad. We created a two-stage test, the Businessman's English Test and Assessment (BETA). The first stage (BETA I) is a 210 item multiple-choice test of English structure, vocabulary, listening comprehension, and reading comprehension. Those who demonstrate a sufficient level of proficiency in those skills are invited to take the second stage (BETA II) which consists of 5 writing tasks (filling out a form, writing a memo, business letter, and telex, and drafting a speech introducing some aspect of Japanese life or culture) and two face to face interviews.

BETA has been used successfully for 25 years but, in the past year, economic considerations have forced us to develop plans for the modification of the test to permit it to be delivered by computer and make part of it computer adaptive. This presentation will introduce the test and will briefly discuss the situation that has made the change to computer delivery necessary and the steps required to make part of the multiple-choice stage of the test computer adaptive. The main discussion will focus on the modifications that will have to be carried out in the writing tasks and, particularly, the speaking tasks in order to permit us to deliver the test by computer. The presentation will conclude by examining possible threats to validity and reliability that are posed when writing and speaking

tasks are computer delivered.

Peer Ratings in EFL Writing Course: User Acceptance and Characteristics

Hidetoshi Saito (Hokusei Gakuen University)

Tomoko Fujita (Rikkyo University)

Objectives. Despite rising interests in application of alternative assessment to foreign language classrooms, the characteristics of peer and self-assessment have not been well understood. In particular, lack of research literature on the use of peer ratings in EFL writing course may prevent teachers from appreciating the utility of the system. Teachers may also be worried about the possibility of students' resistance against rating classmates' writings. The present study addressed the following research questions: 1) What are similarities and differences among peer, self-, and teacher ratings of EFL writing?; 2) Do students like or dislike peer ratings?; and 3) Is students' acceptance of using peer rating influenced by receiving high ratings from teachers and peers?

Method. Sixty-five college students studying English writing in a Japanese college participated in the study. They went through a rater training session and wrote 2 essays in one semester. Each essay was rated by 2 teachers, 3 peer students, and writers themselves with a 4-point rating scale. Students made final revisions using peer and teacher ratings and comments. Upon return of final teacher ratings, students completed an 11-item questionnaire on the utility of peer rating.

Results. While peer and teacher ratings were highly and statistically significantly correlated, peer-self and teacher-self ties were almost negligible. Concerning item difficulty hierarchies, teacher-self tie was found strong. Positive means were found across the user acceptance items, indicating students' positive attitudes towards peer rating. A regression analysis suggested that neither teacher nor peer ratings influence students' acceptance of peer rating.

「英語能力簡易測定のための語彙テスト—開発されたテストの妥当性・信頼性の検証」

片桐 一彦 (東京学芸大学 博士課程)

本研究は、日本の高校生・大学生向け英語能力簡易測定テスト開発を目指した研究プロジェクトの3番目の研究にあたる。

Katagiri (2000a)において、望月 (1998) の語彙サイズ測定テストが英語能力簡易測定テストとしても十分に機能することがわかった。また、筆記語彙サイズ測定テストとリスニング語彙サイズ測定テストの比較をおこない、音声処理を伴わない前者のテストのほうが TOEIC 総合得点との相関係数が高いことがわかった。

片桐 (2000b)では、「頻度」を機軸に作成された既存の筆記語彙サイズ測定テストから、語彙の「項目困難度」を機軸にしたテスト作成を試みた。「語彙の困難度を機軸に作成されたテスト」と「TOEICテスト得点や大手予備校模擬試験英語テスト得点」との間には、中から高い相関が確認された。10分で実施できる「項目困難度を機軸にした語彙テスト (32項目)」を3版作成した。

今回は、1) 実際にその3つの版がお互いに平行テストになっているのかどうか、2) 10分で実施できるそれぞれの版が「英語能力簡易測定テスト」の役割を実際に機能しているのかどうかについて、2つの先行研究も踏まえて発表する。

「パブリックスピーキングテストの分析：ラッシュモデルが教えてくれるもの」

中村 優治 (東京経済大学)

Oral Communication Skills とりわけ public speaking 能力あるいは presentation 能力はビジネス社会では特に重要視されており、その指導も熱心に行われている。しかしながら、この伝達能力の指導と表裏一体をなす public speaking の測定と評価に関しては研究報告はあまり多くなされていない。

本発表では、授業中に行われた public speaking (presentation) のテスト結果をラッシュモデルを用いて分析し、項目応答理論が私たちに何を伝えてくれるのか考えてみたい。

テスト手続きは次のように行われた。被験者(大学生15名)の presentation を評価者5名(教員1名、学生4名)が評価項目11項目(Tatum 1997を基に日本人学生用に簡素化したもの)を用いて評価し、その結果をラッシュモデルを用いて分析した。発表では被験者、評価者、評価項目の3つの facets を細かく分析し、検討する予定である。

「Aパラメーターから見た言語テストの弁別力」

湯浅 克己 (ILC BETAテストセンター)

弊社では、全部で10言語の語学テストを開発し運営しているが、各言語テストの各項目の弁別力を見てもっとも学習人口の多い英語のそれが一番低い傾向がみられる。

弊社で所有する言語テストのうち、受験者数の多い順に5言語(英、仏、独、西、中)を採り、IRT(項目応答理論)の各項目の弁別力(Aパラメーター)を調べてみた。アプリケーションは、BILOG3.0を使用した。

英語の場合は、サンプルの数が多いため、できるだけその他の言語の受験者数に近いものを選んだ。英語とその他の言語では、サブテストの問題数が少しづつ違っている。設問形式も若干の差異があるが、基本的には英語の測定方法を下敷きにしている。中国語は、ヨーロッパ系言語とは異なるが、設問形式はその他の言語と同一になっている。

各テストのそれぞれのサブテストの項目毎の弁別力(Aパラメーター)の平均値は、テスト開発は、英語のほうが専任の執筆チームをもち、予備テストを経て本テストが編集されるのにもかかわらず、ほとんどの領域で英語の弁別力は、その他の言語をかなり下回っていた。

「Test Data Analysis Program: TDAP Ver. 1.0E について」

中村 洋一 (常磐大学)

大友・中村(1996)は、大友(1996)『項目応答理論入門』(大修館書店)で示された言語テストのデータ処理に沿って、基礎統計量・標準得点・古典的テスト理論による項目分析・信頼性係数・項目応答理論による項目困難度パラメータ及び能力パラメータ・項目及びテスト特性曲線・項目及びテスト情報関数

の算出を行うコンピュータ・プログラムとして、Test Data Analysis Program: TDAP Ver. 1.0 を発表した。

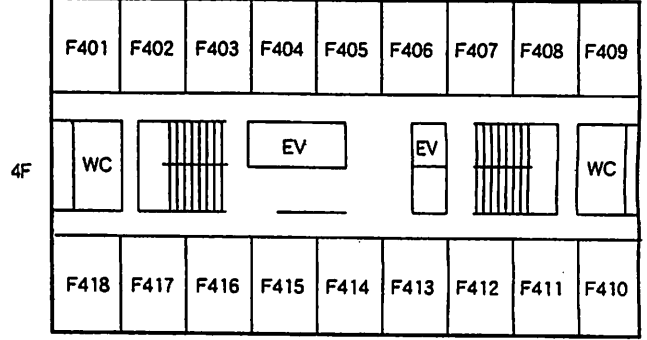
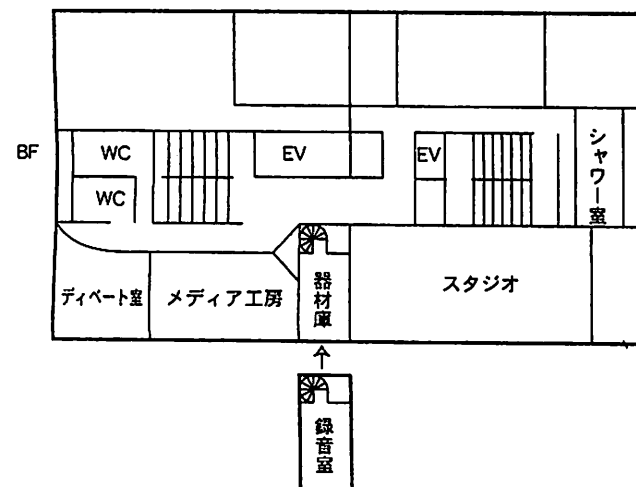
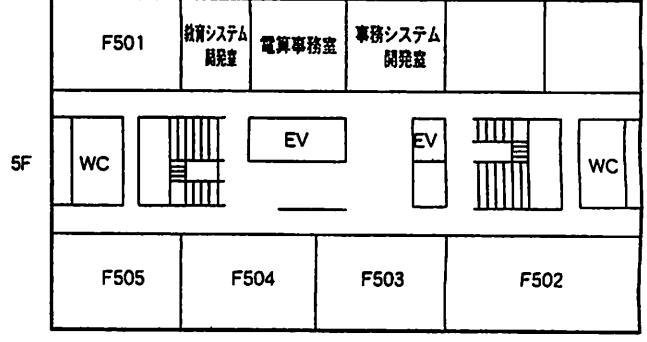
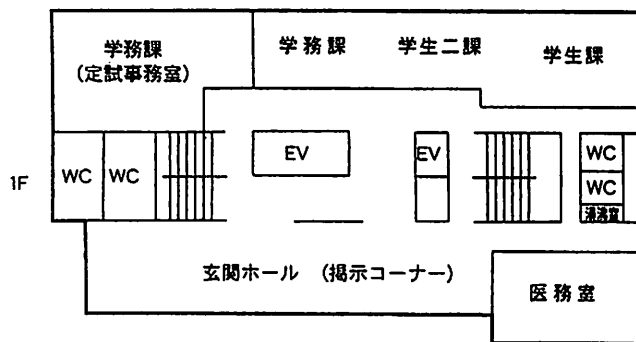
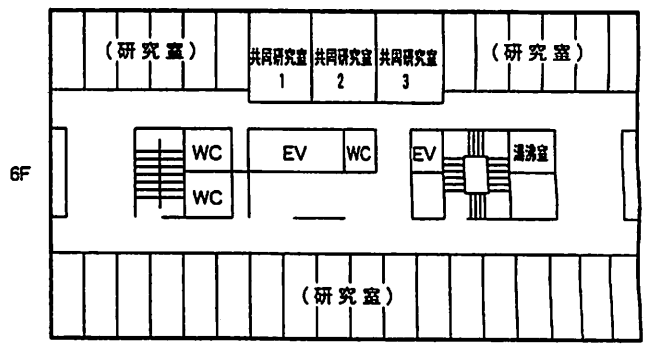
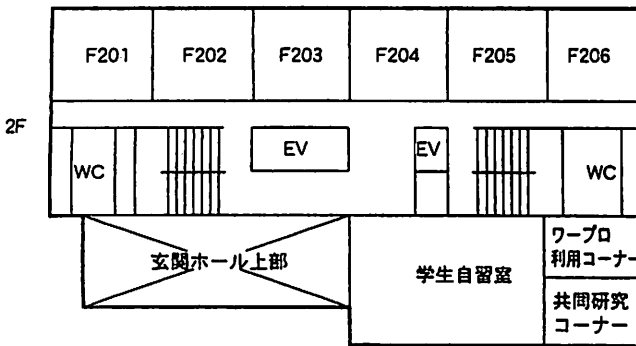
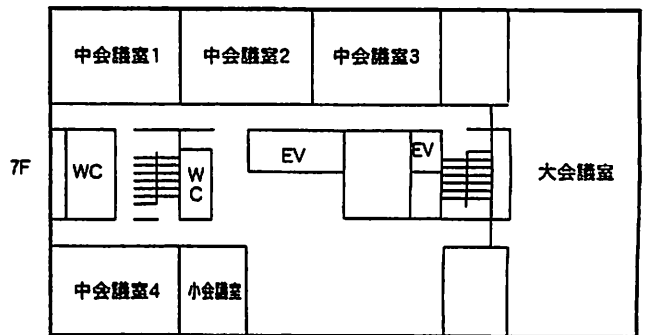
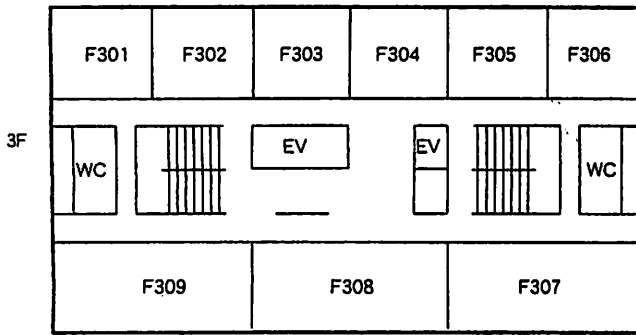
TDAP Ver. 1.0E (2000) は、東電ソフトウェア（株）との共同開発による、TDAP Ver. 1.0 の改良版である。TDAP Ver. 1.0E では、Windows 版 Excel のマクロによってプログラムを走らせ、Ver. 1.0 の課題となっていた、より汎用的なコンピュータ・システムでの使用と、より多くのデータ・セットの処理を可能にした。

発表では、TDAP Ver. 1.0E を使用した言語テストのデータ処理について紹介する予定である。

Classical and IRT Approaches to System Feedback on Language Testing for University Entrance
Steven Ross (関西学院大学)

Feedback systems for high-stakes examinations such as those commonly used in Japan are largely dependent on non-empirical, impressionistic critiques. These critiques are often given as a form of test pre-administration moderation by committees of language specialists, who often lack expertise in SLA, language assessment, or psychometric theory. Another approach is to invite post-administration critiques from secondary school teachers or their representatives, as well as those from cram schools. While these types of critiques offer a modicum of feedback, they cannot suffice to aid the creation of comprehensive positive and negative feedback systems for the cumulative betterment of entrance examinations. The present paper outlines the use of classical test item analysis and item response theory approaches for the purpose of creating synchronic and diachronic composites of language tests on university entrance examinations. These composites, it is argued, are essential for diagnosing test malfunction, item weight determination, and test tailoring.

6号館

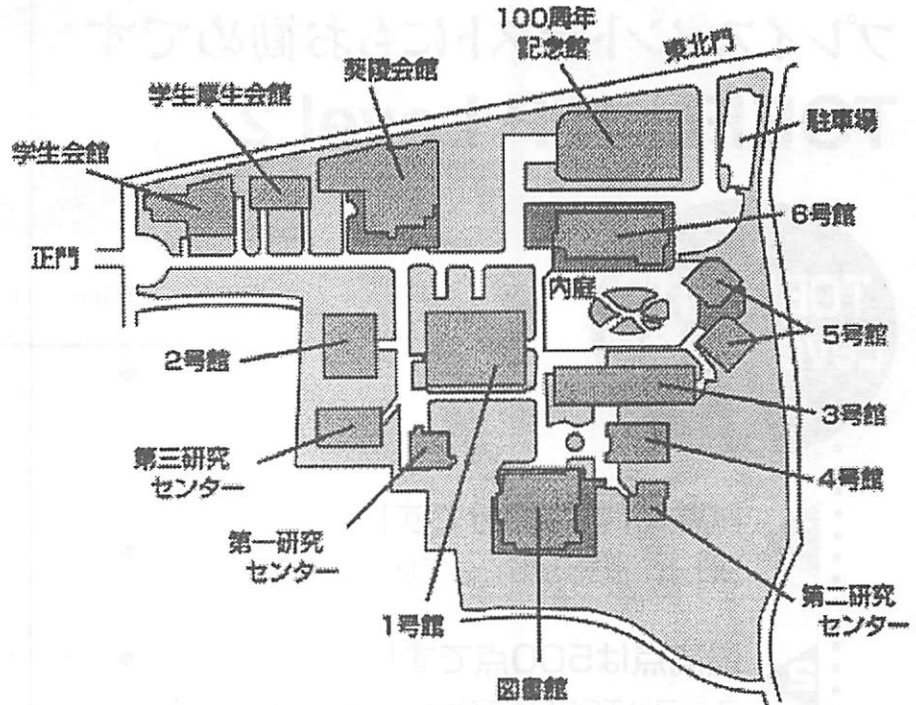


東京経済大学

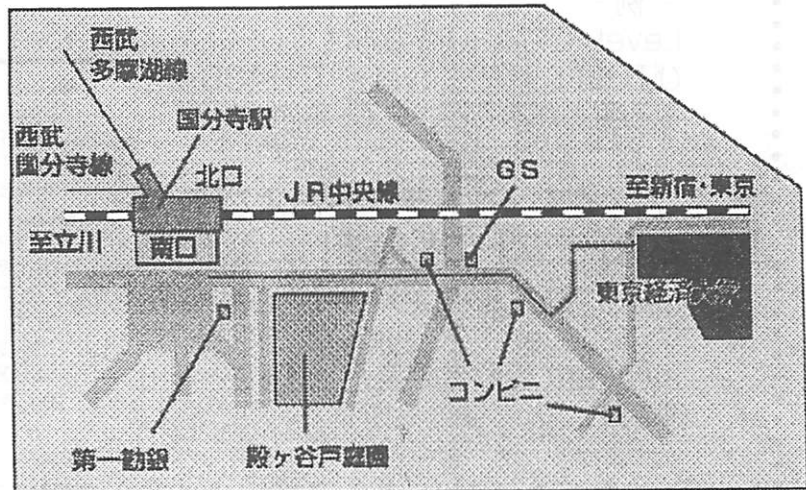
- ・キャンパスMAP
- ・交通経路

国分寺キャンパス

◆ キャンパスMAP ◆



◆ 交通経路 ◆



やさしいTOEFL®

ご存知ですか?

プレイスメントテストにもお勧めです

TOEFL® ITP Level 2



TOEFL® ITP Level 2の特徴

1

試験時間は70分です!

90分の授業の中でも実施できます。

2

最高点は500点です!

スコアはTOEFL、TOEFL-ITP Level 1
と高い相関関係にあります。

— 例 —

Level 2で430点の方は、TOEFLおよび
TOEFL-ITP Level 1でも430点を
取る実力があるとされています。

3

**プレイスメントテストおよびその
ポストテストにも最適!**

Level 2でプレイスメント、そのポスト
テストにLevel 1をご利用いただくことで、
より正確に受験者の実力を測ることが出
来ます。

4

10名から実施出来ます!

10名未満でも10名分の料金をお支払
いただければ実施可能です。

5

**一人当たりの料金はTOEFL®の約
1/4です!**

TOEFL® ITP(Institutional Testing Program)とは?

- TOEFL-ITPはTOEFL作成・運営元であるETSが提供する、TOEFLの団体向けテストプログラムです。問題は、過去のTOEFLで出題されたものを再印刷して利用しており、TOEFLと同様の受験形式で団体独自の目的に合わせてご利用いただくのに最適なテストです。
- TOEFL-ITPスコアに公的な効力はありませんが、出題形式や採点方法はTOEFLに準じており、TOEFLのスコアと高い相関関係にあります。
- TOEFL-ITPは、アメリカの大学付属語学学校(ESL)をはじめ、世界中の学校や教育機関などで多数利用されています。TOEFLの受付・登録業務と同様にカウンセルがETSより委託を受け、日本でのTOEFL-ITP運営・実施を行っています。

TOEFL® ITPは2種類のレベルがあります。

- Level 1 TOEFL®
Level 1 TOEFLは、過去に出題されたTOEFLをそのまま使用していますので、Level 1 TOEFLとTOEFLのスコアは高い相関関係にあります。最高点は677点です。
- Level 2 Pre-TOEFL®
スコアが500点以下の受験者を対象に、過去の問題を再構成して作られたやさしいTOEFLテストです。初めての方でも受験しやすいように、また短時間で実施できるように問題数を少なくしてあります。スコアはTOEFLやLevel 1 TOEFLと高い相関関係にあります。最高点は500点です。

お問い合わせ先

国際教育交換協議会(カウンセル)日本代表部
〒150-8355 東京都渋谷区神宮前5-53-67
コスモス青山ギャラリーフロア

TEL: 03-5467-5478 FAX: 03-5467-2185
(月曜~金曜9:30▶17:30 / 土曜・日曜・祝日休業)
<http://www.cieej.or.jp>

Council Exchanges

Council on International Educational Exchange