

The 11th JLTA Annual Conference

Place Building No. 14, Aichi Gakuin University (Nisshin Campus)

Dates October 28 (Sunday)

Reception (8:30-) [the Second Floor at Building No. 14]

Opening Ceremony (9:00-9:15) [Room 14202]

Presentation Part I (9:20-10:00) & Part II (10:05-10:45)

*Presentation titles are given on pages 5-11.

Room 14301 (A) Papers 1 & 2

Room 14302 (B) Papers 3 & 4

Room 14303 (C) Papers 5 & 6

Room 14306 (D) Papers 7 & 8

Break (10:45-10:55)

Presentation Part III (10:55-11:35) & Part IV (11:40-12:20)

Room 14301 (A) Paper 9 & 10

Room 14302 (B) Paper 11 & 12

Room 14303 (C) Paper 13 & 14

Room 14306 (D) Paper 15 & 16

Lunch Break (Room 14203) & Business Meeting (Room 14202) (12:20-13:20)

Symposium (13:20-15:20)

Theme: The Role of Language Testing in Curriculum Reform

Coordinator: Koizumi, Rie (Tokiwa University)

Panelists: Lee, Soo im (Ryukoku University); Horiuchi, Kayoko (Kanazawa Institute of Technology); Katagiri, Kazuhiko (Senshu University); Koshimuzi, Ikuko (Kofu Daiichi High School, Yamanashi)

Break (15:20-15:30)

Plenary Speech (15:30-16:40) "The Development of Language Tests and Extensive

Reading Activities."

Noro, Tadashi (Aichi Gakuin University) [Room 14202]

Break (16:40-16:50)

General Business Meeting (16:50-17:15)
[Room 14202]

Closing Ceremony (17:15-17:25) [Room 14202]

Banquet (18:00-19:30) [Gakuin Kaikan]

日本語テスト学会第11回全国大会プログラム

10月27日(土)

17:00 ~ 19:00 理事会 (サンプラザ シーズンス 蓬莱の間)

10月28日(日)

8:30~ 受付 (14号館2階)

(PC利用発表者: 14号館3階の各発表教室で機器接続確認)

9:00~9:15 開会行事 (14号館2階 14202室)

総司会 伊藤彰浩 (大会運営委員長・愛知学院大学)

挨拶 Thrasher, Randy (日本語テスト学会会長・沖縄キリスト教学院大学・
国際基督教大学名誉教授)

9:20~11:00 研究発表 (発表30分, 質疑10分) Part I 9:20~10:00

Part II 10:05~10:45

A室 (14号館3階 14301教室) 司会 法月 健 (静岡産業大学)

[1] Part I TOEIC[®] 受験者の英語能力記述ー TOEIC[®] Scale Anchoring Study
三橋峰夫 ((財)国際ビジネスコミュニケーション協会)

[2] Part II Can-do statements 調査と指標テスト：ESP の観点から
小山由紀江（名古屋工業大学）

B室（14号館3階 14302教室） 司会 島谷 浩（熊本大学）

[3] Part I 文法性判断テストによる明示的 / 暗示的知識
島田勝正（桃山学院大学）

[4] Part II 多肢選択形式による語彙サイズ測定テストの誤差の推定方法について
野上康子（(株)教育測定研究所）

C室（14号館3階 14303教室） 司会 小泉利恵（常磐大学）

[5] Part I スピーキングにおける生徒相互評価の妥当性：日本人高校生の場合
深澤 真（茨城県立竹園高等学校）

[6] Part II 速読と精読におけるテキストタイプとタスクタイプの要因の分析
長沼君主（清泉女子大学）
和田朋子（工学院大学）

D室（14号館3階 14306教室） 司会 片桐一彦（専修大学）

[7] Part I 日本語手紙文を評価する際に留意すべき点とはー評価過程の分析をもとに
宮島良子（名古屋大学）

[8] Part II 能力記述による日本語教育カリキュラムの改善
村上京子（名古屋大学）
10:45～10:55 休 憩

10:55～12:20 研究発表（発表30分、質疑10分） Part ・ 10:55～11:35
Part ・ 11:40～12:20

A室（14号館3階 14301教室） 司会 塩川春彦（北海学園大学）

[9] Part ・ 共通教育「英語」の成績評価における課題とその解決に向けた試みー

愛媛大学における事例

折本 素 (愛媛大学) 廣森友人 (愛媛大学)

田中英理 (愛媛大学) 山西博之 (愛媛大学)

山本武志 (ベネッセコーポレーション)

[10] Part ・ 言語コミュニケーション力の三次元的理解

柳瀬陽介 (広島大学)

B室 (14号館3階 14302教室) 司会 中村優治 (慶応大学)

[11] Part ・ A Meta-analysis of Task Type Effects on Listening and Reading Test Performance

In'nami, Yo (Kanda University of Foreign Studies)

[12] Part ・ A Dual-Testlet Approach to Student Placement

Shiotsu, Toshihiko (Kurume University)

C室 (14号館3階 14303教室) 司会 島田勝正 (桃山学院大学)

[13] Part III Validation of the EBB Scales for the Story Retelling Speaking Test

Hirai, Akiyo (University of Tsukuba)

Koizumi, Rie (Tokiwa University)

[14] Part IV Unveiling and Improving Current Speaking Evaluations of NS EFL Teachers in Japan

Sage, Kristie (Komazawa University)

Tanaka, Nozomi (Ochanomizu University)

D室 (14号館3階 14306教室) 司会 Randy Thrasher (Prof. Emeritus, ICU)

[15] Part III Exploring Korean Students' Perceptions of Test Fairness

Jungtae Kim (Yonsei University) & Seo Young Yun (Hankuk Univ. of Foreign Studies)

[16] Part IV Ratings of Korean students's English language oral proficiency

Hyun-Ju Kim (Dankook University)

12:20~13:20 昼食 展示ブース見学 (役員会 : 14202室 休憩室 : 14203室)

13:20~15:20 シンポジウム (14号館2階 14202教室)

テーマ：「カリキュラムの変革における言語テストの役割」

コーディネーター 小泉利恵（常磐大学）

パネリスト 1 李洙任（龍谷大学）

2 堀内香予子（金沢工業大学）

3 片桐一彦（専修大学）・興水以久子（山梨県立甲府第一高校）

15:20～15:30 休憩

15:30～16:40 講演（14号館2階 14202教室）

司会 伊藤彰浩（大会運営委員長・愛知学院大学）

紹介 Thrasher, Randy（日本言語テスト学会会長）

演題：多読指導の効果を測定するテストの開発について

講師：野呂忠司（愛知学院大学）

16:40～16:50 休憩

16:50～17:15 総会（14号館2階 14202教室）

議長選出

報告 中村洋一（JLTA 事務局長・常磐大学）

17:15～17:25 閉会行事（14号館2階 14202教室）

18:00～19:30 懇親会（学院会館）

司会 柳瀬陽介（広島大学）

Abstracts:

Paper Presentation（研究発表）

[1] TOEIC_Y受験者の英語能力記述－TOEIC_Y Scale Anchoring Study

三橋峰夫（財）国際ビジネスコミュニケーション協会

近年言語テストの結果として、単にスコアのみではなくそのスコアを取得した受験者が実際に英語を使ってどのようなことができる（できない）のかという、診断的情報をフィードバックすることが求められている。Can-do statements を用いて受験者（学習者）の言語能力を具体的に記述しようとする試みが、最近各方面で多く見られるのもその流れに沿

ったものであるといえる。但しテスト結果としてフィードバックする場合、当該テストが元々そういった情報を受験者から引き出すことを目的に設計されていなければ—いわゆる **Diagnostic Test** でなければ—テストの結果から直接必要情報は入手できないため、何らかの形で別途そのための調査をする必要が生じる。TOEIC[®]においては2006年5月のリニューアル以降、受験者に対し **Score Descriptors** と **Abilities Measured** という情報を提供している。これは各受験者の受験結果—受験したテスト問題の正誤状況—から導き出しているものである。これを可能にした新たなテスト作成手法である **Evidence-centered design**, 及び **Scale Anchoring Study** について紹介する。

[2] **Can-do statements** 調査と指標テスト:ESPの観点から
古屋工業大学

小山由紀江 名

言語を使用して何ができるかを述べた **can-do statements (CDS)** は、これまでの、学習者がテストの結果によって受動的に能力を「評価をされる」ことが主であった能力評価に対して、自らの能力を自らが評価するという主体的な評価の可能性を示すものである。CDSは **Common European Framework** を始めとして、資格試験のテストスコアを解釈することを目的とする TOEIC[®] や英検などでも制定されているが、学習者のモチベーションを高める評価法として各教育機関で独自のものが開発され始めている。しかし、外部評価テストと CDS の関連付けに関してはまだ試行錯誤が続いていると言ってよいだろう。本研究では、ESP の考え方に基づいて組まれた名古屋工業大学のカリキュラム（一般科学技術英語の習得を英語教育の目標とする）の中で行った CDS のパイロット調査の結果を、その他のテスト（TOEIC[®] 及び期末試験）の結果と比較分析した。一回目のパイロット調査の結果は以下のようにまとめられる。(1) CDS と TOEIC[®] の相関は低い；(2) **reading, listening** という **receptive skill** の自己評価が比較的高く **speaking, writing** という **productive skill** は低い；(3) **interactive** スキルに対する自己評価は5分野のうち最も低い；(4) 5分野全てを通して、自己評価が低いのは、専門性・抽象性の高いもの、**negotiation** や問題解決が必要とされる項目である。本発表では、さらにパイロット調査を踏まえて改訂した CDS の第二回目の調査結果を他のテスト結果と比較分析し、カリキュラムの目標と実施されているテストとの整合性についても論ずる。

[3] 文法性判断テストによる明示的 / 暗示的知識
院大学

島田勝正 桃山学

本研究の目的は、文法性判断テストにおいて、問題文の提示時間の制限の有無により、明示的知識(**explicit knowledge**)と暗示的知識 (**implicit knowledge**)の2つの異なる文法能力を測定することができるかを調べることである。1つの文法範疇毎に2項目の適格文

(grammatical sentence; GR)と2項目の非適格文(ungrammatical sentence; UG)を配し、20 文法範疇、80 項目から構成される文法性判断テスト(Grammaticality Judgment Test; GJT)と、同一の 80 項目に対して UG の誤りの場所を指摘させるテスト(Error Location Test; ELT)を開発した。GJT には、提示時間を制限した Timed GJT と制限のない Untimed GJT の2種類がある。Timed GJT と Untimed GJT および ELT の平均点差は小さく相関も高いという、大学生 81 名から得られた結果は、提示時間の制限の有無が必ずしも2つの文法能力の測定に貢献するものではないことを示唆している。一方、GR と UG の相関は、Timed GJT, Untimed GJT, ELT の3種類のテストのいずれでも低く、GR および UG を正しく判断する能力は異なると考えられる。

[4] 多肢選択形式による語彙サイズ測定テストの誤差の推定方法について

野上康子 (株)教育測定研究所

多肢選択形式のテストを用いて語彙サイズを推定する際の推定誤差を推定する方法を提案する。U 語からなる語彙ユニバースから n 語を抽出し、各語について、選択肢数が k 個の多肢選択形式の問題を作成してテストを構成する場合を考える。受験者は知っている語には必ず正解し、知らない語については、 $1/k$ の確率で正解するものとする。U 個の語彙のうち本当に知っている語彙の数(語彙サイズ)が w であるとき、出題された n 項目に含まれる「知っている語」の数の分布を算出し、これを重みとして、受験者が本当に正解を知っている項目の数 R を与えた場合に観察される正答数 x の分布を合成すると、U 個の語彙のうち本当に知っている語彙の数(語彙サイズ) w に関して、選択肢数 k, 項目数 n の場合の正答数 x の標本分布を導き出すことができる。これを用いて、語彙サイズの標準誤差および信頼区間を推定する。これにより、テストの長さや選択肢数を決める際に有効な情報が得ることができた。

[5] スピーキングにおける生徒相互評価の妥当性：日本人高校生の場合

深澤 真 茨城県立竹園高等学校

スピーキングの評価方法の1つである、生徒相互評価の妥当性についての研究結果を発表する。研究の主な目的は、(1) 教員による評価と比べた生徒相互評価の妥当性の検証、(2) 生徒の熟達度による生徒相互評価の妥当性の違いの検証、(3) 評価人数による生徒相互評価の妥当性の違いの検証の3点である。研究は、日本人高校生 79 名(有効データ数 62)を対象に調査を行い、併存的妥当性、表面妥当性、信頼性の3つの観点から生徒相互評価の妥当性の検証を行った。研究の結果、次の3点が明らかになった。第1に、全体評価において生徒相互評価は、教員による評価と比べても十分な妥当性が認められた。第2に、熟達度の上位群には十分な妥当性が見られ、中・下位群には妥当性が部分的に認められた。第3に、生徒相互評価を行う人数(5, 10, 20 人)については、上・中・下位群全てにおいて

妥当性が部分的に認められた。また、評価人数が増えるに従い、妥当性が高まる傾向も見られた。教育的示唆として、様々な形の生徒相互評価を活用した、実用的で、生徒にとってより安心感がある、よりダイナミックなスピーキング活動と評価の在り方を提案する。

[6] 速読と精読におけるテキストタイプとタスクタイプの要因の分析

長沼君主 清泉女子大学 和田朋子 工学院大学

長沼・和田 (2002, 2003)では、速読に影響を及ぼす要因として、テキストタイプ、トピック、文の長さに加えて、語彙や文法等のミクロ構造からの読みやすさの分析を行った。長沼・和田(2004)では、テキストのマクロ構造に焦点をあて、情報の緊密さから Association (AS), Description (DS), Causation (CS), Problem-solution (PS)の4つのテキストタイプ (Kobayashi, 1995) を比較したが、同一トピックのテキストをリライトしていたこともあり、一貫した傾向は見られなかった。本研究では異なるトピックで、テキストタイプ間の速読力を比較した結果、構造の緩やかなテキスト (AS, DS)で速読力が高い結果となった。また、概要、部分、比較、参照の4つのテストタスクによる理解度の測定では、DSやCSで正答率が高く、テキストタイプは速読力と理解度の2つの軸により分類された。タスクタイプとの関連では、緊密度の高いテキストでは参照型が、低いテキストでは概要型のタスクがより理解確認に適していた。速読テストと同一のテキストを用い、精読テストを行った結果、速読では、概要、参照、比較、部分の順で正答率が低下したのに対し、精読ではタスクタイプによる差が見られなかった。ただし、精読においてもテキストタイプによって正答率に差は見られた。今後は、タスクタイプとの関連に着目していきたい。

[7] 日本語手紙文を評価する際に留意すべき点とは—評価過程の分析をもとに

宮島良子 名古屋大学

近年、日本語教育の分野では海外でもコミュニケーション能力の育成を重視したカリキュラムが多く作成され、書く指導についても、手紙や文通といった項目が中等教育の教科書などで扱われるようになってきている。しかしながら、評価のあり方がパフォーマンステストではなくこれまで同様変わらないのであれば、学習者への波及効果は期待されないというのが現実ではないだろうか。本研究は、日本語作文の中でも、手紙文のように社会言語学的要素を含んだものを課題にした場合、テストとしての信頼性を確保するためには評定者はどのような点に留意しながら評価することが求められるのかを提言するため、まずは、実際にどのように評価しているのかを明らかにすることを目指した。そこで、発話思考法及び半構造化インタビューを用いて、発話プロトコル資料を収集し、作文評価の過程を分析・考察した。実験で使用した課題は「旅行へのアドバイス」であるが、この課題をどのようにとらえるのか、つまり、課題を達成しているとはどういうことであるのかという認識のずれ、呼称の選択、表現の選択といった読み手への待遇表現に関する考え方のずれ、

書き手の意図をどのくらい汲めるかの違いなどが評定結果に影響していることが示唆された。

[8] 能力記述による日本語教育カリキュラムの改善 村上京子 名古屋大学

日本語能力試験では各級に合格したか否かは判定されるが、各級に合格することがその学習者が実際にどのようなことができることなのかは示されていない。現在ヨーロッパ言語共通参照枠（CEF）に基づき、国際交流基金では日本語教育スタンダード構築に向けて動いているが、なかなか具体的な基準は示されるまでに至らない。従来、外国人留学生に対する日本語クラスの案内は、初級・中級・上級など大まかなレベルで示されたり、SJ100のような記号で表示されたりしてきた。しかし、受講希望者や外部の教員にはそのクラスの内容や修了すると何ができるようになるのかわかりにくい。そこで、本学の7レベルの各会話・独話・聴解・読解・作文の35クラスに関して能力記述によるレベル設定をCEFを参考に作成した。この記述と受講生のレベルが一致するかどうかを判定するために、Can-do-statementsを作成し、各クラスで実施した。その結果の分析から、レベル記述の一部変更が必要であることが明らかになった。また、Can-do-statements それ自体の問題も見えてきた。従来の自己評価に比べ、より具体的で対象者が答えやすくテストとしての信頼性は高いが、項目によっては識別力の低いものもあり、記述の改善が必要であることが示された。

[9] 共通教育「英語」の成績評価における課題とその解決に向けた試みー愛媛大学における事例

折本 素 愛媛大学 廣森友人 愛媛大学 田中英理 愛媛大学 山西博之 愛媛大学
山本武志 ベネッセコーポレーション

愛媛大学では平成13年度に英語教育センターを設置し、少人数クラス編成による学生生活中心型の英語授業の実施や、学生の実態を考慮した独自の英語教科書の作成などを通じて、英語による実践的なコミュニケーション能力の育成を推進してきた。これらの取組は、英語教育・学習に対して一定の成果を上げている一方で、各教員による評価の（極端な）ばらつきや成績評価の妥当性の確保といった新たな課題を生みつつある。平成19年度からは、これらの課題の解決に向けて、各授業の成績評価の30%分を英語能力試験（GTEC for Students）に充てる、学生の実態に応じた独自の英語運用能力判断基準（Can-Doリスト）を作成するといった新たな試みに着手している。前者の取組により、教員間の成績評価のばらつき相殺や、毎学期の成績評価（GTEC for Studentsによる30%を除いたもの）とGTECの結果の比較などが可能になり、後者の取組により、愛媛大学の学生の実情に応じたより適切な指導と評価が可能になると考えられる。本発表では、とくに前者の取組に焦

点を当て、英語能力試験を成績評価の一部に取り入れることの意義やその効果などについて報告したい。

[10] 言語コミュニケーション力の三次元的理解 柳瀬陽介 広島大学

言語コミュニケーション力の全体像を理解しようとする哲学的探究は、言語コミュニケーション力の一部分を精確に測定しようとする科学的研究と共に重要である。この発表では、言語コミュニケーション力を (1) 言語知識, (2) 心の理論・関連性理論, (3) 身体性の三次元のベクトルの合力として考える全体像を提示する。この全体像は、これまでの言語コミュニケーション力理論の発展史を踏まえながら, "capacity", "strategic competence" といった用語で現されていた言語知識を使いこなす力を, 認知科学の「心の理論」と「関連性理論」に沿って解釈する試みであり, また 1990 年の Bachman のモデルでは存在していた "psychophysiological mechanisms" の重要性を再主張するものでもある。この三次元的理解により, 言語をコミュニケーションで使う際の相手との相互作用性についての理論的理解が進み, また実践的にも, 各種英語教育で狙われている言語コミュニケーション力を統合的に理解することが期待できる。

[11] A meta-analysis of task type effects on listening and reading test performance

Yo Inumami Kanda University of International Studies

In contrast to a narrative approach to synthesizing study findings, this study conducted a meta-analysis on the effects of three types of tasks on listening and reading test performance. A total of 29 studies retrieved through a thorough search of the literature was the basis for estimates of mean effect sizes for listening and reading task type effects. Results using the fixed and random effects models of meta-analysis show that task type effects vary and are not sufficiently similar across studies except for in L2 listening. The difficulty order of multiple choice (MC) being the easiest, followed by open ended (OE) and then by summary gap filling (SG) tasks is observed across skills, languages, and research designs (independent-group, and repeated-measure with or without counterbalancing). The degree of task type effect difference is medium to large for MC vs. OE tasks and small to medium for OE vs. SG tasks, although the degree of MC vs. SG task type difference is wide-ranging and not very clear. Implications and directions for future research are also discussed.

[12] A Dual-Testlet Approach to Student Placement

Toshihiko Shiotsu Kurume University

As an integral part of an EFL curriculum revision project at a Japanese university, an instrument was needed to place incoming students into three level divisions of the newly introduced Core English course. Practical and theoretical considerations have led

to an in-house development and administration of a placement test, the process of which the presenter wishes to describe. Among the notable features of the test development was an attempt at a close interface between the test items and the prospective contents of instruction. Further, the test was designed such that it consisted of two testlets of distinct difficulty levels corresponding to the higher two of the three level divisions. Student performance on each testlet constituted the primary source of placement decision, i.e., sufficient performance on the Easier Testlet was judged as evidence of readiness for at least the Middle Level, in which case performance on the More Difficult Testlet was also consulted for decision on the student's readiness for the High Level. A Rasch-based calibration of item difficulty was carried out following a pilot testing with 1028 students, and the test revision took account of both such quantitative test data and qualitative feedback data from 629 individuals, who proctored and/or took the pilot test. The completed test demonstrated signs of good functionality overall, when subsequently administered to the target population of 1460 for the actual student placement. The presenter will provide further relevant information on the project while addressing known issues and limitations of the approach.

[13] Validation of the EBB scales for the Story Retelling Speaking Test

Akiyo Hirai University of Tsukuba Rie Koizumi Tokiwa University

Compared with other skilled tests, speaking tests in general have required much more time and cost for administration and scoring, which result in infrequent use of speaking assessment both in class and outside of the classroom. In addition, low reliability and validity for the ratings are expected, especially where raters are not trained and rating descriptors are vague (e.g., Upshur & Turner, 1999). To alleviate this situation, we have developed a highly practical semi-direct speaking test, which was named Story Retelling Speaking Test (SRST), and the practical scoring scales for the test using an Empirically derived, Binary-choice, Boundary-definition (EBB) technique. The EBB scales are easy to use and highly reliable though required less time for rater training (Turner & Upshur, 1996). However, it has been questioned whether EBB scales can be applied to other tasks (in our case, different stories) and whether scores rated by the EBB scales can be closely related to the scores rated by other existing scales. This paper focuses on these validity issues regarding the EBB scales. 43 low-intermediate EFL learners were asked to read and retell each of the four stories in the SRST sessions. Later, their SRST performances were rated by the EBB scales to examine whether the scales can discriminate the test takers similarly across the different stories. Also, their performances were scored by another existing scale and these scores were compared

with scores by the EBB scales. The results will be reported in the presentation, followed by the discussion of the applicability of the EBB scales for the SRST.

[14] Unveiling and improving current speaking evaluations of NS EFL teachers in Japan

Kristie Sage Komazawa University Nozomi Tanaka Ochanomizu University

This paper aims to present small scale quantitative and qualitative research results from a questionnaire that was administered to practicing university teachers in the Kanto region. The aim was to ascertain the authenticity of their speaking curricula; specifically, details about their evaluation methods and degree of internal or external regulation or moderation of their courses. The results presented will have been analysed and theoretical proposals offered for making their curricula more authentic and performance based. The support for this is based on the following. MEXT's Action Plan (2003) promotes the cultivation of "Japanese with English abilities." This study refers to one of its sub goals, "Research relating to English education at university," found in category seven "Promotion of practical research" (2003, pp. 20-21). For Japanese universities, this category suggests educators apply "concrete models" which encourage the use of the English language for work purposes after graduation. Authenticity requires that the interpretation of test scores generalise to the Target Language Use (TLU) domain (Bachman & Palmer, 1996). Hence, university curriculum ought to facilitate performance testing to equip students for using English in the international workplace. According to McNamara's (1996) Pandora's Box (second language communicative ability), use or performance is only one dimension of performance testing; the others being knowledge and real-time use of language. While performance testing is criticized as subjective, McNamara states it can be made objective, providing correct rating procedure is established (McNamara, 2000, p.37). In brief, conditions for performance behaviour, criteria for judging, and raters allocating a grade according to a criterion should be met.

[15] Exploring Korean Students' Perceptions of Test Fairness

Jungtae Kim Yonsei University & Seo Young Yun Hankuk Univ. of Foreign Studies

This study examines how Korean test takers of iBT TOEFL differently view the test based on several factors of test fairness. iBT TOEFL was chosen since it is a form of assessment that measures both language ability and cultural knowledge in order to estimate whether a student is capable of academic pursuit in North American

universities. The

study addresses the factor structure of test fairness perception questionnaire in terms of school policy, test validity, and bias against the test (based on Kunnan, 2000; Jang, 2002) and the relationship between test taker characteristics and the fairness. It also examines the factor structure of culture questionnaire probing test takers' cultural exposure to and knowledge of the target culture. The relationship between the cultural fairness and test taker characteristics is investigated as well. A total of 183 students in English classes at two Korean universities participated in the in-class questionnaires. The data were analyzed using explanatory factor analysis and reliability for extracting the factor structures of the test and cultural fairness and ANOVA and t-test statistics for examining the relationships above. The results of the statistical analyses indicate that test and cultural fairness are multi-faceted and could vary depending on local testing context and test takers.

[16] Ratings of Korean students π English language oral proficiency

Hyun-Ju Kim Dankook University

This study suggests that raters π attitudes toward non-standard Englishes impact the rating process of Korean students π English language oral proficiency. The researcher approaches this study by investigating the attitudes toward the application of World Englishes (WEs) in language teaching and testing from the three different groups of English teachers—the UK English teachers, Malaysian English teachers, and Japanese English teachers. The results show that there are significant differences among the groups of English teachers depending on their teaching contexts in their attitudes toward WEs in language teaching and testing. Through Cochran π s Q test, it was found that different groups of English teachers make different decisions for choosing rating criteria when assessing non-native speakers π English language oral proficiency. The English teachers showing negative attitudes toward WEs perceive grammar, pronunciation, native-likeness as more important rating criteria than the others such as vocabulary, organization, task fulfillment, and appropriateness for the assessment of Korean students π English language oral proficiency compared to the other group of English teachers showing positive attitudes toward WEs. In this study, the researcher argues for a pragmatic approach which looks at appropriate rating criteria for the assessment of English language oral proficiency not from the native speakers π viewpoint only but from the international English speakers π perspectives.

講演

演題：多読指導の効果を測定するテストの開発

野呂忠司（愛知学院大学）

科学研究費の助成を受け、多読の実践・実証研究を始めたところです。多読の効果を測るのに英検、TOEICやTOEFLのテストが多く使われていますが、テキストの種類、流暢さの点で不十分ではないかと考えます。読みの流暢さを測定するのに、読みの速さと正確さ、更なるその流暢さの基盤となる単語認知力について論じ、聴衆の皆様からご意見をいただき、テスト開発に役立てていきたいと思ひます。

シンポジウム

テーマ：カリキュラムの変革における言語テストの役割

コーディネーター 小泉利恵（常磐大学）

パネリスト 李洙任（龍谷大学）、堀内香予子（金沢工業大学）

片桐一彦（専修大学）・水以久子（山梨県立甲府第一高校）

コーディネーター 小泉利恵（常磐大学）「本シンポジウムの目的」

Brown (2005) によると、カリキュラムのデザインには (1) ニーズ分析、(2) 目的と目標の設定、(3) 言語テスト（集団規準準拠テストと目標基準準拠テスト）、(4) 教材開発、(5) 指導、(6) プログラム評価が関わる。本シンポジウムでは、その中の「言語テスト」に焦点を当て、3組のパネリストに、それぞれ所属の教育機関で用いているカリキュラムと言語テスト、そして両者の関係について論じていただく。3組のケース・スタディの発表と質疑応答を通して、効果的なカリキュラム変革を成し遂げるためには、言語テストをどのように使用していくべきか、注意すべき点は何か、カリキュラムのデザインの他の要素とどのように関連付けるべきかについて検討したい。

第1発表者 李洙任 (Soo im Lee) (龍谷大学) 「言語テストの構成概念的妥当性の考察」

本研究では、言語テスト(英語)における妥当性に注目し、言語テストのあり方と点数の解釈、そしてカリキュラム目標設定との関係を考察することを目的とした。テストの得点解釈を正当化するためには、(1) テストの特徴と(2) カリキュラム上の目標、そして(3) 授業内で学習する英語との一致を常時確認する作業が必要である。ある教育機関が「英語によるコミュニケーション能力を身につける」ことを教育目標として掲げた場合、学習者が習得すべき英語能力は「コミュニケーション能力」であり、使用されるテストは「英語によるコミュニケーション能力」を正確に測定・評価する手段でなくてはならない。テスト作成専門家や教員たち、そして教育機関の経営に携わる人たちが、テストの有効利用を怠

ったとき、テストを実施する意義が失われる。よって、カリキュラムで掲げた目標、すなわち構成概念的妥当性を言語テストにおいてどの程度まで具体化できるか、ということは英語教育において最優先課題となる。

第2 発表者 堀内香予子（金沢工業大学）「英語カリキュラムの工夫と TOEIC Bridge Ⅰの活用」

金沢工業大学は全校学生数約 6,800 名の大規模工学系大学である。学生の多くが英語に対して苦手意識を持ち、英語を勉強するという習慣を持たず、また語彙数や基本的文法知識が不足している傾向が見られる。本学基礎英語教育センターではこの現状をふまえると同時に、学生の人生や将来の仕事において成功の糧となる英語コミュニケーションスキルの習得を目標としている。同じ科目を複数の教員が担当するため同一の学習支援計画書（シラバス）によって学習内容、評価基準、授業予定を定めている。特色としては、科目担当者が交代で作成する小テストを毎時間課していることである。目標を細分化することで勉強する範囲が狭まり高得点を得た際の満足感によって動機づけを高めることをねらいとしている。以上の英語必修科目における成果を測るため、平成 17 年度に TOEIC Bridge Ⅰを試行として導入し入学生の英語力を測定した。平成 18 年度入 学生からは TOEIC Bridge Ⅰをプレイスメントテストとして実施し入学時の英語力を測定、これら対象学生の必修科目修了後に再度 TOEIC Bridge Ⅰによって英語力の測定を行った。結果、対象学生の 英語力に伸びがみられた。

第3 発表者 片桐一彦（専修大学） 興水以久子（山梨県立甲府第一高校）

「SELHi における諸テストの活用とその成果と今後の課題」

山梨県立甲府第一高校（以下、甲府一高）では、「縦断的に」（longitudinally: つまり、入学時から 3 年間に渡って約 1 年おきに同一の生徒に対して）、さまざまな英語のテストを実施した。最近では、項目応答理論を駆使した TOEIC Bridge Ⅰや GTEC for students といった比較可能な絶対得点を算出するテストを複数回生徒に受験させ、英語力の伸びを検証する高校が少しずつ散見されるようになってきた。しかし、そうした高校はまだ多くはないようである。甲府一高では、3 年間に渡ってこれらを実施した。また、「望月（1998）語彙サイズテスト」（知っている英単語の数を測定するテスト）や「スピーキング・テスト」（著作権者より許可を得て、英検準 2 級の過去問 [絵のカード] を利用し、描写説明させる口答試験）も実施した。これらの記述統計等の結果に関しては甲府一高出版の『平成 18 年度（3 年次）研究開発実施報告書』に記載され、また、語彙の伸びに関しては片桐（2007）にて、スピーキング能力の伸びに関しては小泉・片桐（2007）にて詳細に分析されている。それらの結果を紹介するとともに、その結果をどのように教育現場に還元でき活用できたか、また逆に何がうまく還元できず活用できなかったかについて報告する。