

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 31 日現在

機関番号：32644

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22520584

研究課題名（和文） 一万人の必須英語プログラム改革の為に有効性の高いプログラム評価システムの開発

研究課題名（英文） Developing effective program evaluation systems of the required English courses for 10,000 students.

研究代表者

藤田智子 (FUJITA TOMOKO)

東海大学・外国語教育センター・教授

研究者番号：80329002

研究成果の概要（和文）：日本の大学英語教育プログラム評価を効率よく有効に行う方法を検討するために、事前事後テストに焦点を当てた。項目応答理論(IRT)の利用により、事後テストは事前の14%の被験者数であっても被験者能力値(θ)の平均値変化が推定可能となる。また、事前事後テストを最適の等化の方法(IRTモデルと等化方法)を選択して実施する重要性を研究した。さらに θ とニーズアンケート結果を用いて分析し、プログラム評価のさらなる証拠とすることも提案する。

研究成果の概要（英文）：In order to seek an effective and efficient method of program evaluation for English language programs at Japanese universities, pre-and post-tests are the main focus. Using item response theory (IRT) makes it possible to estimate the average growth of students' ability level (θ) even when the number of participants in a post-test is only 14% of the number of pre-test participants. Moreover, the importance of selecting the most appropriate way of equating (IRT model and equating methods) is researched. Finally, other evidence of program evaluation is shown utilizing both θ and needs analysis survey results.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	900,000	270,000	1,170,000
2012年度	1,000,000	300,000	1,300,000
総計	3,000,000	900,000	3,900,000

研究分野：人文学

科研費の分科・細目：言語学、外国語教育・教育評価・測定

キーワード：大学英語プログラム評価

1. 研究開始当初の背景

(1) 研究の学術的背景

研究代表者は、博士論文「Validation of a Japanese University English Language Placement Test」(藤田, 2005)や科研費報告書「テスト妥当性を高める為の英語プレースメントテスト作成システム」(藤田, 2007)以来、項目応答理論(IRT)を利用してテストを分析し、より被験者に適したテストとその結果の使用を実行するための研究を行っ

てきた。また、本研究で扱う英語教育プログラムには2010年の改革以前から、主に評価の部分に密接に関わり、ヨーロッパ共通参照枠(CEFR)に準拠した独自の能力記述文(Can-do statement: CDS)を基盤にした新コースを開発し、プログラム改革にも携わってきた。

これらの背景から、この改革の前後でプログラム評価を実施して、1万人規模の大学英語教育プログラムにおける「改革による変

化)を、統計学的な結果による量的分析によって明らかにする評価システムを開発するという着想に至った。

また、この評価活動過程を、シンプルかつシステムティックで、より有効性の高い評価とその結果の利用ができるように開発し、「大学における英語プログラム評価のモデルケースの一つ」となるような研究の基盤を構築することをめざした。

(2)本研究の学術的特徴

Stufflebeam(2000)によると、プログラム評価活動は、操作が簡単であること、費用やエネルギーがかかり過ぎないこと、評価プロセスや情報収集手段が明確であること、などを基準にして行われるべきだと述べている。本研究の目的は1大学のプログラム自己診断を超えて、英語プログラムの評価のシンプルでわかりやすいモデルシステムの確立をめざす。これによって今後のプログラム評価の指標ができ、プログラムごとの評価結果を比較する事も可能になることを期待する。

実際、大学英語教育においてプログラム評価の研究の一部であるニーズ分析や、入口出口テストでプログラム履修前後の英語力の変化を報告した研究はあるが、一万人規模のプログラム評価したものは、多く報告されていない。その1つの理由は、それぞれの英語プログラムには独自の事情があり、自己診断的な目的で実施され、評価結果はその大学だけの為であり、他のプログラムへの汎用性が低いと思われることが挙げられる。しかしこれはプログラム評価の統一基準があれば、他校と結果を比較したり情報交換や相互協調することも可能になるはずである。

2. 研究の目的

(1) 問題点

大学教育の「質の保証」の必要性が緊急課題である昨今、受講生たちにどれだけ教育効果をもたらすか検証すること(プログラムアウトカム評価)(安田, 2010)の重要性が高まりつつある。このアウトカムがわかりやすい形で表わせるものの一つとして、プログラムを履修する前と後でどのように受講生の能力が変化したのか測る事前事後テストがある。

しかしながら、日本の大学教育の中では、この事前事後テストシステムを構築する試みあまり実施されていない。おそらくその理由には、大規模なテストを1学年全員に事前と事後の2回実施することで大きな労力とコストがかかること、また、テストデータに最適な方法を選択してテストの等化や結果分析をするための専門知識が必要とされること。そしてさらに、教員作成によるテストの場合は、質のコントロールが難しいこと

などが挙げられる。

本研究で対象とする大学英語教育プログラムでは、履修学生全員である約5,000人~6,000人が事前テストを受験する。これには約150人の試験監督と100室以上の教室を要する。事後テストにおいても、これと同じ規模で実施することは非常に難しい場合が多い。しかし、項目応答理論(IRT)を利用することによって、事前テスト受験者の約14%にあたる約700人のみが事後テストを受験するとしても、事前事後両テストの受験者の能力値(θ)の変化は推定でき、プログラム評価の証拠のひとつにすることは可能である。本研究では、このように受験者数が事前事後で大きく異なる2つのテストを共通項目デザイン(Angoff, 1984; Ayala, 2009; Kolen & Brennan, 2004)で等化するとき、最も適切なIRTモデルと等化方法を選択する方法を提案する。2通りの項目応答モデルと5通りの等化方法を組み合わせて10種類の等化の方法(ここでは項目応答モデルと等化方法の組み合わせを、等化の方法と呼ぶことにする)を比較する。この中には、教育現場の教員が比較的容易に計算できる等化方法も組み入れた。

また、ここで使用した事前事後テストは、実際に教えている内容をもとに教員が作成した英語リスニングテストである。これらのテスト問題は中間テストや期末テストとして英語プログラムで過去に使用され、既に分析されて項目パラメータも分かっている。これらの問題の中から適切な項目を選び、新たに問題を作成せず編集した2つのテストフォームを事前事後テストとして使用している。これらの事前事後テストを通じて推定された受験者の能力値(θ)を利用して、学年ごとの能力値平均の推移を観察し、また、プログラムニーズに関するアンケート結果と比較して、能力値を伸ばした学生のニーズを理解することによって、より有効なプログラム評価へ繋げることをめざしている。

このように、本研究は、日本の大学教育機関での事前事後テスト実施を妨げていると考えられる多大な労力とコスト、高度に専門知識を要するテスト分析、困難なテストの質のコントロールなどの問題点を少しでも改善し、より有効性が高く効率の良い事前事後テストシステムを、教育機関が取り入れる手がかりの一助となることを目的としている。

(2) 本研究が明らかにしようとする事

事前テスト受験者は5000人以上だが、コストや労力を抑えるために、事後テスト受験者をその14%である約700人で実施するようなどき、2パラメータ(2PL)モデル、3パラメータ(3PL)モデルと5つの等化方法合わせて10通りのうち、どの方法を利用する

のが最適であるのか決定する方法を検討する。また、最適な等化の方法と、最も適さないと判断された方法で、実際のデータを等化して受験者の θ を推定したときに生じる差について調査する。

この事前事後テストによって推定された受験者の θ を利用して、どのようなプログラム評価の証拠を提示することができるのか検討し、習熟度別に集計した受験者の平均能力値変化や、受験者 θ の伸びとプログラムニーズアンケート結果の関係が、プログラム評価の証拠になる可能性はあるのか探る。

日本の教育機関で実施しやすいシナリオで、労力やコストを省いた「プログラム評価のための事前事後テストシステム」の実現可能性を念頭に置き、 θ を利用してどのようなプログラム評価の証拠を提示できる可能性があるのか検討する。

3. 研究方法

(1) 受験者とテスト

ある大学の同じキャンパスにおいて2年間の必須英語教育プログラムを履修する1年生のほぼ全員、約6000人が事前リスニングテストを受験した。このプログラムでは、新入生は事前テストの結果で上級(A) 中上級(U)、中初級(L)、初級(B)の4つのレベルに分けられ、それぞれの習熟度に適した教材を使用して2年間で168時間の必須英語プログラムを履修する。この事前テストを受験した全学生のうち、約700人が事後テストをプログラム終了時に受験した。上記4つの習熟度レベルそれぞれの約10~20%ずつの学生をクラス単位で抽出し、標本の代表性を保つように考慮した。

最初に、約6000人が年度01に事前リスニングテスト30項目を受験した。事前テストから10問を共通項目とし、過去に実施したリスニングテストの中から20問をこの共通項目に加え30問の事後テストを作成した(図1)。この教員作成テストの20問は事前テストの内容と重複がなく、事前テストの項目より困難度がやや高くなるように選択した。

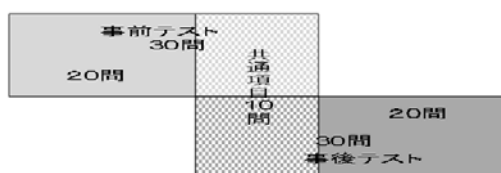


図1. 事前事後テスト

項目反応理論を用いて異なるテストに含まれる項目パラメタの等化をする際には、そこに含まれる共通項目のパラメタの値が不変であること、もしくは、各テストを受験した受験者の能力値が不変であることを仮定

する必要がある。本研究においては、事前テストと事後テストの実施の間に2年の隔たりのあり、しかもその間に受験者が英語を学習していることから、2時点における受験者の能力が不変であると仮定することには無理があると考えられる。他方、共通項目に関しては、対象がリーディングと比べ比較的記憶に残りづらいと考えられるリスニングテストであることや2年の間隔の長さから、両方のテストに含まれる10項目の共通項目の項目パラメタは事前テストの時点から変化していないものと仮定することに、それほど無理はないと考えられる。このような観点から、本研究では、事前テストと事後テストの差異を、それを受験した受験者の能力値の変化として捉える立場を取った。

(2) プログラムニーズアンケート

Busch, et al. (1992) をもとにして作られた Kikuchi (2005) の、ニーズ調査のための100問のアンケートから19問と、独自の問題6問を合わせて25問のプログラムニーズアンケートが作成された。学生は、自分たちの授業での学習方法(10問)、先生の指導方法(7問)や評価(7問)に関する3つのカテゴリーの質問に対して、同意するかどうかのレベルを4段階で回答した。

アンケートを実施する英語教育プログラムの現状に適合し、そこで最近実施されたプログラム改革に関連があるものを優先的に選択した。新たに作成した6問のうち5問は、その英語教育プログラム独自の内容や、プログラム改革に関連ある質問である。

(3) 最適なIRT項目パラメタモデルと等化方法の選択

本研究のデータを分析するのに最適な等化の方法を探るため、2通りの項目応答モデル(2PL, 3PL)と、5通りの等化方法: ①MM, ②MS, ③HB, ④SL, ⑤CR これらをそれぞれに組み合わせて、合計10種類の方法の中から最適な方法を選択するための評価を実施した。項目応答モデルの選択や等化方法の比較の基準としては、項目パラメタの値の安定性を用いることも可能であるが、本研究では、教育機関のプログラム評価における θ の果たす役割を考慮して、その基準として θ に準拠したものを採用することにした。



図2. 等化の方法の評価

まず図2にあるように、評価の前提として共

通項目に着目した。事前テスト 30 項目のデータを BILOG-MG3 (2003) で 2 つの項目応答モデルを使ってそれぞれ分析し、30 項目のパラメタの中から、共通 10 項目のパラメタ (これらを Pre-par とする) を取り出す。また、同様に事後テスト 30 項目のデータを 2 つの項目応答モデルによって BILOG-MG3 で分析し、推定された 30 項目のパラメタから共通 10 項目のパラメタ (これらを Post-par とする) を取り出す。これらの 2 種類のパラメタ

(Pre- and Post-par) を使って前述の①～⑤の 5 通りの等化方法により事後テストを事前テストに等化する。

ここで事前テストと事後テストのデータから共通項目 10 項目分の項目反応を抜き出し、これらを縦に繋いで結合したデータファイル (D-data) を作る。この D-data を用いて、事前テストのパラメタの共通項目部分 (Pre-par) を使用して、すべての受験者の 2 通りの θ を計算し、これらを 2PL $\theta_{_0}$ と 3PL $\theta_{_0}$ とする。

さらに、①～⑤の方法で等化した事後テストのパラメタの共通項目部分 (Post-par) を使い、D-data のすべての受験者 θ を計算し、分かりやすくするために、使用した等化方法の名前を θ に付け 2PL $\theta_{_MM}$, 2PL $\theta_{_MS}$, 2PL $\theta_{_HB}$, 2PL $\theta_{_SL}$, 2PL $\theta_{_CR}$ とし、3PL モデルの 5 つの θ にも同様に 3PL $\theta_{_MM}$ ～3PL $\theta_{_CR}$ のように名づけた。もしもモデルの選択が正しく、しかも等化が適切に行われていれば、これらの θ の値、例えば、2PL $\theta_{_0}$ と 2PL $\theta_{_xx}$ は、理論的には等しくなる値である。従って、2 つの項目応答モデルごとのすべての受験者の $\theta_{_0}$ を真値 ($\theta_{_0}$) とし、これらのモデルの違いと等化方法の違いからなる 10 通りごとの受験者 θ ($\theta_{_i}$) との平均 2 乗平方誤差 (RMSE : Root Mean Square Error) を計算した。10 通りの RMSE の中で、最も 0 に近いものが本研究データの分析に最適な方法だと考えることができる。

4. 研究成果

(1) モデルと等化方法の決定

① RMSE による評価

表 1 は、等化係数を使用して、前述した等化の評価方法によって、2PL モデルと 3PL モデルごとに $\theta_{_0}$ 、ならびに $\theta_{_MM}$ ～ $\theta_{_CR}$ の値を計算して RMSE を求めたものである。この表の中で、RMSE が最も 0 に近い方法が最適の方法であると考えられる。まず、年度 0 1 と 0 2 の項目応答モデルによる違いについては、年度 0 1、0 2 とともに 2PL が 3PL より小さい RMSE を示している。さらに、5 種類の等化方法については、MS 法の RMSE が一番高く平均値は 0.093、次に MM 法と HB 法で 0.091、SL 法は 0.089、最も RMSE が小さい等化方法は CR 法の 0.088 であった。

これらの結果から、本研究のデータにおいて 10 種類の等化の方法のなかで、RMSE が最も小さくなるのは、2PL モデルで CR 法によって等化した場合 (2PL-CR) で、RMSE が最大となるのは 3PL モデルで MS 法によって等化した場合 (3PL-MS) であることが分かった。しかし、5 種類の等化方法間での RMSE の差は非常に小さく、最小の CR 法と最大の MS 法との差はわずか 0.047 であった。なお、RMSE の計算式に現れる二つの θ の差の 2 乗の、モデルと等化法による 2 元配置の分散分析の結果は、各年度とも、モデルも等化法もきわめて小さな p 値で有意な結果となった。しかし、等化法の MS 値はモデルの MS 値の 50 倍以上となり、各年度とも受験者数が大きいこと (N=6630, 5157) を考慮すれば、D-data に関する限り、RMSE に及ぼす等化法による違いは実用的に無視できるほどの大きさであると考えられる。

表 1. 年度 0 1 と年度 0 2 における 10 通りの等化の方法による平均値 θ の RMSE 比較

モデル	等化方法	年度01	年度02
2PL	① MM	0.065	0.092
	② MS	0.066	0.092
	③ HB	0.065	0.090
	④ SL	0.065	0.089
	⑤ CR	0.064	0.089
3PL	① MM	0.104	0.102
	② MS	0.111	0.103
	③ HB	0.110	0.100
	④ SL	0.102	0.099
	⑤ CR	0.100	0.099

項目パラメタを安定して推定するために必要な受験者数は、3PL では 1000 人と言われているとおり、事前テストは 5,000 人以上であっても、事後テストの受験者数が 1000 人に満たない本研究のようなケースでは、やはり 3PL より 2PL モデルが相応しいことが確認できた。共通項目の θ を利用し、RMSE を基準とする評価は、モデルの違いが判定結果に与えるインパクトは大きく、本研究のデータには、2PL が 3PL よりも適合しているという結果を示唆している。今後、繰り返し継続的に本研究と類似したデータで事前事後テストの等化を行う場合は、2PL モデルを使用することが適切である。

どのような方法で等化するか即座に決定できないときは、等化する方法を決める何らかの基準があることが望ましい。本研究では教育現場での θ の果たす重要性を考慮して、共通項目の θ を利用して RMSE による評価方法を判断基準に採用し、データによりふさわ

しい等化の方法を決定することを提案した。

また、このとき IRT モデルの違いによる RMSE へのインパクトは、等化方法のインパクトより大きいことが確認でき、等化の方法を選択するときは、まずは IRT モデルの選択を優先するべきだと考えられる。さらに、等化方法の選択に関しては、本研究で用いた実際のデータにおいて、5つの等化方法の RMSE に大きな違いはなかった。従って、本研究の目的の一つである、高度な専門知識を必要としないテスト分析を行うには、複雑な計算を要する項目特性曲線に基づいた HB 法、SL 法、CR 法よりも、計算がシンプルな項目困難度の推定値の平均値や標準偏差を利用する MS 法や MM 法のほうが適切だと言える。

(2) 2通りの方法での学生平均能力値 θ

最適とされたモデルと等化方法の組み合わせである 2PL-CR と、最も適切でないとい評価された 3PL-MS により計算された等化後の 30 項目の項目パラメタを用いて、リスニング事前事後テストを両方受験した年度 01 の 688 人と、年度 02 の 772 人分のデータを分析し、それぞれの θ を計算した。表 2 は、2PL-CR 法で推定した年度 01 と年度 02 の事前と事後の間の学生平均能力値 θ の変化である。また表 3 は、3PL-MS 法を用いて同じように 2 セットの事前事後テストによって推定された受験者能力値変化の平均値である。同じデータであるにもかかわらず、使用するモデル・等化方法によって結果がかなり変わることが確認された。

表 2 2PL-CR による能力変化

年度	θ の変化平均値	SD
01	0.411	0.822
02	0.286	0.881

表 3 3PL-MS による能力変化

年度	θ の変化平均値	SD
01	0.269	0.767
02	0.135	0.808

これによると、より適切な等化の方法である 2PL-CR 法で推定した事前事後テストの θ の変化は、年度 01 で $\theta=0.142$ 、年度 02 で $\theta=0.151$ 3PL-MS 法による推定値より大きい伸びを示した。この理由は定かではないが、RMSE の比較に際しては、共通項目 10 項目のみを用いて推定された θ の安定性・不変性を調べたのに対し、(2) に於いては、事前・事後テストのそれぞれに含まれる全ての項目 (30 項目) を用いて θ の推定を行いその伸びを調べている。したがって、共通項目以外の部分に等化方法の違いの影響が出たと解釈することも可能である。

今回の本研究の実データを基にした RMSE 比較結果では、大きな違いはないものの 5 種類の等化方法のなかでは、CR 法が安定して最も誤差の少ない推定を行える可能性を示した。さらに、最近のシミュレーション研究 (eg. Arai & Mayekawa, 2011 ; 光永・前川, 2012) でも、CR 法による等化の優位性を示す結果が得られている。計算の煩雑さはあるが、PLINK (2007) などのパッケージを用いると比較的簡単に等化係数を求めることはできるので、教育機関において、項目バンクを作成したり、繰り返し等化を行う場合はオプションの一つとして考慮する価値はあると考える。

(3) プログラム評価の証拠

① 習熟度レベル別 θ の伸び

両年度通して θ の伸びが最も低かったレベルは L (中初級) であった。特に年度 02 では、他のレベルと比べ中初級レベルの伸びの落ち込みが激しく、このレベルの θ の伸びが非常に低いために、他のレベルの θ の伸びも影響を受けて年度 01 に比べ、年度 02 が全体的に低い θ の伸びを示す結果となっている。事前事後テストによる調査結果により、中初級レベルだけ θ の伸びが低いことが確認できた。これはプログラム評価の証拠の一つとなり、原因を解明する必要がある。大規模な英語教育プログラムの中で、B (初級) レベルは、他のレベルよりややクラスサイズが小さく平均 25 人から 28 人になるなど特別な対応を受けているが、中初級レベルのクラスサイズは他のレベルと同じ対応である。この他に何か要因があるのか、学生・教員へのインタビュー調査も含めて詳しく調査する必要がある。

② プログラムニーズアンケート

事前事後テストで推定した受験者の能力値 θ の変化量を従属変数とし、プログラムニーズアンケートの回答結果を独立変数とした強制投入法に基づく重回帰分析を行った。偏回帰係数の推定値の絶対値は、全体的にあまり顕著なものではなかったが、年度 01、02 ともに、質問 13 「先生が授業を楽しくしてくれるとよく勉強できる。」については、能力値が伸びた学生ほど、この質問に同意している傾向を示した。この傾向は、先行研究 Kikuchi (2005) の結果とも一致する。また、変数増加法による独立変数の選択を行った結果、質問 17 「外国人の先生から英語を学ぶのが好きだ。」も両年度を通じて能力値の伸びを予測する傾向を示した。

年度別に見ると、質問 8 「ひとつのスキルだけ (「リーディング」ならそれだけ) を学ぶ授業が良い。」に対して、年度 02 では負の偏回帰係数の推定値を示しており、 θ の伸

びが大きかった学生ほど、質問8に対してネガティブな反応を示している可能性がある。これは年度01と年度02の間でカリキュラムを変更し、以前はスキル別であった授業の形態を、年度02から2スキル合併型(例:リスニング&スピーキングコース)に変えていることと関係があるかもしれない。今回のプログラムニーズアンケートは、回答のための時間が10分以内であったこともあり、十分な情報が得られなかったが、そのプログラムを履修して能力値が伸びた学生の声を聴くための手段としてプログラム評価上意味のある証拠となる可能性はある。

(4) まとめと今後の研究

本研究では、2学年分の実際の事前事後リスニングテストのデータを使ったために、これをすぐに一般化して結論づけることは難しい。しかし、テストの等化を実行するとき、どの項目応答モデルを選択するかは分析結果に影響をもたらすと言える。この結果からも、テスト分析者は安易に分析を実行せず、データと分析の目的を良く確認してから、用いる項目応答モデルと等化方法を選択するべきだと言うことを再認識した。しかし、先行研究にもあったように等化方法の比較研究は、シミュレーション研究、事例研究、どちらか一方だけでは不十分で両方の研究を合わせて実施するべきである。今後は、本研究のデータでシミュレーション研究を実施する必要がある。

本研究が、労力を省きながらも少しでも有効性が高く、効率の良い「プログラム評価のための事前事後テストシステム」を教育機関が取り入れるひとつの例となれば幸いである。

5. 主な発表論文等

〔雑誌論文〕(計 9件)

- ① プログラム評価のための最適項目応答モデルと等化方法の選択。
藤田智子・前川眞一、『研究紀要 J L T A (日本言語テスト学会) Journal』第15号、(2012年) 153-171. 査読有
- ② 英語教育における習熟度レベルと Can-do statements.
藤田智子、Chapter3. (3) Can-do statements における規準設定、77~85. 『言語テストの規準設定』日本英語検定協会 英語教育センター委託研究 2012 年度報告書
- ③ Can-do-statements が英語の授業において果たす役割。
藤田智子、Chapter3. (4) Can-do statements における規準設定、86~94. 『言語テストの規準設定』日本英語検定協会 英語教育センター委託研究 2012 年度報告書

- ④ IRT によるテストの規準設定
藤田智子、Chapter4. (1) テスト理論 (IRT と NTT) と規準設定、95-103. 『言語テストの規準設定』日本英語検定協会 英語教育センター委託研究 2012 年度報告書.
- ⑤ テストによる規準設定の実践: プレースメントテストの開発
藤田智子、Chapter4. (2) テスト理論 (IRT と NTT) と規準設定、104-112. 『言語テストの規準設定』日本英語検定協会 英語教育センター委託研究 2012 年度報告書.
- ⑥ A Comparison Between Common Item Equating with Pre-and Post-Reading and Listening Tests.
T.Fujita & S. Mayekawa, E-Learn 2011, 査読有、2011、26 - 32. Best paper award 受賞
- ⑦ The best IRT logistic model for a pre-and post-testing system with previous tests, T.Fujita & S. Mayekawa, SITE, 査読有、24-30、2011
- ⑧ .テスト開発における内容関連妥当性の証拠、藤田智子、『人事試験研究』、査読無、216、13、2010
- ⑨ Consideration of Test Validity and English Language Entrance Examinations in Japan, T.Fujita、東海大学外国語教育センター紀要、査読有、29、31-39、2010

〔学会発表〕(計 3件)

- ① A Comparison Between Common Item Equating with Pre-and Post-Reading and Listening Tests. T.Fujita & S. Mayekawa, E-Learn 2011, Hawaii US. Best paper award 受賞
- ② The best IRT logistic model for a pre-and post-testing system with previous tests, T.Fujita, paper presented at SITE 2011, Nashville Tennessee.
- ③ The effects of the arrangement of questions in can-do self-checklists on students' responses and their ease of use., T.Fujita, paper presented at the 32nd Language Testing Research Colloquium (LTRC) at University of Cambridge, 2010

6. 研究組織

- (1) 研究代表者
藤田智子 (FUJITA TOMOKO)
東海大学・外国語教育センター・教授
研究者番号: 80329002