

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4月1日現在

機関番号：15501
 研究種目：若手研究(B)
 研究期間：2010～2012
 課題番号：22720203
 研究課題名（和文） 適性試験型の日本語論述課題の開発と項目反応理論による尺度化・得点等化
 研究課題名（英文） A development and IRT scaling of academic aptitude test items for assessing Japanese language writing abilities.
 研究代表者
 大澤 公一 (OSAWA KOICHI)
 山口大学・大学教育機構・講師
 研究者番号：20555320

研究成果の概要（和文）：

本研究では、(1) 高等教育における学習意欲や適性、スキルとの相関関係、あるいはそれら自身の間接的な測定までを視野に入れた日本語論述課題の開発を行い、(2) 客観試験で測定される日本語能力との相関関係および採点者の評定バイアスの除去を考慮に入れた項目反応理論による項目反応データの尺度化の方法を提案した。研究の遂行に際しては、日本留学試験や日本語能力試験に代表されるような得点等化が必要となる年複数回実施の大規模試験における応用・実装可能性が重視された。

研究成果の概要（英文）：

In this study (1) academic aptitude test items for assessing Japanese language writing abilities were developed based on a survey of the correlations amongst motivation for learning, academic aptitudes and skills required in higher (university) education, and (2) psychometric method based on item response theory (IRT) to scale multiple ratings data with statistical adjustment of raters effects in calibrating examinees ability and item parameters was investigated through the attention of applying the results to large scale educational assessments such as EJU and JLPT requiring test score equating.

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	500,000	150,000	650,000
2012年度	1,500,000	450,000	1,950,000
年度			
年度			
総計	3,100,000	930,000	4,030,000

研究分野：日本語教育

科研費の分科・細目：言語学、日本語教育

キーワード：日本語論述試験、項目反応理論、得点等化、大学入学適性試験

1. 研究開始当初の背景

(1) 韓国語母語話者の日本語学習者数は世界最大の約91万人であり、世界の日本語学

習者全体の30.6%を占めている。学習者数の規模を反映し、韓国では大学修学能力試験（以下CSAT）、日本留学試験（以下EJU）、日本語能力試験（以下JLPT）といった外国語母

語話者の日本語能力を測定・評価するための大規模公的試験に対する需要が、ヨーロッパ諸国あるいは英語以外のインド・ヨーロッパ語族に属する諸言語を母語とする諸国に比べて相対的に大きい。

(2) CSAT, EJU, JLPT は実施の目的や受験対象者などがすべて異なるが、大学入試として、あるいは日本語能力の公的な認定試験として社会的な影響力が非常に大きい。しかしながら、CSAT と EJU に関しては試験の統計的な特徴が明らかにされておらず、利用者に対する説明責任が十分に果たされていない。JLPT に関しては上級レベルにおける口頭能力試験や論述能力試験の導入に対する社会的要請が大きいものの、2010 年に予定されている試験の改定においてはどちらも含まれていない。また、3 つの試験の難易度を直接比較することができず、不都合である。

(3) 近年の大学入学志願者層の多様化に伴い、大学入試においては受験者の学力や能力、適性やスキルなどの多面的な評価が求められるようになった。そのような多面的な評価を行う方法の一つとして、論理的思考力や推論・分析力、表現力、読解力などを総合的に測定・評価するための論述能力試験の重要性が高まってきている。

(4) (1)～(3)の社会的背景を踏まえ、既存の大規模公的日本語試験で測定される日本語能力と、日本語運用能力の自己評価尺度（日本語 Can-do Statements, 以下 CDS）や高等教育において必要とされる適性、スキルなどの自己評価との関係を明らかにするための大規模研究調査が実施された。その結果、CSAT, EJU, JLPT の難易度が初めて直接比較できるようになり、CDS や高等教育で必要な適性、スキルとの相関関係も明らかとなった。ここで各試験と高等教育の適性、スキルとの相関に着目すると、多肢選択式の客観試験で測定される日本語能力と適性、スキルとの相関が非常に低いことが明らかとなった。従って、近年の大学入試を取り巻く社会的背景(3)を鑑みても、受験者の学力以外の適性や資質、スキルなどを多面的に評価あるいは反映する、適性試験型の試験を開発しなければならないという研究課題が実証的に明らかになった。

2. 研究の目的

(1) 高等教育における適性やスキル（論理的思考力や推論・分析力、表現力、読解力などを含む個人の特性）との相関関係、あるいはそれら自身の間接的な測定を視野に入れた適性試験型の日本語論述課題を開発する。

(2) 客観試験で測定される日本語能力との相関関係および採点者のバイアス補正を考慮に入れた、項目反応理論による評定データの尺度化および得点等化法の数理統計学的な理論研究を行う。大学入試を想定した受験者の能力や適性などの多面的な測定・評価ツールを提案する。

3. 研究の方法

研究計画は下位計画(1)～(3)に分割され、並行して推進された。

(1) 多次元項目反応モデルを基礎とした評定者のバイアス補正を考慮に入れた評定データの測定モデル、および等化方法の提案を行う。（平成 22～23 年度）

(2) 現行 EJU 記述課題の解答データ分析を行う。先行研究で収集された現行 EJU の 400 字記述課題の答案を独自の採点基準によって採点・数量化する。EJU 読解領域（客観試験）との同時受験者データを利用して、EJU 記述課題が測定する言語能力と高等教育における適性、スキルとの相関関係を項目反応理論によってモデル化する。これを適性試験型の論述課題を開発するための基礎データとする。（平成 22～23 年度）

(3) 適性試験型の日本語論述課題および評定基準の開発、モニター調査の実施、尺度化および得点等化方法の実証データによる検証を行う。日本の大学入試で出題された小論文課題を収集し、(2)の結果を加えて高等教育で必要となる適性、スキルとの関連性を課題の内容・形式の両面から考察し、適性試験型の論述課題と評価基準を開発する。韓国および日本において新論述課題と EJU, JLPT の既出問題を同時受験してもらい、試験の潜在構造を統計モデル(1)によって分析する。（平成 22～24 年度）

[平成 22 年度]

下位計画(1)では文献・設備備品等の調達および環境整備を行い、統計モデルの理論研究を推進する。その際、国際的な大規模教育調査の経験が豊富な豪州の公的教育研究機関 ACER (Australian Council for Educational Research) の附属機関である Psychometrics Institute (心理・教育測定センター) からの助言を受けた。

下位計画(2)では EJU 記述課題の採点を行い、項目反応理論による試行的な分析を実施した。採点基準に関しては、現行 EJU あるいは平成 22 年の改定 EJU の採点基準を参考に、TOEFL など世界の大規模言語試験における論述試験の評価基準は分析的評価ではなく総合評価が主流であることも鑑みて、実際の

規模言語試験への応用可能性を重視した総合評価尺度の開発を進めた。採点基準となる評定尺度をより精緻化するために、当初採用する予定であった豪州 ACER で用いられている GAMSAT 評定尺度に加え、米国 TOEFL や英国 IELTS, GCSE, GCE 等で用いられている論述試験の採点方法に関する情報収集を行った。

下位計画(3)適性試験型の日本語論述課題の開発に関しては、日本の大学入試で過去に出題された小論文課題を収集する一方で、英国における公的試験 GCSE, GCE の日本語科目のサンプル項目を入手し、内容面の検討を行った。また、日本国内の研究協力者とのディスカッションを通して、モニター本試験で用いる修学能力試験 CSAT, 日本留学試験 EJU, 日本語能力試験 JLPT の既出項目からなる客観試験の構成・版組に関する事前調査に計画を前倒しして着手した。

[平成 23・24 年度]

下位計画(1)では研究協力者との議論を行いながら、前年度の下位計画(2)の分析結果も考慮して具体的な統計モデルの理論研究を推進した。評定データ用の項目反応モデルおよび得点等化法の研究については、研究協力者との議論を必要に応じて行いながら、多相評定者モデル (Multi-faceted Raters Effect Model; Linear Logistic Test Model, LLTM) および階層評定者モデル (Hierarchical Rater Model, HRM) の応用可能性について文献研究および理論研究を推進した。本研究においては最終的には大規模試験における応用可能性を重視するという立場を取るものの、現実的な要請として小サンプルの試験実施計画における

下位計画(2)では前年度の IRT 試行分析の結果を考察し、計画(1)および計画(3)にフィードバックを行った。

下位計画(3)では研究協力者の助言を受けながら、下位計画(2)の分析結果の考察と共に、収集した小論文試験の内容分析を行い、適性試験型の論述課題と採点基準を開発した。論述課題開発については、日本や海外の大学入試で出題された小論文試験や面接試験の課題、英国 Cambridge Assessment で開発・実施されている Oxford TSA (Thinking Skills Assessment) の項目などを参考として、平成 24 年度の海外モニター本試験に投入するための複数の日本語論述課題を準備した。

その後、論述課題と同時に受験してもらう客観試験の版を作成した。客観試験は、申請者の先行研究で項目反応理論による尺度化および等化が行われ、統計的な特性が判明している現行 JLPT の上位級および EJU の既出問題から構成された。計 395 題の問題項目データベース (文字・語彙, 文法, 読解領域) の中から、受験者の言語能力を特に良く識別

する優れた問題を 50 題抽出した (平成 23 年度)。平成 24 年度は、適性試験型の論述課題と JLPT, EJU の客観試験を用いて日本および韓国でモニター試験を行った。韓国では日本留学を希望する日本語能力中上級者を対象として調査を行った。

4. 研究成果

(1) 研究の主な成果

高等教育において必要とされる適性・資質、スキルと関連付けられた適性試験型の日本語論述課題として 8 題の論述課題をモニター試験に投入した。採点基準は大規模試験への応用を重視した総合評価尺度を用い、テストデータの尺度化に用いる項目反応モデルは多次元ランダム係数多項ロジットモデルを拡張した多次元多相評定者モデル、および素点での評定尺度上で評定者効果をモデル化する階層評定者モデルの 2 種類によって分析を行った。

分析の結果、数値計算にマルコフ連鎖モンテカルロ法を用いる階層評定者モデルは大規模試験における採点には適さないことが分かり、ロジット尺度上で受験者の能力や項目困難度、評定者効果を分離する多相モデルによる分析が実用上は効果的であることが示唆された。

(2) 得られた成果の国内外における位置づけとインパクト

これまで、高等教育における適性やスキルとの関係に着目した適性試験型の論述能力試験の研究開発が、既存の公的試験を対象とした大規模な実証的データ分析の結果に基づいて行われた例は見られない。本研究の最大の特徴は、既存の大規模公的試験の統計的な特徴を項目反応理論による統一的な枠組みで表現した上で、客観試験の延長線上に位置づけられる発展的上位試験としての新しい論述能力試験の構築を目指しているという点にある。

本研究では、研究成果を評価するための参照基準 (大規模試験のモニター調査に基づいた客観データに基づく参照基準) が明確にされている。その結果、手探り状態で全くの新しい試験を単発的に開発して終わるのではなく、研究の推進過程において EJU や JLPT に関する統計的フィードバックが得られたことも研究成果の一つである。今後、研究結果は広く社会に公開されるため、試験の統計情報開示という公的試験が持つ社会的説明責任の一端が間接的にでも果たされることになるだろう。最後に、本研究が今後の大規模日本語試験における論述課題導入のための、そして日本の大学入試改善のための基礎研究となることが期待されるものである。

(3) 今後の展望

大規模試験における実装可能性を重視する限りにおいては、評定データに適用する項目反応モデルは LLTM 型の多相評定者モデルが（得点等化を別途に行う必要が出てくるとい可能性を考慮すると）現実的である。しかし、多相モデルでは評定者内信頼性をモデル化することができないほか、実践的なテスト実施デザインを考慮に入れた評定者特性の構造をモデル化する必要があるため、理論的には階層評定者モデルの実用性をより高めるための基礎的研究が望まれる。

階層評定者モデルの理論的展開としては、評定者（集団）の背景や受験者の属性、あるいはテスト項目（群）の特異な性質、時系列的な観点から見た評定者のパフォーマンス（例：評定基準の日内変動）などが考えられるだろう。柔軟なモデル構成が許されれば、個別のテスト場面だけではなくテスト実施前の評定者トレーニングなどの場面においても有効な効果測定ツールとして利用することができるだろう。合わせて、大規模データに対して複雑なモデル構成でパラメタ推定を効率的に行うことができる数値計算方法（アルゴリズム）の研究を進めていくことが今後の課題として挙げられる。

本研究において開発された適性試験型の日本語論述課題については、測定対象となる能力が比較的高次の論理的思考能力を前提とした論述能力であった。高等教育において必要となる適性・資質との連携を取るとい観点からは課題内容は適切であったと考えられるが、今後は日本語での発信能力の水準が中級程度の学生にも適した難易度の課題を開発するとともに、分析に際して受験者の属性の偏りによるパラメタ推定のバイアスが発生しないよう、広範な国・地域におけるモニター試験を実施して実証的なエビデンスを積み重ねていくことが求められる。

採点基準については、本研究においては総合評価を念頭に置いた評価基準を開発したが、日本語学習の現場においては形成的評価を行う流れの中での診断的・分析的評価を行うことに対する要望が大きいのと思われる。今後は、診断的評価に基づく連続測定の実現するための評価基準（完全習得地図）の研究を進めていくことが必要である。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）

- ① 大澤公一，テストの科学と大規模公的日本語試験に果たす役割 — 学習者の日本語能力の育成と評価のために —，北京師範大学外文学院日本語教育研究所第

二回シンポジウム論文集，招待論文，vol. 5，2012，pp. 26-37.

- ② 大澤公一，日本語能力における非音声領域の客観的測定および言語運用能力の主観的評定に共通する潜在特性の項目反応モデリング，日本語教育，査読有，150号，2011，71-85.

〔学会発表〕（計 1 件）

- ① 大澤公一，高等教育で求められる適性や資質・スキルに関する実証的研究 — 韓国語母語話者による自己評定データの項目反応モデリング —，日本教育心理学会第 54 回総会，2012 年 11 月 23 日～25 日，琉球大学。（沖縄県中頭郡）

6. 研究組織

(1) 研究代表者

大澤 公一 (OSAWA KOICHI)
山口大学・大学教育機構・講師
研究者番号：20555320