

科学研究費助成事業 研究成果報告書

平成 29 年 8 月 16 日現在

機関番号：14301

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25770190

研究課題名(和文)日本語論述能力の連続測定モデルに基づく診断的教育測定エンジンの基礎的研究

研究課題名(英文)A development of diagnostic psychometrics engine based on continuous measurement model of Japanese language writing abilities

研究代表者

大澤 公一(Osawa, Koichi)

京都大学・高大接続・入試センター・特定准教授

研究者番号：20555320

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：日本語論述能力を対象とし、次世代型の教育測定モデルである「連続測定」の概念に基づいた診断的教育測定エンジン開発のための基礎研究を行った。代表者の先行研究の成果である大規模データを基礎として、(1)学習者が「日本語を書く」単元において習得すべき事項を系統立てた「完全習得地図」を構成する。(2)この地図に基づいて学習者の学習の軌跡と能力プロファイルの発達的变化を連続的に測定するカリキュラム一体型の教育測定モデルを提案する。その際、(3)診断的な教育評価を行うことで単純な能力測定に止まらず、測定がそのまま未習得部分の効率的な学習に繋がるような学習支援にまで踏み込んだ測定モデルの開発を試みた。

研究成果の概要(英文)：In this study a diagnostic psychometrics engine based on continuous measurement model of Japanese language writing abilities was developed. Based on the results of formerly conducted large-scale survey, (1) a mastery map that systematically describes and defines the items Japanese language learners are supposed to gain in the field of writing in Japanese was constructed. (2) Based on the mastery map, a psychometrics model was proposed to assess continuously developmental changes in their ability profiles and learning path. (3) By employing diagnostic assessment the model made not just an psychometrics engine for ability estimate but rather an educational assessment model that directly links measurement and effective learning of students' unachieved abilities.

研究分野：教育心理学

キーワード：日本語論述能力 連続測定モデル 診断的評価 項目反応理論 得点等化 大学入学適性試験 高等教育

1. 研究開始当初の背景

本研究の社会的背景として以下の(1)(2)の状況が挙げられる。

- (1) 日本語学習の目的は高度化・多様化を続けている。それに伴い、日本語能力の測定評価を行う大規模公的試験においては「読む・書く・聴く・話す」4技能の試験科目の提供が社会的に強く求められている。しかし日本語能力試験(JLPT)では口頭・論述試験が実施されず、日本留学試験(EJU)では測定範囲の狭い古典的テスト理論による論述試験の提供に止まっている。
- (2) 項目反応理論(IRT)やテスト等化法に代表される現代テスト理論が実用の段階に入り、EJUやJLPTでも実装されている。今後はコンピュータ適応型テスト(CAT)への進化や、テスト得点の質的解釈といった教育測定としての理論的・実践的精緻化が期待される。

社会的背景(1)(2)を踏まえ、文部科学省(大澤, 2007~2008)および科学研究費補助金(大澤, 2010~2012)の助成を受け、大規模公的日本語試験で測定される日本語能力を客観的な参照基準とした新しい日本語論述試験の開発、および項目反応理論による尺度化・得点等化法の開発を目的とした大規模研究調査が実施された。日本語学習者数の多い韓国、中国、香港、インドネシアで合計5,500人以上のサンプルが収集されデータ分析が行われた。また、上記の研究では新しい日本語論述試験の妥当性を検証するために高等教育において必要となる適性・資質、スキルや言語運用能力4技能に関するCan-do Statements尺度との相関関係を明らかにするための調査データや、既存の大規模公的日本語試験の既出問題に対する解答データも同時に収集されている。その結果として、論述試験の単純なCAT化を超え、次世代型の連続測定モデルによる日本語論述能力の診断的評価に挑戦するための基礎研究データがある程度整備された。

2. 研究の目的

日本語論述能力の測定について、次世代型の教育測定モデルである「連続測定」の概念に基づいた診断的教育測定エンジンを開発するための基礎研究を行うことが本研究の目的である。大澤(2007~2009; 2010~2012)の研究プロジェクト成果である大規模データを基礎として、以下の3点に関する下位研究を設定する。

- (1) 学習者が「日本語を書く」単元において習得すべき事項を系統立てた「完全習

得地図」を構成する。日本語教育学的には論述能力の学習内容かつ評価項目(採点基準)を設定することに相当し、テスト理論的には潜在能力次元の数と相関関係を事前定義することに相当する。

- (2) 完全習得地図に基づいて学習者の学習の軌跡と能力プロファイルの発達的变化を連続的に測定する、カリキュラム一体型の教育測定モデルを提案する。先行研究において考察された多次元・多相IRTモデルに時系列解析と認知診断モデルの概念を取り入れる。
- (3) その際に診断的な教育評価を行うことで、単純な能力測定に止まらず、測定がそのまま未習得部分の効率的な学習に繋がるような、学習支援にまで踏み込んだ教育測定モデルを構築する。完全習得地図の学習・評価項目の達成水準に対して、合否分割点法(standard-setting method)を応用して定量的なテスト得点を質的な基準に対応付ける。これによってテスト得点に具体的な意味が付与され、学習者の能力の診断的な評価(能力プロファイリング)が可能となる。

3. 研究の方法

本研究は下位研究(1)~(3)より構成される。各下位研究は並行して推進されるが、初年度は先行研究データの再分析および理論研究に基づく(1)完全習得地図の項目構成に、次年度は先行研究を拡張した認知診断的・縦断的解析を行うための(2)教育測定モデルの開発に、最終年は日本語論述能力のモニター試験を実施し、実際の時系列データに対して縦断的・認知診断モデルによる能力プロファイリングを行う(3)診断的能力評価に重点が置かれている。

本研究は申請者の大規模調査研究を発展させるものであり、基礎分析に用いるデータセットや本試験のためのテスト項目は確保された状態にある。計画推進上の課題点は如何にして妥当な完全習得地図を構成できるかにあるが、そのための日本語教育学、テスト理論、教育評価、教育心理学に関する研究協力者、および公的研究機関のネットワークが国内外に整備されている。

[平成25年度]

下位研究(1)では、代表者の先行研究で収集された高等教育において必要となる適性・資質・スキルや言語運用能力4技能についての自己評定尺度(Can-do Statements尺度、CDS)の調査データに対して潜在変数モデル(因子分析、共分散構造分析)による探索的データ解析を適用する。言語論述能力を構成・説明すると考えられる因子を中心にモデルを構成し、完全習得地図のベースとなる学

習・評価項目を探索的に抽出する。その一方で、日本語教育や言語教育に関する文献研究を推進し、国内外の研究協力者の助言を得ながら学習・評価項目の精緻化を行う。

下位研究(2)では、認知診断モデルおよび縦断データ解析モデルに関する文献や先行研究を入手し、理論的研究を推進する。項目反応モデルに関する実践的な支援として、豪州の公的教育研究機関 ACER (Australian Council for Educational Research) の付属機関である Psychometrics Institute (心理・教育測定センター) からの助言を受けた。

下位研究(3)では、合否分割点法 (standard-setting method) に関する諸文献を入手し、理論的研究を推進する。検討対象とした手法は以下の通りであった。Angoff, Extended Angoff, Ebel, Nedelsky, Jaeger, Bookmark & item mapping, Direct consensus, Contrasting groups, Borderline group, Item by Item (paper selection), Holistic (body of work, booklet classification, generalized examinee centered), Hybrid (analytical judgement, integrated judgement, work classification), Judgemental policy capturing, Dominant profile, Item cluster, Beuk, de Gruijter, Hofstee.

[平成 26・27 年度]

下位計画(1)では、前年度末に構成された完全習得地図の項目構成を下位研究(2)に反映する。その後は適宜文献研究を継続して行い、研究協力者の助言等も得ながら最終的な項目構成の完了を目指す。

下位研究(2)では、研究(1)の結果(潜在能力次元の数および相関関係の構造)を認知診断モデルにおける Q 行列の形で数量的に表現する。その後、モデルに投入する母数を確定させ、数値計算アルゴリズムや等化法などの検証を行う。先行研究で得られているデータに対して推定実験を行う。完全習得地図が確定した後に、最終的な Q 行列を確定させる。モニター本試験データにモデルを適用し、認知診断および時系列解析を行う。

下位研究(3)については、平成 26 年度は研究(1)の結果と連動して採用候補となる合否分割点法の候補を選定し、先行研究データに試験的に適用する。データ分析結果からモニター試験で採用する手法に精緻化する。平成 27 年度は日本において先行研究で得られた新しいタイプの日本語論述試験を用いたモニター試験および質問紙調査を行う。下位研究(2)で得られた診断的・縦断的データ解析を行い、連続測定モデルの試作版を構成する。

4. 研究成果

[平成 25・26 年度]

下位研究(1)については、先行研究データの再分析および理論研究に基づく完全習得地

図の項目構成を行った。言語論述能力を構成する要素を中心とした学習・評価項目群の構成を目指したが、主として日本語教育における「書く力」の教育課程に関連した学習・評価項目を取り入れながら項目の精緻化を進めた。学習者が「日本語を書く」単元において習得すべき事項の体系化を進める一方で、実際のテストにおける評価ルーブリック(採点基準)の試作を進めた。その際は、日本語論述能力の中・上級レベルの能力者を識別するような評価項目の構成を目指した。

下位研究(2)認知診断的・縦断的データ解析を行うための教育測定(数理統計)モデルの開発に関しては、項目間多次元性を基礎とした項目反応モデルの拡張型を中心に理論的研究を進めた。特に、モデル母数の周辺最尤推定問題やマルコフ連鎖モンテカルロ法によるパラメタ推定に関する文献研究を進めた。また、先行研究データに対する因子分析的研究を進め、言語論述能力と相関関係の見える言語能力因子の同定に関する実証的研究を進めた。

[平成 27 年度]

連続測定モデルにおいて学習者が「日本語を書く」単元において習得すべき事項を系統立てて記述した学習事項および評価基準である完全習得地図の構成については、前年度までに作成してきた素案に対して日本語教育学の観点から研究協力者のアドバイスを随時得ることと平行し、日本語教育や英語教育に関連する理論研究および文献研究を追加して実施した。それらを踏まえ、モニター試験で用いる完全習得地図の素案を構成した。この素案に基づき、完全習得地図の評価基準項目を認知診断的・縦断的評価を行うための教育測定モデルである多次元多相項目反応モデルに反映させるための Q 行列の構成をモニター試験に投入するテスト項目の選定と平行して進めてきた。数理モデル自体に対しても、テスト理論に関する理論研究やシミュレーション実験、文献研究を随時進めることによって、より精緻化されたモデル構成を行うべく主としてモデル母数の設定について様々な観点からの検討を行ってきた。

本来であれば、平成 27 年度(研究計画最終年度)中に診断的能力評価法の研究として日本、韓国、中国、香港、インドネシアにおいて代表者が作成した新しいタイプの日本語論述試験項目を用いたモニター試験および質問紙調査を行い、認知診断的観点からデータ解析を行う計画となっていた。しかしながら、代表者の一身上の理由により平成 27 年度下半期以降の研究活動続行が一時的に不可能な事態となり、モニター試験の実施を含めた研究期間を一年間延長し、翌平成 28 年度に再開することとした。本調査の実施延期に伴い、調査対象国の研究協力者との各種調整を行った。

[平成 28 年度]

日本語の論述能力に関する学習内容および測定・評価項目のマッピングに相当する完全習得地図の構成に関しては、本研究期間中に議論が進んだ日本の大学入試改革との関連性・整合性を担保するために、より教育的な構成概念を包含するものとなるよう構成概念に変更を加えた。研究計画立案当初に想定していた比較的単純な日本語学、日本語教育学的な「ことば」に焦点を置く項目観点に留まらず、一般的・広範的な意味での教育測定学と高大接続との関連性を考慮したうえで、言語としての日本語が持つ諸相（文字・語彙、文法、表現形式など）だけでなく、大学などの高等教育機関に進学して自身の専門分野を効率的に学習していくために必要とされる資質について、中等教育課程における学習教科・科目間に共通して獲得されることが期待される能力や技能を整理して分析の評価項目としてリスト化を試みた。

日本語論述能力を客観的に尺度化するための教育測定モデルに関しては、潜在変数モデルを基礎とし、多次元項目反応モデルの適用を主眼に置いて検討を引き続き行ってきた。将来的には個別の問題項目の持つ能力識別情報も母数化する方向で研究を進展させたいと考えているが、本研究においては識別情報の差異をモデル母数として導入することは困難であった。多次元項目反応モデルにおける識別力母数の解釈が必ずしも容易ではないため、適切な評価項目数の設定基準や反応データが持つ情報水準の減少バランスなど、今後の研究における指針をいくつか得ることができた。

診断的能力評価に関しては、standard setting method の比較実証研究を十分な質・量を保証したパネル調査をもって行うことができず、数量的な能力値の質的な意味付けの付与に関しては更なる基礎的研究の必要性が明らかとなった。本研究の技術的な側面に関連していえば、項目反応モデルの母数と分割点（cut point）の対応付けを行うという方針に則ると Bookmark & item mapping の手法を多次元・多値型のテスト得点プロファイルに拡張することが自然であると考えられる。

一方で、本研究期間中に研究活動の中断をはさみ、本務校に復職後の一か月後に所属大学が移動となるという研究環境の大幅な変化が研究最終年度にあり、当初に計画していた国外でのモニター試験を実施することができなかった。国内の調査に関しても代表者の前任校で行った調査に留まっている。この点に関しては、高等教育における学習活動の連続測定モデルを提案するという研究目的に照らし合わせると、所属大学が変更となったことにより教育・学習成果のポートフォリオ管理の方針や手法、さらには所属学生に期待される能力・資質の種類や水準も異なるものとなっていることから、完全習得地図の構

成から修正する必要が出てきている。また、診断的評価を実践的に運用するプラットフォームも新たなものを構築する必要がある。このような状況を鑑みて、海外におけるモニター試験の実施内容および調査対象者も再考する必要が出てきている。

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 1 件)

伊藤理英・大澤公一、大規模言語テストにおける項目解答行動の分析の試み - 韓国語母語話者を対象として - , 関西外国語大学研究論集, 査読有, vol.100, 2014, pp.97-113.

〔学会発表〕(計 1 件)

Osawa, K. & Ito, R. (2014). A trial study of test taking behavior in large-scale language tests. - Analysis of responses from Japanese language learners in Korea. Paper presented at International Conference on Japanese Language Education (ICJLE) 2014, University of Technology, Sydney, AU.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

該当なし

6. 研究組織

(1) 研究代表者

大澤 公一 (OSAWA, Koichi)

京都大学・高大接続入試センター・特定准教授

研究者番号：20555320

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし

(4) 研究協力者

伊藤 理英 (ITO, Rie)

関西外国語大学・准教授

李 庸伯 (LEE, Yong-Baek)

韓国教育課程評価院・研究委員