

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 20 日現在

機関番号：82616

研究種目：若手研究 B

研究期間：2010 年度～2012 年度

課題番号：22700303

研究課題名（和文）

テストの科学的運用のためのテストデータ用分析システムの構築

研究課題名（英文）

System development for educational test data analysis

研究代表者

大久保智哉（Okubo Tomoya）

独立行政法人 大学入試センター 研究開発部 助教

研究者番号：80512136

研究成果の概要（和文）：

本研究は、高まりつつあるテストへの需要とその科学的運用への期待を背景に、テストデータを項目反応理論に基づいて容易に分析できるプログラムを研究・開発することを目的としてすすめられた。本研究において、1. 項目反応理論における項目パラメタのパラメタ推定を可能とするプログラムを開発した、2. 多変量型のベータ分布を定義し、その有限混合分布を多変量項目反応理論における潜在特性値の分布として採用することを提案した。3. 推定された項目パラメタに基づいて、コンピュータ型テストを実行できるテストシステムを開発した。

研究成果の概要（英文）：

In this research, the system for educational testing data was developed. The developed system enables us to estimate parameters of various types of items based on Item Response Theory which is widely used in psychology and educational psychology. In this research, 1. The parameter estimation programme for Item Response Theory was developed. 2. Multivariate-beta distribution was formulated, and the finite mixture distribution of the distribution was also proposed, which covers entire range of the multivariate testing scores. 3. The Computer-based Testing system has been developed, which shows items to respondents via WEB server, and stores responses of the respondents.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	1,000,000	300,000	1,300,000
2011 年度	1,200,000	360,000	1,560,000
2012 年度	600,000	180,000	780,000
総計	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：情報学／統計科学

キーワード：行動計量分析

1. 研究開始当初の背景

近年、テスト(試験)に対する期待と重要性は増している。わが国においては2007年より全国学力・学習状況調査が実施され、米国では2002年からThe No Child Left Behind法のもと、3年次から8年次までの生徒全員に対してReadingとMathの試験が毎年実施されている(2007年からはScienceも含まれる)。米国におけるテストの特徴として、項目の作成・テストの編集・分析結果の統計処理・事後分析など全ての段階でテスト専門家が携わっている点が挙げられる。

現在、わが国では大学入試センター試験、法科大学院適性試験、全国学力・学習状況調査など大規模テストが実施されている。これらは、項目ごとの配点をつけ、その合計得点を学力とするいわゆる「素点主義」によってテストが運用されている。しかしながら、

1. 試験結果の複数年度利用の推進
2. テストの複数会場、複数時期での実施と受験機会の拡大

などについては、テスト得点が相互のテスト間で比較可能であることが前提条件となるため、現在の「素点主義」のもとでのテスト運用では実現が不可能である。この「素点主義」で得点を比較するには「全受験者に対して同一の問題から構成されたテスト」という条件を満たす必要がある。一方、海外ではすでに多くの公的試験において項目反応理論(Lord, 1968)とよばれる統計モデルに基づきテストを科学的に運用しているため、素点主義の限界を乗り越え、受験機会の拡大やテスト結果の経年比較が可能になっている。わが国では、医学部共用試験 CBT など、一部では項目反応理論を用いたテストの運用が広まっているが、未だ発展途上といっていよいよ。

テストの品質保証のためには、統計理論に基づいた能力測定と項目分析が不可欠である。テストが測りたいものを(妥当性)正確に(信頼性)効率よく(有用性)測定できているのか?という点を追求する必要がある。海外ではテストの品質を高めるために項目反応理論がすでにテストの標準理論として用いられるようになってきている。この項目反応理論では、異なる時期に実施された異なる問題から構成されるテストの得点を比較可能にするというメリットがある。このことは「全受験者に対して同一の問題から構成されたテストをおこなう」という厳しい制約から脱却できることを意味している。したがって、項目反応理論によるテストの運用はテストの運用や分析の幅が圧倒的に増えることを意味している。

しかし、わが国においては、テスト理論の普及が十分ではない。その背景には

1. わが国がもつ大問形式などの独自の出題形式

2. テスト運用に関する専門家の不在などが指摘されている(石塚, 2003)。

大問形式とは、同一のトピックに関して続けて複数個の問題を設定するような形式である。この形式の特徴として大問内の連続する項目において、前の項目が不正解だとそれ以降の項目も不正解になりやすくなるという測定の独立性が保証されていない点を指摘できる。項目反応理論は、測定の独立性を前提とした上でパラメタを推定するため、大問形式による出題で項目間の独立性が保証されない場合には推定量にバイアス(偏り)がかかるという大問題がある。一方で、大問形式はその文脈効果によって認知的に高次な能力を問うことができるというメリットも報告されている(Wainer & Kiely, 1987)。

さらに、テスト理論に対する認知度が海外に比して圧倒的に低いことを指摘する研究者もいる。たとえば、項目反応理論を扱うことのできるソフトウェアは外国製の商用で、高額な上、英語による説明しかないことが、現場のテスト関係者の理解・使用を阻害しているという指摘がある。さらに、実際に利用が容易な環境にないことがテストの科学的運用を阻害していると考えられる。

2. 研究の目的

本研究の目的は、高まりつつあるテストへの需要とその科学的運用への期待を背景に、テストデータを統計理論に基づいて容易に分析できるプログラムを開発し、WEBサーバ上で自在にテスト関係者が使用できるシステム環境を構築することである。WEB上で簡単に操作することのできる環境を構築することによって、大学関係者のみならず多くのテスト関係者が、テストを選抜のためのものではなく、教育に資する情報を提供するものとして認知し、実際にそう利用できるようになることを期待している。併せて、項目反応理論の理論的拡張も試みる。

3. 研究の方法

本研究は、3か年の計画に基づいておこなわれる。まず、平成22年度には、システム的设计図を詳細に確定させる。また、テストデータ分析のための項目反応理論の基幹プログラムを作成する。また、平行して項目反応理論の理論的な拡張をおこない、より実データにフィットするように改良することもおこなう。次に、平成23年度には、ユーザーインターフェイスの開発を開始する。そして、最終年度にあたる平成24年度には、協力者

にシステムの評価を依頼し、その結果に基づきシステムを改良する。具体的には、それぞれの年度における目標と計画を下記のように定める。

具体的な平成 22 年度には以下の研究内容を計画した。

1. システム仕様の策定
2. 項目反応理論に基づいた項目パラメタの推定プログラムの作成
3. ベータ分布の拡張による項目反応理論の理論的拡張

まず、平成 22 年度においては、システムの仕様を策定する。研究では WEB 上にサーバを構築し、WEB アプリケーションの形でシステムを利用する。なお、WEB 上にサーバを構築することによって、セキュリティの点で心配を持つ利用者がいることが想定されるため、セキュリティは高度な設定にする。また、項目反応理論に基づく項目パラメタの推定プログラムは、C++言語を用いて記述する。統計モデルの実装は通常多くの時間がかかるが、申請者がすでに研究・開発をおこなっている R 言語によるプログラミングを移植する形をとる。さらに、項目反応理論の理論的拡張については、多変量に拡張された項目反応理論の今後の普及を前提に、潜在特性値の(多変量)分布の表現をより柔軟にするために、ベータ分布を基底に新たな多変量潜在特性分布の提案をおこなう。このことにより、対象とするデータによってはより適切な潜在特性分布を想定できるようになる。

次に、平成 23 年度には以下のような計画とする。

1. ユーザーインターフェースの開発
2. WEB 上にサーバを構築・試験的運用開始

平成 23 年度は、平成 22 年度におこなったシステムの基幹となる分析プログラムを使うためのユーザーインターフェースの開発をおこなう。この際、テストの実務家が使いやすさを念頭に、わかりやすく・使いやすいインターフェースの開発をおこなう。プログラムを完成させた上で、サーバを構築し、システムの試験的な運用に移る。なお、システムの試験的運用期間には、外部に対して公開するのではなく、一部の研究協力機関に試験的に扱ってもらうことを計画している。そこで、大きなシステムのバグやシステムの問題点を洗い出すことを試みる。また、研究成果については積極的に海外学会や国内学会において発表をおこなうことを計画とする。

最後に、平成 24 年度に計画しているのは以下の内容である

1. システムの本格的運用開始
2. システムの評価
3. 学会発表・シンポジウム開催等による広報活動

研究最終年度は、システムの本格的運用を

開始する。なお、十分な分析結果の信頼性とセキュリティ上の安全性を確認した上で公開する。また、システムの評価を複数の協力者・協力機関に対して依頼する。さらに、最終年度は国際学会、国内学会において研究成果を発表する。また、それらの評価の際に得られた結果をシステムに反映することを計画する。

4. 研究成果

本研究において、

1. 項目反応理論における項目パラメタの推定に際して用いることができる、プログラムを開発した。
2. さらに、多変量型のベータ分布を定義し、その有限混合分布を多変量項目反応理論における潜在特性値の分布として採用することを提案した。
3. 開発されたプログラムを用いて項目特性を推定された項目を用いて、実際にコンピュータ型テストを実行できるテストシステムを開発した。

まず、項目パラメタ推定用のプログラムでは、これまでに開発された多くの項目反応モデルを統合的に利用できるようなものが開発された。具体的には、Graded Response Model (Samejima, 1969), Generalized Partial Credit Model (Masters, 1982; Muraki, 1992), Nominal Categories Model (1972), Ordered Nominal Categories Model (Okubo & Mayekawa, 2009)などの項目反応モデルが項目ごとに設定が可能となっている。さらに、これらの項目反応モデルを用いて項目パラメタを推定する際には、パラメタ等値の制約や、パラメタを固定パラメタとして設定するなどの詳細なモデルの設定をおこなえるようになっているのが本プログラムの特徴となっている。加えて、多母集団同時分析と有限混合分布を母集団に設定できることがより詳細な分析を可能にしている。これらの研究成果については、国際学会等において報告された。

さらに、将来的に多変量項目反応理論が用いられるようになることを想定し、能力分布として従来用いられている多変量正規分布のみならず多変量ベータ分布を能力分布として採用できるよう、項目反応理論の理論的な側面の研究をおこなった。さらに、提案された分布の有限混合モデルへも拡張がおこなわれた。これにより、多変量項目反応理論についても将来に実装が完了した際には、想定する項目特性分布に多変量正規分布のみならず、多変量ベータ(の混合)分布を指定できるようになる。このことは今後、状況に応じて適切な能力分布を選択できるようになることと期待されている。

そして、従来の計画にはなかったものの、本研究においてネットワークを利用した Computer-based Testing を可能にするテストインテグレーションシステムを開発した。このシステムでは、項目特性値や項目に関する情報をデータベースに保持し、受験者の反応とその項目の特性値をもとにして、次の項目を選択するような適応型テストの利用をも可能にしている。この WEB 型の Computer-based Testing の普及は、本研究において開発された項目反応理論のための項目パラメータ推定プログラムの普及を手伝うものであると同時に、それ自体が本研究の目的にある「項目反応理論を用いたテストの普及」に効果的に役立つものであると評価している。

今後の研究課題の一つは、多変量項目反応理論をシステムへ実装することである。多変量項目理論とは、現在多く利用されている項目反応理論の多変量拡張版である。今後の研究の展開次第では多く利用される現場が増えると思われる。また、本研究では項目パラメータの推定に焦点が当てられたが、今後は推定された後の項目パラメータを他の集団に準じて推定された項目パラメータと等化する際に利用者をサポートできるような機能も実装していく予定である。

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. Okubo, T. (2012). Computer-based testing at ACT. *DNC research note. RN-12-01.*
2. Okubo, T., & Mayekawa, S. (2011). Modeling for applying mixed-multivariate beta distribution to discrete score distribution. *DNC research note. RN-11-03.*

[学会発表] (計 9 件)

1. Okubo, T. (2012). Some extensions of a test data analysis platform based on Multi-dimensional Item Response Theory. *20th International Conference on Computational Statistics (COMPSTAT 2012)*. Cyprus.
2. Okubo, T. (2012). Applying mixed-multivariate beta distribution to item response theory. The 77th Annual Meeting of the Psychometric Society. Nebraska.
3. Okubo, T., & Mayekawa, S. (2011). Fitting mixed multi-dimensional beta

distribution to scored data. The 76th Annual Meeting of the Psychometric Society. Hong Kong.

4. Okubo, T., Hiramura, T., & Mayekawa, S. (2011). Development of a web-based integrated platform for test analysis. The 58th World Statistics Congress of the International Statistical Institute. Dublin.
5. Okubo, T., & Mayekawa, S. (2011). R-based programme for integrated models in item response theory. International Conference on Mathematics & Statistics 2011. Athens.
6. Okubo, T., & Mayekawa, S. (2011). Applying mixed-multivariate beta models and log-linear models to discrete test score distributions. The 4th International Conference of the ERCIM WG on computing & statistics. London.
7. Okubo, T., Okada, K., & Mayekawa, S. (2010). A Bayesian solution for a mixed vector model. The 75th Annual Meeting of the Psychometric Society. Georgia.
8. Hiramura, T, Okubo, T., & Mayekawa, S. (2010). Development of Web-based integrated test analysis platform. Computational Statistics 2010. Paris.
9. Murata, I., Okubo, T., & Muta, H. (2010). The international comparative Study of Relationship between activities other than school lessons and mathematics achievement: Using R-based programme for nominal categories model. International Congress of Applied Psychology 2010. Melbourne.

[図書] (計 2 件)

1. 大久保智哉. (2011). テストの標準誤差. 統計応用の百科事典. 松原望(編). pp. 380-381. 丸善出版.
2. 大久保智哉. (2011). ベイズ分析における P 値とその他の証拠の指標の役割. ベイズ統計分析ハンドブック. pp. 156-175. 朝倉書店.

6. 研究組織

(1) 研究代表者

大久保智哉 (Tomoya Okubo)

研究者番号 : 80512136