

## 科学研究費助成事業 研究成果報告書

平成 28 年 5 月 11 日現在

機関番号：13101

研究種目：基盤研究(C) (一般)

研究期間：2011～2015

課題番号：23501139

研究課題名(和文) 分割統治法を応用した新しいリンク技術の確立に関する基礎研究

研究課題名(英文) A Basic Study on Development of Novel Linking Techniques Using Divide and Conquer Algorithm

研究代表者

佐藤 喜一 (Sato, Yoshikazu)

新潟大学・教育・学生支援機構・准教授

研究者番号：00300517

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：等化の手法を用いると、形式的には任意の二つの異なるテストの対応づけ(linking)が可能である。しかしながら、意味ある対応づけを保証するには、テスト得点だけでなく対応づけ得点にもある程度の高い信頼性が必要である。本研究では、単一グループデザインのもとでの線形対応づけに焦点を絞り、対応づけ得点のための測定の標準誤差と信頼性指数を提案した。さらに、実際の大規模テストの結果を利用し、信頼性の観点から線形対応づけの対応づけ可能性を検討した応用例を示した。

研究成果の概要(英文)：Any two different tests can be linked formally by applying equating methods. However, not only test scores but also linked scores should have a high degree of reliability to ensure meaningful linking. In this study, we focus on linear linking under single group design and propose a standard error of measurement and an index of reliability for linearly linked scores. Furthermore, we present applied examples that examine the linkability of linear linking from the perspective of reliability with the results of actual large-scale tests.

研究分野：教育測定学

キーワード：リンク 対応づけ 等化 古典的テスト理論 信頼性 線形等化法 対応づけ可能性

1. 研究開始当初の背景

本研究では、異なる二つのテスト間のスコアを結びつけることを対応づけ (linking) と呼ぶことにする。とくに、二つのテストが同一の構成概念を測定しているなどの条件を満たす場合には「等化」と呼ばれる。等化の例として、TOEFL の異なる二つの版 (同一テストの異なる版) の対応づけがあげられる。一方、実際のテストの運用場面では、上述した等化の枠組みに収まらないスコア間の対応づけが必要とされるようになってきた。その例として、法科大学院の入学者選抜に利用される実施主体の異なる二つの適性試験の対応づけがあげられる (柴山・野口、2004)。さらに教育現場でも、小中学校対象の全国学力調査と地方自治体における学力調査との対応づけなどに関心が高まりつつある (石井、2008 など)。教育現場の形成的評価をはじめ、実施主体の異なる試験から意味のある有益な教育情報を引き出すためには、対応づけの際の誤差評価を含めた包括的な対応づけ技術の確立が必要である。

2. 研究の目的

本研究は、等化を含む対応づけ技術の向上に資することを目的とし、既存の等化手法に分割統治法 (divide and conquer algorithm) の考え方を応用した新しい対応づけ技術を確立するための基礎研究を実施するものである。

3. 研究の方法

当初は、二つのテストを適当に分割して対応づけすると対応づけの標準誤差を減少させることができるという佐藤・柴山 (2010a) の示唆に基づき、対応づけの標準誤差に着目して研究を進めていた。しかしながら、初年度 (平成 23 年度) の研究発表等から新たな知見が得られたため、基礎研究の方向性を対応づけ得点の信頼性という新たな視点を対応づけ可能性分析に応用することに軌道修正した。線形等化法を利用した対応づけの場合に焦点を絞り、対応づけ得点の測定モデルを構築するとともに、対応づけ得点の信頼性に関する指標を定式化した。さらに、いくつかの大規模試験の対応づけ可能性分析を通し、提案指標が対応づけ得点の信頼性評価に役立つことを例示した。

4. 研究成果

(1) 表記法および基礎事項

本稿では、任意の確率変数  $U, V$  の統計量について、以下のような表記法を一貫して用いることにする。

- $E(U), E(V): U, V$  の期待値
- $\sigma^2(U), \sigma^2(V): U, V$  の分散
- $\sigma(U, V): U, V$  の共分散
- $\mu(U), \mu(V): U, V$  の平均

$\sigma(U), \sigma(V): U, V$  の標準偏差

$\rho(U, V): U, V$  の相関係数

古典的テスト理論 (Holland & Hoskens、2003) によれば、テスト  $X$  における観測得点  $S_X$  は、真の得点  $\tau_X$  とそれとは無相関の測定誤差  $\varepsilon_X$  を用いて、

$$S_X = \tau_X + \varepsilon_X \quad (1)$$

と表される。ここで、 $\tau_X, \varepsilon_X, S_X$  は有限の平均と分散をもつ確率変数とする。測定誤差については、 $E(\varepsilon_X) = 0$  が仮定されるとともに、 $\sigma^2(\varepsilon_X)$  は測定誤差分散、その正の平方根  $\sigma(\varepsilon_X)$  は測定の標準誤差と呼ばれる。

テスト  $X$  の信頼性係数は、分散比  $\sigma^2(\tau_X) / \sigma^2(S_X)$  として定義される。信頼性指数は、信頼性係数とは別の定義をもつ信頼性指標の一つであり、観測得点とその真の得点との相関係数  $\rho(S_X, \tau_X)$  として定義される。なお、信頼性指数の 2 乗  $\rho^2(S_X, \tau_X)$  は、個々のテストの場合、信頼性係数の定義  $\sigma^2(\tau_X) / \sigma^2(S_X)$  と同値である (池田、1973、p.77)。

(2) 対応づけ得点の測定モデル

第一の研究成果は、対応づけ得点の測定モデルを記述したことである。いま、適切なデータ収集デザインのもとでテスト  $X$  をテスト  $Y$  に対応づけする場面を想定しよう。本研究では、対応づけ得点は共通尺度の真の得点を別のテストを通して間接的に測定した結果であると考えられる。このとき、対応づけ得点  $l_Y(S_X)$  は、真の得点  $\tau_Y$  と測定誤差  $\varepsilon_Y'$  を用いて、

$$l_Y(S_X) = \tau_Y + \varepsilon_Y'$$

と表される。ここで、 $\tau_Y, \varepsilon_Y', l_Y(S_X)$  は有限の平均と分散をもつ確率変数とする。ただし、(1)式のように、 $\tau_Y$  と  $\varepsilon_Y'$  の間に特別な関係を仮定しない。

(3) 対応づけ得点の信頼性に関する指標

第二の研究成果は、線形等化法を利用した対応づけの場合に焦点を絞り、対応づけ得点の信頼性に関する指標を提案したことである。上述の測定モデルから出発し、以下の指標を定式化した。

対応づけ得点の測定誤差の期待値は、

$$E(\varepsilon_Y') = \mu(S_Y) - E(S_Y)$$

と表される。線形等化法を利用する場合、その基本的な性質から  $E(\varepsilon_Y') = 0$  である。

対応づけ得点の測定誤差分散は、

$$\sigma^2(\varepsilon_Y') = \sigma^2(S_Y)[1 + \rho^2(S_Y, \tau_Y) - 2\rho(S_X, S_Y)]$$

と表される。ただし、 $\rho^2(S_Y, \tau_Y)$  はテスト Y の信頼性係数である。上式の正の平方根  $\sigma(\varepsilon_Y')$  が対応づけ得点の測定の標準誤差である。

対応づけ得点の信頼性指数は、対応づけ得点とその共通尺度における真の得点との相関係数と定義すれば、

$$\rho(l_Y(S_X), \tau_Y) = \frac{\rho(S_X, S_Y)}{\rho(S_Y, \tau_Y)}$$

と表される。

#### (4) 対応づけ可能性分析への応用

第三の研究成果は、いくつかの大規模試験の対応づけ可能性分析を通し、提案指標が対応づけ得点の信頼性評価に役立つことを例示したことである。ここでは、法科大学院全国統一適性試験 (Japan Law School Admission Test, JLSAT) の対応づけ可能性分析の例を示す。

2011年に実施された JLSAT の試験結果を利用し、同じテストの版 X, Y を単一グループデザインのもとで線形等化する状況を想定する。この例では、両回受験者 4,915 名から 1,000 名を共通受験者として無作為に抽出し、第 1 回の試験結果を版 X (72 項目)、第 2 回の試験結果を版 Y (72 項目) として 2 値の項目反応データを作成した。なお、実際には、本適性試験の第 1 回と第 2 回はもう少し複雑なデータ収集デザインのもとで等パーセントイル等化されている (佐藤・柴山, 2010b)。この例は、あくまでも対応づけ可能性の検討例を示すための状況設定であることに注意されたい。なお、試験結果のより詳しい内容については、適性試験管理委員会 (2011) を参照されたい。

表 1 に、共通受験者 1,000 名の正答数得点についての統計量を示す。表 2 と表 3 に、尺度 X, Y における版 X, Y の測定の標準誤差と (信頼性指数)<sup>2</sup> を示す。太字でない数値は、それぞれの尺度におけるテスト得点の信頼性を表し、Cronbach の係数を用いて推定された。また、太字の数値は、それぞれの尺度における等化得点 (対応づけ得点) の信頼性を表し、前述の提案指標を用いて推定された。

版 X, Y の信頼性係数を正しく評価できるなら、表 2 と表 3 を用いて以下のような対応づけ可能性分析が考えられる。たとえば、尺度 Y を共通尺度とする場合、版 Y の受験者におけるテスト得点の測定の標準誤差は 3.74 であり、(信頼性指数)<sup>2</sup> は 0.823 である。それ

に対し、版 X の受験者における等化得点の測定の標準誤差は 5.15 であり、(信頼性指数)<sup>2</sup> は 0.673 である。これらの数値から、尺度 Y においては、版 Y のテスト得点の方が版 X の等化得点より信頼性が高いことがわかる。さらに、表 2 と表 3 は、どちらの尺度を共通尺度とすればよいかについて一つの判断材料を与えてくれる。たとえば、テスト得点も等化得点も (信頼性指数)<sup>2</sup> > 0.7 としたい場合、あるいはテスト得点と等化得点の (信頼性指数)<sup>2</sup> の値をなるべくそろえて報告したい場合、尺度 X を共通尺度とした方がよいことを判断できる。

表 1 正答数得点の基礎統計量

	平均	標準偏差	相関係数
版X	39.5	7.65	0.744
版Y	43.3	8.90	

版X: 第1回、版Y: 第2回

表 2 テスト得点と等化得点の信頼性 (尺度X)

	尺度X	
	測定の標準誤差	(信頼性指数) <sup>2</sup>
版X	3.76	0.758
版Y	<b>3.98</b>	<b>0.730</b>

**太字**: 等化得点の信頼性

表 3 テスト得点と等化得点の信頼性 (尺度Y)

	尺度Y	
	測定の標準誤差	(信頼性指数) <sup>2</sup>
版X	<b>5.15</b>	<b>0.673</b>
版Y	3.74	0.823

**太字**: 等化得点の信頼性

#### (5) 今後の展望など

提案指標の利用は、厳密には単一グループデザインまたは等価グループデザインのもとで線形等化法を利用する場合に限られる。また、実際のテストの等化には、線形等化法だけでなく等パーセントイル法も広く利用されている。今後の課題の一つは、非線形な等化法と提案指標との関係を検討することである。そのような制限はあるものの、本研究を通して、これまで十分な議論がなかった対応づけ得点の信頼性評価というテーマに着手したことは一つの成果といえよう。

本研究の主たる成果は、佐藤・柴山 (2014) にまとめられている。全期間を通じた研究成果は、分割統治法の考え方を応用した新しい対応づけ技術の開発へとつながるものである。

#### < 引用文献 >

柴山 直、野口 裕之、「対応づけ」の理論と計算アルゴリズム、適性試験委員会 (編) 法科大学院統一適性試験テクニカ

ル・レポート 2004、商事法務、2004、pp.53 - 72

石井 秀宗、全国学力・状況調査の項目分析的検討、平成 19 年度「全国学力・学習状況調査」分析報告書、千葉県検証改善委員会、2008、pp.98 - 111

佐藤 喜一、柴山 直、下位テストから構成されるテスト間の等化において合計と等化の順序が与える影響について、日本テスト学会第 8 回大会発表論文抄録集、2010a、pp.168 - 171

佐藤 喜一、柴山 直、対応づけ得点のための信頼性指標の提案 対応づけ可能性分析への応用、日本テスト学会誌、Vol.10、No.1、2014、pp.69 - 80

池田 央、心理学研究法第 8 巻 テスト、東京大学出版会、1973

Holland, P. W.、 & Hoskens, M., "Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test, Psychometrika, Vol.68, No.1, 2003, pp.123-149

適性試験管理委員会、2011 年法科大学院全国統一適性試験実施報告書、日弁連法務研究財団、2011 年 12 月 15 日、<[https://www.jlf.or.jp/jlsat/pdf/20111215\\_houkoku.pdf](https://www.jlf.or.jp/jlsat/pdf/20111215_houkoku.pdf)> (2016 年 5 月 11 日)

佐藤 喜一、柴山 直、下位テストから構成されるテスト間の等化におけるブートストラップ法を援用した等化の標準誤差の評価、日本テスト学会誌、Vol.7、No.1、2010b、pp.85 - 98

## 5 . 主な発表論文等

[雑誌論文](計 1 件)

佐藤 喜一、柴山 直、対応づけ得点のための信頼性指標の提案 対応づけ可能性分析への応用、日本テスト学会誌、査読有、Vol.10、No.1、2014、pp.69 - 80

[学会発表](計 4 件)

佐藤 喜一、柴山 直、対応づけ可能性のネガティブチェックの試み 対応づけ得点の信頼性に着目して、日本テスト学会第 13 回大会、2015 年 9 月 10 日、関西大学千里山キャンパス(大阪府吹田市)

Sato, Y., & Shibayama, T., "Equating by subtest scores" strategy for JLSAT

scoring, Law School Admission Council, Mar/21/2014, Law School Admission Council (Newtown, PA)

佐藤 喜一、柴山 直、線形等化の向きを決定するための判断基準の提案 平均二乗誤差減少率の定式化、日本テスト学会第 10 回大会、2012 年 8 月 22 日、東京医科歯科大学(東京都文京区)

佐藤 喜一、柴山 直、テストの分割等化が等化結果に与える影響について、日本テスト学会第 9 回大会、2011 年 9 月 11 日、岡山大学(岡山県岡山市)

## 6 . 研究組織

(1)研究代表者

佐藤 喜一(SATO, Yoshikazu)  
新潟大学・教育・学生支援機構・准教授  
研究者番号：00300517

(2)研究分担者

柴山 直(SHIBAYAMA, Tadashi)  
東北大学・大学院教育学研究科・教授  
研究者番号：70240752