

平成 31 年 4 月 24 日現在

機関番号：17102

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K04052

研究課題名(和文) 対応づけ得点の信頼性に着目した新しい対応づけ可能性分析法の開発

研究課題名(英文) Development of a Linkability Analysis Method on Reliability of Linked Scores

研究代表者

佐藤 喜一 (Sato, Yoshikazu)

九州大学・アドミッションセンター・教授

研究者番号：00300517

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：二つの異なるテストを「対応づけ」すると、テストXの80点がテストYの85点に相当するなど、テストX、Yのスコアの対応関係を知ることができる。その際、意味ある対応づけを実現するためには、さまざまな観点から「対応づけ可能性(対応づけの実行可能性)」を吟味する必要がある。本研究の目的は、「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導入することである。対応づけ得点の測定モデルを構築し、対応づけ得点の信頼性を評価するための基本的な指標を定式化した。さらに、それらの指標により、対応づけ得点の信頼性という観点から対応づけ可能性分析が可能なることを例示した。

研究成果の学術的意義や社会的意義

大学英语とTOEICやTOEFLとの対応づけや、全国学力調査と地方自治体の学力調査との対応づけなど、近年、ますます対応づけへの関心が高まりつつある。対応づけを伴うテストの場合、私たちの関心の多くは共通尺度における受験者の得点、すなわち対応づけ得点にある。本研究では、これまで十分な議論のなかった「対応づけ得点の信頼性」に着目し、定式化した指標を対応づけ可能性分析に応用した。これらの研究成果により、意味ある対応づけが可能なるテストの組み合わせかどうかを判断できるため、対応づけの濫用を防ぐなどの効果が期待できる。

研究成果の概要(英文)：When two different tests are linked, we can see the correspondence between the scores of Test X and Test Y, such as 80 points of Test X correspond to 85 points of Test Y. For meaningful linking, it is necessary to examine the linkability, or the feasibility of linking, between the two tests in a comprehensive way. The purpose of this study was to incorporate a new perspective on the reliability of linked scores into linkability analysis. We defined a measurement model of linked scores and then formulated basic indices for the reliability of linked scores. In addition, we showed that the indices enable us to perform linkability analysis in terms of the reliability of linked scores.

研究分野：教育測定学

キーワード：教育測定 古典的テスト理論 対応づけ 等化 対応づけ可能性(対応づけの実行可能性) テストの信頼性 対応づけ得点の信頼性

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

(1) 二つの異なるテストを「対応づけ (linking)」すると、テスト X の 80 点がテスト Y の 85 点に相当するなど、テスト X、Y のスコアの対応関係を知ることができる。このように、複数のテストを対応づけすることにより、テスト単体よりも一層、立体的で有益な教育情報を引き出すことができる (Dorans, Pommerich, & Holland, 2007; 柴山, 2006)。たとえば、法科大学院入試における二つの適性試験の対応づけ、大学英語と TOEIC や TOEFL との対応づけ、全国学力調査と地方自治体の学力調査との対応づけなど、近年、ますます対応づけへの関心が高まりつつある (熊谷・五島・中畝・柴山, 2007; 石井, 2010; 柴山・野口, 2004)。

(2) 等化 (equating) の技術を利用すれば、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、どんなテストの組み合わせでも対応づけしてよいわけではなく、意味ある対応づけを実現するためには、さまざまな観点から対応づけ可能性 (linkability) あるいは対応づけの実行可能性 (feasibility of linking) を吟味する必要がある (Dorans, 2000; Dorans, 2004; Dorans & Holland, 2000; Feuer, Holland, Green, Bertenthal, & Hemphill, 1999; Hanson, Harris, Pommerich, Scoring, & Yi, 2001; Holland & Hoskens, 2003; Holland, 2005)。その中でも、テストの信頼性は対応づけ結果の正しい解釈のために重要な要素の一つである。

(3) これまでの対応づけ可能性分析では、テストの信頼性として個々のテストの信頼性を評価するのが常であった。しかしながら、対応づけを伴うテストの場合、私たちの関心の多くは共通尺度における受験者の得点、すなわち対応づけ得点にある。それにもかかわらず、対応づけ得点の信頼性については現在まで十分な議論がなく、尺度変換によってテストの信頼性はどのように変化するか議論が留まっていた (Almehrizi, 2013; Cronbach & Shavelson, 2004; Feldt & Brennan, 1989; Kolen & Brennan, 2014, Chapter 9; Kolen & Lee, 2011; Segall, 1994)。

2. 研究の目的

本研究の目的は、「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導入することである。単一グループ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、対応づけ得点 $l_Y(X)$ の信頼性に関する基本的な指標として、測定 $RMSE$ (root mean square error) や信頼性係数などを定式化した。さらに、定式化された指標の推定時の性質を調べるとともに、それらの指標を対応づけ可能性分析に応用できることを例示した。

3. 研究の方法

(1) 本研究では、「対応づけ得点は、共通尺度となるテスト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」と考える。すなわち、テスト X をテスト Y に対応づけするとき、対応づけ得点 $l_Y(X)$ を観測得点、共通尺度となるテストの真の得点 T_Y を $l_Y(X)$ の真の得点とみなし、その差を測定誤差 $D_Y = l_Y(X) - T_Y$ と定義して測定モデルを記述する。この測定モデルから出発し、対応づけ得点の信頼性を評価するための指標を新たに定式化した。

(2) シミュレーション研究により、定式化された指標を対応づけ可能性分析に応用できることを確認した。項目反応理論 (item response theory) を利用し、さまざまな対応づけ場面のテスト結果をコンピュータ内に生成した。対応づけ得点の信頼性を評価することにより、意味ある対応づけが可能なテストの組み合わせかどうかを判断した。

4. 研究成果

(1) 単一グループ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する。テスト得点 X 、 Y が古典的テスト理論 (Feld & Brennan, 1989; Haertel, 2006) に従うと仮定する。対応づけ得点 $l_Y(X)$ を観測得点、共通尺度となるテスト Y の真の得点 T_Y を $l_Y(X)$ の真の得点とみなし、その差を対応づけ得点の測定誤差 $D_Y = l_Y(X) - T_Y$ と定義すれば、対応づけ得点の測定モデルは、

$$l_Y(X) = T_Y + D_Y$$

と記述できる。ここで、 X 、 T_Y 、 D_Y 、 $l_Y(X)$ は、期待値が存在する確率変数または確率関数である。実際の対応づけ場面を考慮し、一般には測定誤差の期待値が $E(D_Y) \neq 0$ として取り扱う。

(2) 上記の測定モデルから出発し、対応づけ得点 $l_Y(X)$ の信頼性に関する基本的な指標を定式化した (表 1)。テスト Y の信頼性係数 $\rho(Y)$ の推定値として、係数 (Cronbach, 1951) を用いると、受験者数が大きいとき、測定 $RMSE$ の推定値はその真値の下界となり、信頼性係数 $\rho[l_Y(X)]$ の推定値はその真値の上界となる。対応づけ得点の信頼性がある基準値より高いことを評価できるように、測定 $RMSE$ の上界と信頼性係数 $\rho[l_Y(X)]$ の下界を推定できる式も導出した (表 2)。

表1 対応づけ得点 $l_Y(X)$ の信頼性に関する基本的な指標

指標	式
測定誤差の期待値	$E(D_Y) = E[l_Y(X) - Y]$
測定誤差分散	$\sigma^2(D_Y) = \sigma^2[l_Y(X) - Y] - \sigma^2(Y)[1 - \rho(Y)]$
測定の標準誤差	$\sigma(D_Y)$
測定のMSE	$E(D_Y^2) = E\{[l_Y(X) - Y]^2\} - \sigma^2(Y)[1 - \rho(Y)]$
測定のRMSE	$\sqrt{E(D_Y^2)}$
信頼性係数	$\rho[l_Y(X)] = \frac{\sigma^2(T_Y)}{\sigma^2(T_Y) + E(D_Y^2)}$
信頼性指数	$\rho[l_Y(X), T_Y] = \frac{\rho[l_Y(X), Y]}{\sqrt{\rho(Y)}}$

注) $\rho(Y)$ = テストYの信頼性係数 .

表2 上界・下界と線形等化法による近似式

指標	式
測定のMSEの上界	$E(D_Y^2) \leq E\{[l_Y(X) - Y]^2\} \approx 2\sigma^2(Y)[1 - \rho(X, Y)]$
信頼性係数の下界	$\rho[l_Y(X)] \geq \frac{\sigma^2(Y)}{\sigma^2(Y) + E\{[l_Y(X) - Y]^2\}} \approx \frac{1}{3 - 2\rho(X, Y)}$

注) 上段の正の平方根が測定のRMSEの上界である .

(3) 単一グループ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、定式化された指標が対応づけ可能性分析に応用できることを例示した。三つの例の項目反応データについては、Rasch モデル (Rasch, 1960) を用いてそれぞれ生成した。

例 1 の結果を表 3 と表 4 に示す。表 3 は、テスト X, Y の正答数得点についての基本統計量と信頼性を推定した結果である。表 4 は、対応づけ得点 $l_Y(X)$ の基本統計量と信頼性を推定した結果である。表 4 の信頼性係数の欄をみると、どちらの等化法を利用して下界の推定値が 0.7 を上回っている。信頼性係数が 0.7 以上を一つの基準とすれば、対応づけ得点の信頼性の観点からは、どちらの等化法で対応づけしてもよいと判断できる。

例 2 の結果を表 5 と表 6 に、例 3 の結果を表 7 と表 8 に示す。例 2 については、対応づけ得点の信頼性係数の上界が 0.7 を下回っていることから、どちらの等化法でも対応づけしないほうがよいと判断できる。例 3 については、対応づけ得点の信頼性係数の推定区間に 0.7 が含まれることから、テストの利用目的によって判断が分かれるところである。

表3 テストの基本統計量と信頼性 (例1)

	項目数	平均	標準偏差	測定の標準誤差	係数
テストX	50	27.2	6.22	2.50	0.839
テストY	50	27.9	5.84	2.40	0.832

注) $N=1,000$. $r=0.851$.

表4 対応づけ得点 $l_Y(X)$ の基本統計量と信頼性 (例1)

対応づけ	平均	標準偏差	測定のRMSE		信頼性係数	
			下界	上界	下界	上界
線形等化法	27.9	5.81	2.10	3.19	0.770	0.865
等パーセントイル法	27.9	5.79	2.08	3.17	0.772	0.868

表5 テストの基本統計量と信頼性 (例2)

	項目数	平均	標準偏差	測定の標準誤差	係数
テストX	50	24	6.22	2.49	0.840
テストY	50	25	6.41	2.48	0.851

注) $N=1,000$. $r=0.717$.

表6 対応づけ得点 $l_Y(X)$ の基本統計量と信頼性(例2)

対応づけ	平均	標準偏差	測定のRMSE		信頼性係数	
			下界	上界	下界	上界
線形等化法	25.0	6.26	4.08	4.77	0.644	0.678
等パーセンタイル法	25.1	6.38	4.13	4.81	0.640	0.672

表7 テストの基本統計量と信頼性(例3)

	項目数	平均	標準偏差	測定の標準誤差	係数
テストX	25	10.2	3.35	1.73	0.732
テストY	25	16.6	3.28	1.74	0.719

注) $N=1,000$. $r=0.741$.

表8 対応づけ得点 $l_Y(X)$ の基本統計量と信頼性(例3)

対応づけ	平均	標準偏差	測定のRMSE		信頼性係数	
			下界	上界	下界	上界
線形等化法	16.2	3.29	1.65	2.40	0.652	0.740
等パーセンタイル法	16.5	3.09	1.52	2.31	0.669	0.771

(4) 単一グループデザインに限られるものの、対応づけ得点の信頼性という観点から対応づけ可能性分析が可能となった。本研究の成果により、意味ある対応づけが可能なテストの組み合わせかどうかを判断できるため、対応づけの濫用を防ぐなどの効果が期待できる。今後の課題の一つは、定式化された指標の実用性を確認するため、さらに分析例を蓄積することである。本研究の成果は、対応づけ得点の性質解明という基礎研究に留まらず、新しい対応づけ技術の開発などの応用研究へと発展が見込まれる。

<引用文献>

- Almehrizi, R. S. (2013). Coefficient alpha and reliability of scale scores. *Applied Psychological Measurement, 37* (6), 438-459.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3), 297-334.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64* (3), 391-418.
- Dorans, N. J. (2000). *Distinctions among classes of linkages* (College Board Research Note No. RN-11). New York: The College Board.
- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement, 28* (4), 227-246.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement, 37* (4), 281-306.
- Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.) (2007). *Linking and aligning scores and scales*. New York: Springer Science+Business Media, LLC.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: Macmillan.
- Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon measures: Equivalence and linkage among educational tests* (Report of the Committee on Equivalency and Linkage of Educational Tests, National Research Council). Washington DC: National Academy Press.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education and Praeger.
- Hanson, B. A., Harris, D. J., Pommerich, M., Sconing, J. A., & Yi, Q. (2001). *Suggestions for the evaluation and use of concordance results* (ACT Research Report No. 2001-1). Iowa City, IA: ACT, Inc.
- Holland, P. W. (2005). Assessing the validity of test linking: What has happened since uncommon measures? In C. A. Dwyer (Ed.), *Measurement and research in the accountability era*. Mahwah, NJ: Erlbaum.

- Holland, P. W., & Hoskens, M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*, 68 (1), 123-149.
- 石井秀宗 (2010). 平成 21 年度文部科学省企画公募委託研究「学力調査を活用した専門的な課題分析に関する調査研究」『地域におけるデータ等を補完的に用いた調査分析手法の調査研究』報告書.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.
- Kolen, M. J., & Lee, W. (2011). Psychometric properties of raw and scaled scores on mixed-format tests. *Educational Measurement: Issues and Practice*, 30, 15-24.
- 熊谷龍一・五島譲司・中畝菜穂子・柴山 直 (2007). 新潟大学における「共通英語」の成績と TOEIC および大学入試センター試験との関係. *日本テスト学会誌*, 3 (1), 91-97.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Segall, D. O. (1994). The reliability of linearly equated tests. *Psychometrika*, 59 (3), 361-375.
- 柴山 直 (2006). 学力テスト. *日本経済新聞* 2006 年 10 月 16 日朝刊, 寄稿論考.
- 柴山 直・野口裕之 (2004). 「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*, 53-72. 商事法務.

5. 主な発表論文等

〔学会発表〕(計 5 件)

- 佐藤喜一 (2018). 法科大学院全国統一適性試験の歩み 等化法の開発経緯とその後の研究 . *日本テスト学会第 16 回大会発表論文抄録集*, 34-35 .
- 佐藤喜一・柴山 直 (2015). 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 回大会発表論文抄録集*, 44-47 .
- 佐藤喜一・柴山 直 (2016). 対応づけ得点の信頼性係数について . *日本テスト学会第 14 回大会発表論文抄録集*, 116-119 .
- 佐藤喜一・柴山 直 (2017). 対応づけ得点の信頼性を評価するための簡便な指標 . *日本テスト学会第 15 回大会発表論文抄録集*, 156-159 .
- Sato, Y., & Shibayama, T. (2018, April). Linkability analysis focused on reliability of linked scores. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

6. 研究組織

〔研究分担者〕

研究分担者氏名: 柴山 直
ローマ字氏名: (SHIBAYAMA, tadashi)
所属研究機関名: 東北大学
部局名: 大学院教育学研究科
職名: 教授
研究者番号 (8 桁): 70240752

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。