

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—

○佐藤喜一*

*新潟大学入学センター

柴山 直**

**東北大学大学院教育学研究科

1 はじめに

等化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、意味ある対応づけを保証するには、対応づけ可能性 (linkability) の吟味が重要である。その検討内容は、構成概念の類似性、テストの関係性、測定精度、スコアの公平性など多岐にわたる。

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。本発表では、対応づけ得点の信頼性に関する指標を定式化するとともに、それらの指標を対応づけしない方がよいかどうかの判断に応用した例を示す。この対応づけ可能性のネガティブチェックにより、対応づけの濫用を防ぐ効果が期待できる。

2 定式化

2.1 測定モデル

古典的テスト理論によれば、テスト X における観測得点 S_X は、真の得点 τ_X と独立な測定誤差 ε_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である。テスト Y についても、(1) 式と同様に定義できる。測定誤差 $\varepsilon_X, \varepsilon_Y$ については、それぞれ期待値が 0、互いに独立が仮定される。

いま、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る。このとき、(1) 式を模倣することにより、対応づけ得点 $l_Y(S_X)$ は、真の得点 τ_Y と測定誤差 ε_Y' を用いて、

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

と表される。ここで、 τ_Y, ε_Y' は有限の平均と分散をもつ確率変数であり、 $l_Y(S_X)$ は確率変数 S_X の関数である。(1) 式の場合とは異なり、測定誤差については特別な条件を仮定しない。

2.2 測定 RMSE

対応づけ得点の測定誤差の期待値は、

$$\begin{aligned} E(\varepsilon_Y') &= E[l_Y(S_X) - \tau_Y] \\ &= E[l_Y(S_X)] - E(S_Y) \end{aligned} \quad (2)$$

と表される。

対応づけ得点の測定誤差分散は、

$$\begin{aligned} \sigma^2(\varepsilon_Y') &= \sigma^2[l_Y(S_X) - \tau_Y] \\ &= \sigma^2[l_Y(S_X)] + \sigma^2(\tau_Y) \\ &\quad - 2\sigma[l_Y(S_X), \tau_Y] \end{aligned} \quad (3)$$

と表される。実際に (3) 式を利用するには、第 3 項の共分散をいかに計算するかが鍵である。 $l_Y(S_X)$ と ε_Y の共分散は、 $\mu(\cdot)$ を $\cdot$ の平均とすれば、

$$\begin{aligned} &\sigma[l_Y(S_X), \varepsilon_Y] \\ &= E(\{l_Y(S_X) - \mu[l_Y(S_X)]\}[\varepsilon_Y - \mu(\varepsilon_Y)]) \\ &= E\{l_Y(S_X)\varepsilon_Y - \mu[l_Y(S_X)]\varepsilon_Y\} \\ &= E[l_Y(S_X)\varepsilon_Y] - \mu[l_Y(S_X)]E(\varepsilon_Y) \\ &= E[l_Y(S_X)\varepsilon_Y] \end{aligned}$$

と計算できる．ところで，互いに独立な確率変数 X_1, X_2 の関数 $u(X_1), v(X_2)$ について $E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$ がいえる．いま， S_X と ε_Y は互いに独立なので，

$$\sigma[l_Y(S_X), \varepsilon_Y] = E[l_Y(S_X)]E(\varepsilon_Y) = 0$$

が得られる．このとき， $\sigma[l_Y(S_X), \tau_Y] = \sigma[l_Y(S_X), S_Y]$ に注意すれば，(3) 式は，

$$\begin{aligned} \sigma^2(\varepsilon_Y') &= \sigma^2[l_Y(S_X) - S_Y] \\ &\quad - \sigma^2(S_Y)[1 - \rho^2(S_Y, \tau_Y)] \end{aligned} \quad (4)$$

と書き換えられる．ただし， $\rho^2(S_Y, \tau_Y)$ はテスト Y の信頼性係数である．(4) 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．

対応づけ得点の測定の二乗平均誤差 (mean square error, MSE) は，

$$\begin{aligned} E\{[l_Y(S_X) - \tau_Y]^2\} \\ = \sigma^2(\varepsilon_Y') + [E(\varepsilon_Y')]^2 \end{aligned} \quad (5)$$

と表される．(5) 式の正の平方根が対応づけ得点の測定の二乗平均平方根誤差 (root mean square error, RMSE) である．

2.3 信頼性指数

対応づけ得点と共通尺度の真の得点との相関係数を対応づけ得点の信頼性指数と定義すれば，

$$\begin{aligned} &\rho[l_Y(S_X), \tau_Y] \\ &= \frac{\sigma[l_Y(S_X), \tau_Y]}{\sigma[l_Y(S_X)]\sigma(\tau_Y)} \\ &= \frac{\sigma[l_Y(S_X), S_Y]}{\sigma[l_Y(S_X)]\sigma(S_Y)\rho(S_Y, \tau_Y)} \\ &= \frac{\rho[l_Y(S_X), S_Y]}{\rho(S_Y, \tau_Y)} \end{aligned} \quad (6)$$

と表される．(6) 式は， $l_Y(S_X)$ を S_X とすれば，形式上は希薄化の修正公式と一致する．しかしながら， $l_Y(S_X)$ は S_X の関数である点と， $l_Y(S_X)$ の方が S_X より測定モデルの仮定が緩い点は異なる．いわば (6) 式は，希薄化の修正公式をやや一般化した式といえよう．

実際に対応づけ得点の信頼性を評価するには，(6) 式の 2 乗を利用するとよいであろう．なぜなら，個々のテストの信頼性係数は信頼性指数の 2 乗と等しく，その延長上に対応づけ得点の信頼性を評価できるからである．

なお，対応づけに線形等化法を利用する場合，(4) 式と (6) 式を整理すると佐藤・柴山 (2014) の導出結果に一致する．

3 応用例

3.1 α 係数による推定値

「対応づけ得点の信頼性指数の 2 乗」と「対応づけ得点の測定の RMSE」を推定するには，信頼性係数の推定値が必要である．ここでは，信頼性係数の推定値として Cronbach の α 係数を用いた場合，これらの推定値にどんな影響を与えるのかを考察する．

前者の指標は，(6) 式の 2 乗として推定される．このとき，分母に信頼性係数の推定値が必要である．受験者数が大きい場合， α 係数は当該テストの信頼性係数の一つの下限值であることが証明されている．それゆえ，受験者数が大きい場合，(6) 式の 2 乗の推定値は，その真値に対する一つの上限值となる．

一方，後者の指標は，(5) 式の正の平方根として推定される．そのうち，第 1 項に相当する (4) 式の推定に信頼性係数の推定値が必要である．前者と同様に，(5) 式の正の平方根の推定値は，その真値に対する一つの下限值となる．

3.2 対応づけ可能性のネガティブチェック

テストの現場では，テストの高い信頼性を保証するため， α 係数が α_C 以上という基準がしばしば設定される (例： $\alpha_C = 0.7$)．ここでは，この基準値 α_C を対応づけ得点の信頼性評価にも利用し，対応づけしない方がよいかどうかを判断する一つの方法を例示する．以下の条件を順番に確認し，どちらか一方でも条件を満たすとき，テスト X をテスト Y に対応づけしない方

表 1: テスト X, Y の基礎統計量 (例 1)

	受験者数	項目数	平均	標準偏差	測定の標準誤差	α 係数	相関係数
テスト X	1,000	50	23.9	6.51	2.54	0.848	0.714
テスト Y		50	25.8	5.95	2.39	0.839	

表 2: 尺度 Y における対応づけ得点の信頼性 (例 1)

	平均	標準偏差	測定の RMSE	信頼性指数の 2 乗
線形等化法	25.8	5.97	3.81	0.608
等パーセンタイル法	25.8	5.90	3.80	0.606

がよいと判断する。

(条件 1) 信頼性指数の 2 乗 $< \alpha_C$

(条件 2) 測定の RMSE $> \sigma[l_Y(S_X)]\sqrt{1 - \alpha_C}$

(条件 1) は、対応づけ得点の信頼性指数の 2 乗と基準値 α_C との比較である。前節の考察から、 α 係数を用いた左辺の上限値が α_C 未満ならば、その真値も明らかに α_C 未満である。この場合、対応づけ得点の信頼性は十分とはいえず、対応づけは控えた方がよいと考えられる。

(条件 2) の右辺は、対応づけ得点と等しい得点分散をもつテストにおいて信頼性係数が α_C のときの測定の標準誤差に相当する。(条件 1) と同様に、 α 係数を用いた推定値が (条件 2) を満たすならば、その真値も明らかに (条件 2) を満たす。この場合、対応づけ得点の測定の RMSE は許容誤差より大きいので、対応づけは控えた方がよいと考えられる。

(条件 1) の不等号が逆向きの場合、左辺の上限値が α_C 以上であっても、その真値が α_C 以上だとは限らない。同様に (条件 2) の不等号が逆向きの場合も、測定の RMSE の真値について確かなことはいえない。それゆえ、二つの条件が両方とも成り立たない場合、対応づけ可能性についての判断は保留とする。

3.3 対応づけの例

例 1 では、共通受験者にとって適切な難易度のテスト X, Y について、テスト X からテスト Y へ対応づけする場面を想定した。項目反応データは、Rasch モデルを利用して生成した。共通受験者 1,000 名の能力値は、相関係数 0.85 の 2 変量標準正規分布から無作為に抽出した。項目困難度は、テスト X, Y とも -4~4 の一様分布から 50 項目ずつを無作為に抽出した。

表 1 に、テスト X, Y の正答数得点についての基礎統計量を示す。表 2 に、線形等化法と等パーセンタイル法を用いた場合の対応づけ得点の信頼性を示す。なお、表 2 の結果は、テスト Y の得点範囲を超えた対応づけ得点を 0 点から 50 点 (満点) の範囲に制限し、さらに四捨五入して対応づけ得点を整数化したときの結果である。 $\alpha_C = 0.7$ として上述の条件を適用してみると、表 2 において (条件 1) が成り立つことがわかる。したがって、どちらの等化法を利用したとしても、テスト X をテスト Y へ対応づけしない方がよいと判断できる。

3.4 垂直尺度化の例

例 2 では、同一の構成概念を測定するものの難易度が異なる二つのテストを垂直尺度化 (vertical scaling) する場面を想定した。項目反

表 3: テスト X, Y の基礎統計量 (例 2)

	受験者数	項目数	平均	標準偏差	測定の標準誤差	α 係数	相関係数
テスト X	1,000	50	22.7	6.26	2.49	0.841	0.806
テスト Y		50	41.8	5.21	2.23	0.817	

表 4: 尺度 Y における対応づけ得点の信頼性 (例 2)

	平均	標準偏差	測定の RMSE	信頼性指数の 2 乗
線形等化法	41.7	4.94	2.15	0.813
等パーセンタイル法	41.8	5.25	2.03	0.851

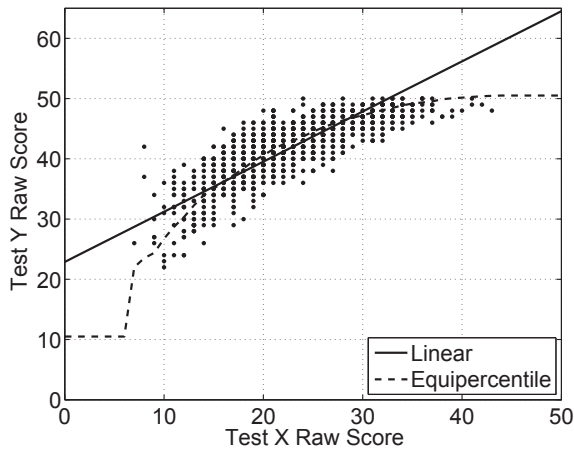


図 1: 対応づけ関数 (例 2)

応データは、例 1 と同様に生成した。その際、2 変量標準正規分布の相関係数は 1 とし、テスト X, Y の項目困難度はそれぞれ-4~4 と-7~1 の範囲に設定した。

例 1 と同様に、表 3 と表 4 に結果を示す。表 4 の結果は、例 1 と同様の方法で対応づけ得点を整数化したときの結果である。図 1 には、正答数得点の同時分布と各等化法の対応づけ関数を示す。 $\alpha_C = 0.7$ とすれば、表 4 において (条件 1) は成り立たないことがわかる。また、(条件 2) の右辺は、線形等化法と等パーセンタイル法の場合でそれぞれ 2.70 と 2.88 であり、(条件

2) も成り立たないことがわかる。したがって、この場合、対応づけ可能性についての判断は保留となる。実際に対応づけするかどうかは、さらに別の観点から考察する必要があるだろう。

表 4 において、推定時の α 係数の値は共通なので、等化法を比較することには意味がある。仮に対応づけするとした場合、測定の RMSE に注目すれば、等パーセンタイル法を用いた方がよいと判断できる。これは、図 1 の対応づけ関数から得られる直観と一致するであろう。

4 おわりに

単一グループデザインのもとでの対応づけにおいて、定式化した指標は対応づけ可能性のネガティブチェックに役立つ可能性が示された。

謝辞 本研究は JSPS 科研費 23501139, 15K04052 の助成を受けたものです。

参考文献

佐藤喜一・柴山 直 (2014). 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, 10(1), 69–80.

(ysato@ge.niigata-u.ac.jp)