測定・評価・研究法部門

心理学と心理測定における信頼性について

— Cronbach の α 係数とは何なのか, 何でないのか —

岡 田 謙 介

(専修大学)

Reliability in Psychology and Psychological Measurement, with Focus on Cronbach's Alpha

Kensuke OKADA

(SENSHU UNIVERSITY)

Studies in educational psychology in Japan are reviewed from the viewpoint of reliability. Cronbach's alpha is one of the most frequently used —if not fully understood— measures of reliability. The main purpose of this note is to give an overview of what alpha is and is not. Specifically, it is shown that (1) alpha is the mean of all split-half reliabilities, (2) alpha is a lower bound of reliability, and (3) alpha equals reliability under essential tau-equivalence. It is also shown that (1) alpha is not a measure of unidimensionality or homogeneity, (2) there are criticisms on understanding alpha as a measure of internal consistency, (3) alpha depends on a variety of characteristics including number of items, and (4) alpha is not a greatest lower bound. Alternative measures of reliability and future issues are discussed.

Key Words : reliability, Cronbach's alpha, split-half method, essential tau-equivalence

測定の信頼性は、実証的研究のおよそすべてに関わる問題である。本稿ではまず、信頼性の観点から直近1年間の教育心理学研究を概観する。近年のわが国における心理学諸分野のレビューからも確認されたように、研究場面において最もよく利用される信頼性の指標として Cronbach の α 係数がある。しかし、 α 係数とはどのような指標なのかについては、心理学者の間で必ずしも理解がされていなかったり、誤解がされていることも少なくない。そこで本稿では、 α 係数がどのように解釈できる指標なのか、またどのように解釈してはいけない指標なのかを論じる。具体的には、前者については α 係数が (1) 可能なすべての折半法による信頼性の平均であること, (2) 信頼性の下界の一つであること, (3) 本質的タウ等価の条件のもとで信頼性と一致することを述べる。後者については、 α 係数が (1) 大きいことが一次元性 (等質性) の根拠とはならないこと, (2) 内的一貫性の指標とされることが多いが近年批判も高まっていること, (3) 項目数など様々な要因に依存すること, (4) 信頼性の「下限」ではないことを述べる。最後に、 α 係数に代わる信頼性の推定法と今後の展望、そして信頼性を高めるような測定の重要性を述べる。

キーワード：信頼性, Cronbach の α 係数, 折半テスト, 本質的タウ等価

はじめに

測定の信頼性は、(データを) 測定することによって行われる研究すべてに関わる問題である。—— Cronbach (1951) による α 係数の論文は、このような書き

出しではじまる (括弧内筆者)。その出版から 60 年以上を経ても、この含意はまったく色あせていない。およそデータを取得して行う研究において、測定の信頼性および妥当性が確認されていることは、実証的な研究を行う上で担保されるべき前提であり、また基盤でも

ある。

直接物理的な測定をすることが困難な対象を扱うことの多い教育心理学研究では、諸科学の中でもとくに信頼性と妥当性をめぐる議論が重要視されるといってよいであろう。心理学研究における信頼性にまつわる研究の歴史は長い。しかし、その評価・検証の方法論をめぐる問題は、依然として今日的な問題であり続けている。1年間の教育心理学研究を振り返る上で、信頼性の観点は一本の縦糸になり得る。

とくに、尺度を作成する研究においては、その信頼性の検証は不可欠な工程の一つと言える。直近1年間の『教育心理学研究』誌を見ても、山田・小泉・中山・宮原 (2013) では小中学生用規範行動自己評定尺度を開発し、中間 (2013) は恩恵享受的自己感尺度を作成し、また村上・櫻井 (2014) は児童用アタッチメント機能尺度を作成して、それぞれ Cronbach の α 係数によってその信頼性を検証している。堀口・小玉 (2014) は社会的活動参加動機づけ尺度を作成し、こちらでは初回の4週間後に再検査を依頼することによって再検査信頼性も検証している。この際には、(1) 初回調査の回答を封筒に入れ厳封し、封筒に自身の回答だと判断できる記号の記入を求めて回収した後に、(2) 初回調査の回答を取り出し、初回調査の質問紙・再検査用の質問紙・封筒に同じ通し番号を付記して再検査用の質問紙を封筒に入れ、(3) 4週間後に回答者が記載した記号をもとにそれぞれの回答者に再度配付回収する、という工夫により、個人情報保護と再検査信頼性の検証を両立している。

同様に、本年度の日本教育心理学会第56回総会においても、尺度開発の研究ではその大半で信頼性が検討されている (井原・牧, 2014; 輕部・佐藤・杉江, 2014; 河村・河村, 2014; 光浪, 2014; 武蔵・河村, 2014; 鍋田・山本, 2014; 杉本, 2014; 田口・橋本・三浦・熊谷, 2014; 高橋・森本, 2014; 湯浅・大久保, 2014)。

またもちろん尺度を開発する研究だけでなく、既存の尺度を用いたり、再構成したりして研究を行う上でも、研究対象となった標本における信頼性を検討することは重要である。直近1年間の『教育心理学研究』誌に掲載された論文の中でも、たとえば石井 (2013) や中間 (2013) においては先行研究を引用する形で、使用した尺度の信頼性が担保されていることを明示的に述べている。

信頼性に関する議論は、もちろんいわゆる心理尺度だけにとどまるものではない。たとえば、栗田・楠見 (2014) においては、皮膚電位や心拍といった生理指

標を用いて障がい者に対する無意識的・無意図的な態度を測定する方法論について、潜在指標としての妥当性が確保されているものの、参加者の心理的・身体的状態や、騒音などの実験室環境による影響を受けやすく、信頼性に問題があることを指摘している。また、反応時間を用いた潜在指標についても信頼性に関して多くの先行研究があり、その精度が検証されていることを指摘している。こうした測定法もちろん、教育心理学における測定の信頼性をめぐる議論の俎上にあげていくことは重要である。

本年度の総会において、高本 (2014) は、2001年から2012年までの12年間に公開された、『心理学研究』『教育心理学研究』『社会心理学研究』『実験社会心理学研究』『パーソナリティ研究』の国内誌5誌を対象とし、心理尺度を作成した論文中で報告された信頼性係数を調べた。その結果、該当するのべ159件の論文のうち、信頼性の指標として Cronbach の α 係数を報告したものが146件 (91.8%)、再検査信頼性を報告したものが66件 (41.5%) であった。その他の手法としては、項目反応理論を用いたもの (4件)、 ω 係数を用いたもの (3件)、一般化可能性理論を用いたもの (1件)、折半法を用いたもの (1件) がそれぞれあったが、その数はわずかであった。

このように、国内の心理尺度を作成する論文は、その90%以上という大部分 (ほかの指標と併用することはあっても) で、Cronbach の α 係数によって信頼性を評価していることがわかる。

一方で近年、Cronbach の α 係数を信頼性の指標として用いることへ、注意を促す論文が数多く出版されている。たとえば、2009年に心理統計学の専門誌である *Psychometrika* 誌に信頼性をめぐる一連の論文が掲載されたが、これは Sijtsma (2009a) による論文「Cronbach の α 係数の使用、誤用、そしてそのきわめて限られた有用性について (On the use, the misuse, and the very limited usefulness of Cronbach's alpha)」にはじまり、Green & Yang (2009), Bentler (2009), Revelle & Zinbarg (2009) らが基本的にそれに賛同する立場から α 係数の代替案についてのレビューや提案を行い、再度 Sijtsma (2009b) がまとめと将来への展望を行うという体裁であった。

一方で、方法論を専門とする者を除く多くの心理学研究者の間では、 α 係数は知られているものの、その意味や利用法について十分な理解が得られていることは、必ずしも多くないように思われる。また、 α 係数について論文や教科書等に書かれている説明の中には、

必ずしも現在では適当でないと考えられている事項もある。

そこで本稿では、教育心理学をはじめとする心理学研究で非常に多く利用され、またしばしば批判され、しかし多くの研究者に必ずしも十分理解されているとは言いがたい Cronbach の α 係数について、(1) α 係数はどのような歴史的経緯で誕生した指標であるか、(2) α 係数はどのように解釈できるのか、そして(3) α 係数はどのように解釈できないのか、という3点について、整理して概観することを目的とする。

なお、妥当性をめぐる教育心理学研究については、本『教育心理学年報』誌において村山 (2011) の妥当性概念を整理した網羅的なレビューと、内田 (2012) による当該年度を中心とした研究動向のレビュー、南風原 (2012) のチュートリアル論文があり、あわせて参照いただきたい。

信頼性と Cronbach の α 係数

信頼性とは

信頼性を数理的に定義する枠組みには、古典的テスト理論に基づくもの、一般化可能性理論に基づくもの、項目反応理論に基づくものなどいくつかの枠組みがある。もっとも、Cronbach (1951) が α 係数を提出した当時には、このいずれの枠組みも確立していなかったことに、注意をする必要がある。現代では、通常 α 係数は古典的テスト理論に基づいて理解され、解釈される。したがって、本稿でも古典的テスト理論の枠組みを前提に、Cronbach の α 係数について述べる。なお本節の執筆にあたり、前川 (2008)、村木 (2011)、Junker (2012) を参考にした。

古典的テスト理論 (classical test theory ; Lord & Novick, 1968 ; Novick, 1966) においては、テストの観測値 X は、真値 T と、それとは独立な測定誤差 E の和によって構成されたと考える。すなわち、

$$X = T + E \quad (1)$$

である。本稿における「テスト」は「質問紙 (による測定)」と置き換えることが可能であり、教育心理学の典型的な応用場面においては、観測値 X は質問紙の合計得点 (尺度得点) に該当すると考えてよい。また、測定誤差 E は測定時の身体的および心理的な条件、テスト時の環境といった回答に影響をおよぼす様々なランダムな要因の影響が総合されたものであり、真値 T とは独立であると考えられる。このとき、独立な確率変数の和の分散の性質より、観測値の分散 $Var(X)$ について

$$Var(X) = Var(T) + Var(E) \quad (2)$$

が成り立つ。このとき、測定の信頼性 (reliability)、もしくは信頼性係数 (reliability coefficient) とは、測定値の分散 $Var(X)$ に占める真値の分散 $Var(T)$ の割合、すなわち

$$\rho = \frac{Var(T)}{Var(X)} = 1 - \frac{Var(E)}{Var(X)} \quad (3)$$

と定義される。

この定義より、測定誤差に由来する分散である $Var(E)$ が相対的に小さければ小さいほど、テストの信頼性は大きくなることがわかる。仮に測定誤差の分散 $Var(E)$ が 0 であれば、信頼性係数 ρ は 1 となり、その測定は事実上いつも直接真値を測定できていることになる。

また、信頼性係数 ρ は定義上、その最小値が 0、最大値が 1 である。すなわち ρ は

$$0 \leq \rho \leq 1 \quad (4)$$

を満たす。

しかし、古典的テスト理論における観測値 X についての (1) 式はあくまでモデルであり、実際には分析者は観測値 X のうちどれだけが真値でどれだけが測定誤差かを直接的に知ることができない。この困難の中で、(3) 式の信頼性係数 ρ をデータからどのように求めるかという問題に、心理学は古来から取り組んできた。そして、以下で述べるような平行テスト法、再テスト法、折半法といった様々な方法が考案されてきた。

平行テスト法

平行テスト法 (parallel test method) とは、真値が等しく、また誤差の分散も等しいテスト (これを平行テスト parallel test と呼ぶ) を 2 セット作成し、それぞれの測定値を用いて信頼性を求める方法である。すなわち、2 度の測定値を X_1, X_2 とし、古典的テスト理論のもとで $X_1 = T_1 + E_1, X_2 = T_2 + E_2$ としたとき、2 つのテストが平行テストであるとは

$$T_1 = T_2 = T \quad (5)$$

$$Var(E_1) = Var(E_2) = \sigma_E^2 \quad (6)$$

ということである。

仮に 2 セットの平行テストを準備することができれば、

$$\begin{aligned} Cov(X_1, X_2) &= Cov(T_1 + E_1, T_2 + E_2) \\ &= Cov(T_1, T_2) + Cov(T_1, E_2) \\ &\quad + Cov(E_1, T_2) + Cov(E_1, E_2) \\ &= Var(T) \end{aligned} \quad (7)$$

となる。また、(5)式と、真値と誤差が独立であることより、

$$\begin{aligned} \text{Var}(X_1) &= \text{Var}(T) + \text{Var}(E_1) \\ &= \text{Var}(T) + \sigma_E^2 \\ &= \text{Var}(X_2) \end{aligned} \quad (8)$$

が成り立つ。したがってこのとき、(3)式の信頼性は

$$\begin{aligned} \rho &= \frac{\text{Var}(T)}{\text{Var}(X)} \\ &= \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)}\sqrt{\text{Var}(X_2)}} \\ &= \text{Cor}(X_1, X_2) \end{aligned} \quad (9)$$

となり、 X_1 と X_2 の相関係数が信頼性に一致する。よって、平行テスト間の相関係数の実現値によって、テストの信頼性を推定できることになる。

しかし、2セットのテストがたしかに平行テストであること、つまり真値と誤差分散が等しいテストになっていることを示すのは応用上非常に困難である。したがって、残念ながら現実的には平行テスト法を用いるための前提条件を満たすのは難しい。

再テスト法

再テスト法 (test-retest method) とは、平行形式のテストを2セット用意する代わりに、同一のテストを時間間隔をあけて2度測定する方法である。2度の測定を独立に行うことができれば、その2度の測定は平行テストとなり、したがって前項と同じ議論により、2度の測定の相関係数によって真の信頼性を推定することができる。

しかしながら、現実にはテストの回答者に1度目の測定を行ったことによる学習効果や順序効果が生じてしまい、2度の測定が独立であるという前提条件を満たすことは困難な場合が多い。したがって再テスト法も同様に、実用上の問題を抱えていることになる。

折半法

以上述べたように、平行テスト法や再テスト法の前提を実際のテストで満たすのは困難である。そこで、古くから使われていたのが折半法 (split-half method) である。これは、平行テストを作成したり、同じテストを2度実施する再テストを行ったりする代わりに、一組のテストに含まれる項目を、2つの部分 X_1 と X_2 とに折半するものである。折半した2つの部分の合計得点は $X = X_1 + X_2$ である。また、 X_1 と X_2 との相関係数を $\text{Cor}(X_1, X_2)$ とする。このとき、Spearman-Brown の公式 (Brown, 1910; Spearman, 1910) を用いると、折半法による信頼性 ρ_{sh} を

$$\rho_{sh} = \frac{2\text{Cor}(X_1, X_2)}{1 + \text{Cor}(X_1, X_2)} \quad (10)$$

と求めることができる。Spearman と Brown によっ

てこの公式が発表されて以来、折半法は40年間にわたってテストの信頼性を推定するうえで標準的な方法であった、と Cronbach (1951, p.298) は述べている。

なお、Spearman-Brown の公式は、テストを折半つまり2分割を行う代わりに k 個の部分に分割し、そのうちの X_i , X_j を利用して信頼性を求める場合には、

$$\rho_{sh2} = \frac{k\text{Cok}(X_i, X_j)}{1 + (k-1)\text{Cor}(X_i, X_j)} \quad (11)$$

と一般化した形で表すことができる。この形をもって Spearman-Brown の公式と呼ばれることもある。(10)式の折半法による公式は、(11)式において分割数 k を $k=2$ とおいた場合に相当することがわかる。

しかし折半法では、信頼性係数の推定値はテストに含まれる項目をどのように折半するかに依存する。したがって、 ρ_{sh} がただ一つのユニークな推定値とならない。たとえば20項目からなるテストの信頼性を求めるとき、折半法では10項目ずつからなる X_1 , X_2 の2セットを作ることになる。このとき、 X_1 にどの10項目を採用するかによって、折半法による信頼性 ρ_{sh} は変わってしまう。Brownell (1933) や Kuder & Richardson (1937) らは、このようにユニークな推定値を与えないという点で折半法による信頼性を批判した。

そこで Kuder & Richardson (1937) は、Spearman-Brown の公式を一般化し、2値データを対象に、一度だけ測定したデータから求められることのできる Kuder-Richardson の20式 (KR-20) として知られる信頼性の推定量を提案した。また Guttman (1945) は、信頼性の下界を研究する中で、同じ推定量に到達した。なお下界については次節で議論する。

この信頼性の推定量を α (coefficient alpha, α 係数) と命名し、その性質を深く議論したのが Cronbach (1951) である。以上の経緯からもわかるように、Cronbach の α 係数自体は、Cronbach (1951) のオリジナルというよりも、それ以前から知られていた量である。

Cronbach の α 係数

Cronbach の α 係数は、最もよく知られた信頼性の指標 (信頼性係数を推定するための指標) と言って過言ではないであろう。Sijtsma (2009a) は、Cronbach の α 係数の論文 (Cronbach, 1951) は、掲載誌である *Psychometrika* 誌のウェブサイトから最も多くダウンロードされた論文の一つであり、またその引用回数 (Web of Science による) は20世紀の最も偉大な発見と

呼ばれる Watson & Crick (1953) による DNA の二重らせん構造を発見した論文すらをも上回ることを報告している。

なお、Cronbach (2004) は、2001 年 6 月に行われた Cronbach と Richard J. Shavelson との対談の内容を編集し、Cronbach の死後に出版された論文である (Shavelson, 2004)。2001 年は Cronbach (1951) が出版されて 50 年目にあたり、その中で Cronbach 自身から見た α 係数の発展が語られているため一読の価値がある。

1951 年の Cronbach の論文に話を戻す。彼は、Kuder & Richardson (1937) の提案する 20 式に注目し、これを 2 値変数に限らず一般化したうえで α 係数と呼ぶことにした。 α という命名の由来は、この後に β , γ , と次なる信頼性係数の研究を進めていく予定であったからであるとされる。

α 係数は、テスト項目数を n として、 i 番目の項目の得点 $X_i (i=1, \dots, n)$ の分散 $Var(X_i)$ 、およびすべての項目の合計得点 $X = \sum_{i=1}^n X_i$ の分散 $Var(X)$ の関数となっている。具体的には α 係数は次によって定義される。

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n Var(X_i)}{Var(X)} \right) \quad (12)$$

なお、あまり望ましくないことではあるが表記の都合上、本項では添え字付きの X_i が表す対象が i 番目の項目の得点であったり、 i 番目の部分テストであったり、 i 番目の平行テストであったりしているの、その文脈に注意していただきたい。ここでの X_i は i 番目の項目の得点である。

先にも紹介した Sijtsma (2009a) は Kuder & Richardson (1937) の被引用回数が 400 を超えることも指摘し、もちろん Cronbach (1951) には及ばないものの十分尊敬に値する数字であると述べている。

Cronbach の α 係数は何であるか

Cortina (1993) は、 α 係数についてよく語られる性質を以下の 5 つにまとめている：(a) α 係数はすべての折半法による信頼性の平均である (b) α 係数は信頼性の下界である (c) α 係数は本質的タウ等価のもとで信頼性と一致する (d) α 係数は一次元性の指標である (e) α 係数は KR 20 を一般化したものである。

このうち、(e) についてはすでに述べたとおり、KR 20 を 2 値データ以外へと一般化したのが α 係数といえる。また、(d) については、この考えは次節で

述べるように、あまり適切ではないと考えられる。そこで、本節では残りの (a)~(c) の性質にしぼって、Cronbach の α 係数とは何であるのかをみていく。

α 係数はすべての折半法による信頼性係数の平均である。

これは Cronbach (1951) が強調した、 α 係数についての重要な性質である。とくに、折半した 2 つの部分得点の分散が等しい場合には、(1) 式の Spearman-Brown の公式によって求めた信頼性係数の平均が、 α 係数と一致する。また、折半した 2 つの部分得点の分散が等しくない場合は、この相違を考慮に入れて Spearman-Brown の公式を一般化した Flanagan (1937) と Rulon (1939) の公式

$$\rho_{sh3} = \frac{4Cor(X_1, X_2)\sqrt{Var(X_1)Var(X_2)}}{Var(X)} \quad (13)$$

によって折半法による信頼性係数を求めると、やはりこの平均が α 係数と一致する。

先に述べたとおり、折半法による信頼性係数の推定は、Cronbach (1951) 以前は幅広く利用されてきた方法であった。しかし、折半法による信頼性係数は、項目全体をどのように折半するか、というその仕方、つまりどの項目が (たまたま) どちらの群に入ったのかということに依存して変動してしまう。しかし、可能なすべての折半の仕方を考えて信頼性の平均をとれば、その平均値はただ一つのユニークな値である。そして、それが α 係数と一致する、というわけである。このことは、信頼性の指標としての α 係数の地位を保障する性質であると解釈できる。

α 係数は信頼性の下界の一つである。

α 係数は信頼性の下界の一つである、というこの標題が意味するところはつまり、信頼性 ρ に対して、 α 係数が

$$\alpha \leq \rho \quad (14)$$

を満たすということである。このことは、Cronbach (1951) よりも遡り、Guttman (1945, p.257) により示されている。Guttman もやはり Cronbach と同じ動機づけを持っていたようであり、論文中でテストを用いた独立な測定を 2 度繰り返すことは現実的に困難なため、1 度だけの測定からでも入手できる情報は何かを考えるのが重要であること、そして信頼性の下界ならば 1 度だけの測定からでも求められることを強調している。そして、彼は λ_1 から λ_6 と呼ぶ、6 種類の信頼性の下界 (lower bound)、つまり真の信頼性を超えることはない下側からの推定量を導入した。 α 係数は、この Guttman が与えた 6 つの下界のうちの 3 番目の

下界 λ_3 に対応する。

Guttman (1945) の論文においては、彼の 6 つの下界のうち、 λ_1 から λ_3 までの 3 つは一連の流れの中で導入されている。すなわち、彼はまずもっともシンプルに与えられる下界 λ_1 を述べ、次にそれよりもより過小推定のバイアスが小さい下界 λ_2 を導入する。しかしここで、 λ_1 の算出には項目ごとの分散 $Var(X_i)$ と全体の分散 $Var(X)$ だけしか必要としないのに対し、 λ_2 の算出にはこれらに加えて各項目間の共分散 $Cov(X_i, X_j)$ も必要であった。現代であれば分散や共分散はコンピュータを用いて瞬時に計算することができるが、当時はとくに項目の数が多い場合、項目間のすべての共分散を計算するのは手間であったに違いない。そこで、項目間の共分散を必要としないが、 λ_1 よりも過小推定のバイアスが小さい下界 λ_3 が導入されるのである。

Guttman は、これら 3 つの下界の間には

$$\lambda_1 < \lambda_3 \leq \lambda_2 \leq \rho \quad (15)$$

の関係があることを示している。

ここで λ_3 が α 係数であることを先に述べた。したがって(15)式は、 λ_3 つまり Cronbach の α 係数には真の信頼性 ρ を過小推定するバイアスがあることを含意している。もちろん、過小推定の度合いは、小さければ小さいほど望ましい。そして、 α 係数 ($=\lambda_3$) を含む 3 つの下界の間には(15)式の関係がある。したがって、Guttman (1945) の時代からすでに、 α 係数よりも過小推定の程度の意味で望ましい信頼性の指標 λ_2 があることはわかっていたのである。

ただし前述のとおり、 α 係数 ($=\lambda_3$) の算出は項目ごとの共分散の計算を必要としないという点で、 λ_2 よりも容易である。この手間の差は、計算機の使える現代では全く問題にならないものである。しかし、発表当時においては現実的な問題であったであろうことが、Guttman (1945) の筆の運びよりうかがえる。

α 係数は本質的タウ等価のもとで真の信頼性に一致する

この標題は、(14)式の不等号において等号が成り立つ条件、つまり 1 度の測定から求めた α 係数が、信頼性と一致するための条件についての命題であり、Novick and Lewis (1967) によって示された。

ここでタウ等価と言うところのタウとはテストの真値のことであり、(1)式において T によって表した量である。「本質的」がつかない、タウ等価 (tau-equivalent) な 2 度の測定とは、先に述べた平行テストの前提から、誤差分散が等しいという仮定を落とした

ものである。つまり、タウ等価な測定では $Var(E_1) \neq Var(E_2)$ であることが許される。そして、本質的タウ等価 (essential tau-equivalent) な 2 度の測定とは、タウ等価の仮定から、さらに 2 度の測定間の真値が、個人によらない定数 b だけ異なることを許す場合である。つまり、そこではさらに $T_1 = T_2 + b$ という関係が許される。

Cronbach の α 係数は、この本質的タウ等価という条件が満たされるとき (もちろんそれより仮定の強いタウ等価の条件が満たされるときにも) に、信頼性 ρ に一致する。タウ等価もしくは本質的タウ等価な 2 つのテストのことを、平行テストの要件よりは仮定が緩いという意味で、弱平行テストと呼ばれることもある。本質的タウ等価の仮定は依然として強い仮定であるが、平行テストの仮定に比べればかなり仮定を弱め、より現実的に即していると言える (Furr & Bacharach, 2013)。

Cronbach の α 係数への批判

α 係数は必ずしも一次元性・等質性を反映しない

α 係数の値が大きいことをもって、尺度が一因子的・一次的 (unidimensional) であることの証拠とする議論は、わが国に限らず論文等でしばしば見られる。しかし実際には、 α 係数が大きいことをもって一次元性の証拠とすることは必ずしも適切ではない (Revelle, 1979; Schmitt, 1996; Zinbarg, Revelle, Yovel, & Li, 2005)。信頼性の研究では等質性 (homogeneity) という用語で一次元性 (unidimensionality) のことを表すので (Cortina, 1993; Schmitt, 1996)、 α 係数が大きいことを尺度の等質性の証拠として扱うこともまた適切ではない。本節ではこれらのことを、まず数式によって、ついで具体例を使って述べる。

(12)式によって与えられる、 α 係数の定義にもう一度戻ることにする。ここで、分母の合計得点の項目全体の分散 $Var(X)$ について、確率変数の分散の和の性質より、以下が成立する。

$$\begin{aligned} Var(X) &= \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) \\ &= \sum_{i=1}^n Var(X_i) + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} Cov(X_i, X_j) \quad (16) \end{aligned}$$

これを使うと、(12)式の Cronbach の α 係数は以下のように書き換えることができる。

$$\alpha = \frac{n}{n-1} \left(\frac{2 \sum_{i=2}^n \sum_{j=1}^{i-1} Cov(X_i, X_j)}{Var(X)} \right) \quad (17)$$

この分子は、 $\sum_{i=2}^n \sum_{j=1}^{i-1} Cov(X_i, X_j)$ は、異なる項目間すべての共分散の総和である。別の言い方をすれば

これは項目間の共分散行列の下三角部分の総和である。

(7)式は、Cronbach の α 係数は、分子にある「すべての項目間の共分散の総和」の単調関数になっていることを表している。総和をその足し合わせた数という定数で割れば平均になるので、 α 係数は「すべての項目間の共分散の平均」の単調関数である、と述べても同じことである。したがって、項目数と合計得点の分散を一定とすれば、テストの背後にある構造が何次元的存在であるかによらず、項目間の共分散の総和が同じであれば α 係数は同じ値をとることになる。

Table 1 はこのことを示す具体例である。Table 1 には 3 種類の、次元性は異なるが同じ α 係数の値を持つ相関行列を示した ((a), (b) は Schmitt, 1996 より)。6 項目に対して、(a) の相関行列はすべての項目間の相関係数が 0.5 で、一次元的 (unidimensional) な場合である。(b) の相関行列は項目 1, 2, 3 の間、および項目 4, 5, 6 の間の相関係数が 0.8 であり、それ以外の項目間の相関係数は 0.3 という、二次元的 (bidimensional) な場合である。そして、(c) の相関行列は、項目 1 と 2、項

Table 1 Sample correlation matrices which have the same Cronbach's alpha values.

(a) Unidimensional case ($\alpha = .86$)

Item	1	2	3	4	5	6
1						
2	.50					
3	.50	.50				
4	.50	.50	.50			
5	.50	.50	.50	.50		
6	.50	.50	.50	.50	.50	

(b) Bidimensional case ($\alpha = .86$)

Item	1	2	3	4	5	6
1						
2	.80					
3	.80	.80				
4	.30	.30	.30			
5	.30	.30	.30	.80		
6	.30	.30	.30	.80	.80	

(c) Tridimensional case ($\alpha = .86$)

Item	1	2	3	4	5	6
1						
2	.90					
3	.40	.40				
4	.40	.40	.90			
5	.40	.40	.40	.40		
6	.40	.40	.40	.40	.90	

Note. (a) and (b) are taken from Schmitt (1996, Table 1).

目 3 と 4、項目 5 と 6 の間の相関係数がそれぞれ 0.9 であり、それ以外の項目間の相関係数は 0.4 という三次元的 (tridimensional) な場合である。

これら 3 つの内的な構造 (次元性) は文字通り三者三様に異なる。しかし、このすべてにおいて、 α 係数の値は $\alpha = 0.86$ と等しい値となるのである。これは、下三角部分の 15 個の共分散の総和が、(a) では $0.50 \times 15 = 7.50$ 、(b) では $0.80 \times 6 + 0.30 \times 9 = 7.50$ 、(c) では $0.90 \times 3 + 0.40 \times 12 = 7.50$ と、すべて同じ値になるからである。このように、 α 係数はテストの内的な構造を反映するのではなく、むしろ項目間共分散の総和 (もしくは平均) を反映するにすぎないのである。

以上のように、Cronbach の α 係数は一次元性・等質性の低いデータでも大きくなる。この意味で、 α 係数を一次元性・等質性の証拠とすることは適切ではないと言える。

α 係数を内的整合性・内的一貫性の指標と呼ぶことには批判がある

洋の東西を問わず、Cronbach の α 係数は内的整合性 (内的一貫性, internal consistency) の指標であると説明されることは多い。教育心理学研究誌でもこのような記載は少なからず見られる。しかし、前項で説明したように α 係数は一次元性の指標とは言えないことから、 α 係数を内的整合性の指標と呼ぶことを批判する声も高まっている (Sijtsma, 2009a)。わが国においてもすでに、たとえば南風原・市川・下山 (2001, p.77) は、「この係数 (著者注: α 係数) は内的整合性が低くても、項目数がある程度多ければ高い値をとるため、純粋に内的整合性の指標とは言えない」と述べている。

そもそも Sijtsma (2009a) が指摘するように、「内的整合性」という言葉の定義は曖昧である。それが何を指すのか、研究者の間でも数理的に明確な定義は存在していない。これは「信頼性」や「一次元性」が数理的に定義され、量的に評価されるのとは対照的である。心理学の研究論文でも、Cronbach の α 係数を内的整合性の指標であるとするものは多いが、内的整合性という概念について深い議論がなされることは少ない。中には、 α 係数で測っているものこそが内的整合性である、といった同義循環的な議論すら見られることがある。

α 係数が内的一貫性の指標であると言われるようになったのは、Cronbach (1951) が自ら論文内でそう述べたからである。内的一貫性という用語はそれ以前から存在していた。しかしながら、冒頭でも述べたとおり、当時は現代のように信頼性という概念がそもそも

数理的に明確に確立されていなかった。それどころか、古典的テスト理論 (Lord & Novick, 1968 ; Novick, 1966) の枠組みすらも存在していなかったのである。

Cortina (1993) は、信頼性をめぐる概念を整理し、Cronbach (1951) が同義的に用いていた「等質性」 (homogeneity) と「内的整合性」 (internal consistency) を明確に区別する必要があると論じた。Hattie (1985) も同じ趣旨を論じている。

現在では、内的整合性は、項目が相互に関連している度合いを指すとされる (Crano & Brewer, 1973 ; Green, Lissitz, & Mulaik, 1977)。一方で、等質性 (homogeneity) とは1次元性のこと、つまり項目の背後にある因子が1因子と考えられる度合いを指す (Green et al., 1977 ; Lord & Novick, 1968)。ここから、Green et al. (1977) や Cortina (1993), Schmitt (1996) らが指摘するように、内的整合性は等質性の必要条件であるが、十分条件とは言えないことになる。

Schmitt (1996) は、 α 係数を一次元性の指標とすることは明確に否定する一方で、内的整合性については詳しく論じていない。これは上記のようにその定義が数理的に明確でないことにも由来すると考えられる。

一方、Sijtsma (2009a) は α 係数を内的整合性の指標と呼ぶことにははっきりと反対している。その理由は、大別すると、第1に信頼性の推定量としては α 係数よりもよい指標が存在するから、第2に前項でも見たように α 係数の値がテストの内的な構造だけを単純に反映しているとは言いがたいからである。そして、これを受けた Bentler (2009), Green & Yang (2009), Revelle & Zinbarg (2009) らの論文を再度概観した上で、Sijtsma (2009b, p.169) はこうした後続の著者らも α 係数が内的一貫性の指標ではないという自分の考えに基本的に賛同したようだ、とまとめている。

筆者は α 係数が内的整合性の指標ではない、という強い立場をとるものではないが、内的一貫性が数理的に明確な用語ではないことには賛同する。 α 係数を内的一貫性の指標だと主張する場合には、その定義と根拠を明確に論じていく必要があると思われる。

α 係数は項目数などさまざまな要因に依存する

多くの推定量がそうであるように、 α 係数は様々な要因の影響を受けてその値が変化する。とくに項目数については、分散の和の性質などを利用して、Cronbach の α 係数を以下のように式変形するとわかりやすい。まず、 α 係数は

$$\alpha = \frac{n\bar{C}}{\bar{V} + (n-1)\bar{C}} \quad (18)$$

と変形できる。ここで、 $\bar{V}$ は n 個の項目分散の平均 $\bar{V} = \sum_{i=1}^n \text{Var}(X_i)/n$ であり、 $\bar{C}$ は $\frac{n(n-1)}{2}$ 組の項目間の共分散の平均である。これをさらに相関係数の式の形にすると、次のようになる。

$$\alpha = \frac{n\bar{r}}{1 + (n-1)\bar{r}} \quad (19)$$

ここで、 $\bar{r}$ は $\frac{n(n-1)}{2}$ 組の項目間の相関係数の平均である。相関係数の平均が $\bar{r} = 0.2, 0.5, 0.8$ のそれぞれの場合について、(19) 式の形で書いた α 係数を、項目数 n に対してプロットすると、Figure 1 になる。ここからわかるように、Cronbach の α 係数は、項目間の平均相関係数が一定であったとしても、テストに含まれる項目数 n が多ければ多いほど、より大きな値をとる。したがって、単に Cronbach の α 係数の値を大きくしたいならば、テストの項目数を多くした方が有利である、ということになってしまう。

このほか、 α 係数は Likert 尺度のカテゴリ数や誤差の構造にも依存すること (Jenkins & Taber, 1977)、尺度を作る上で「望ましくない」項目を削除することが不当に α 係数の値を高めうること (Kopalle & Lehmann, 1997) など、必ずしも望ましくない性質が先行研究において指摘されている。

α 係数は信頼性の「下限」ではない

本稿ではすでに、 α 係数が再検査によって得られる真の信頼性を、1回の測定だけから得られる量を用いて下側から推定する、信頼性の下界 (lower bound) の一つであることを述べた。わが国の文献においては、しばしば α 係数が信頼性の下限である、と述べられているものがある。しかしながら厳密に言えば、数学的には下界 (lower bound) と下限 (infimum) は異なる概念である。たとえば、ある集合 $\{8, 5, 3\}$ に対して、「3」はこの下限である。一方、「2」は下界であるが下限ではない。「1」も同様である。このように、下限はただ1つに決まるが、下界は多数存在しうる。下限とは下界のうちで最大のものを指す用語である。したがって下限のことを最大下界 (greatest lower bound) と呼ぶこともある。下限と下界の違いを、Figure 2 に模式的に示した。

Cronbach の α 係数と信頼性の下限 (下界ではなく) との関係については、Revelle & Zinbarg (2009) に詳しい。Cronbach の時代に等質性と内的整合性が区別されていなかったのと同じように、わが国においても以前は下界と下限という用語があまり区別されてい

Figure 1 Cronbach's alpha as a function of number of items, when the mean correlation coefficients among items are 0.2, 0.5, and 0.8.

Figure 2 Schematic diagram of the difference between infimum (下限) and lower bounds (下界).

かったのかもしれない。もしくは、敢えて厳密だが難解な下界という語を避け、より日常用語に近い下限という言葉が用いられていたのかもしれない。しかし、諸分野間の交流や越境が進む現代においては、数学的に適切な表現を用いて、今後は α 係数を信頼性の下界と呼んでいくことが望ましいのではないかと考える。

結論と今後の展望

本稿のまとめ

本稿では、まず信頼性の定義を確認した後に、Cronbach (1951) の論文から 60 年以上を経てなお、広く利用し続けられている信頼性の指標である α 係数について、その歴史を簡単に紐解き、平行テストや折半法といったそれ以前の信頼性指標における問題点を克服する指標として提案された経緯を概観した。続いて、 α 係数がどのように解釈できる量なのか、そしてどのように解釈してはいけない量なのかを概観してきた。

α 係数の適切な解釈としては、 α 係数はすべての折半法による信頼性の平均になっていること、信頼性の下界の一つであること、そして本質的タウ等価という条件のもとで信頼性に一致することを述べた。また α 係数の必ずしも適切でない解釈としては、 α 係数が高いことを一次元性 (等質性) の証拠とはできないこと、 α 係数は内的一貫性の指標とされることが多いが近年批判も高まっていること、 α 係数は様々な要因に

依存し、とくに項目数の単調増加関数となること、そして α 係数は信頼性の下界であるが、下限ではないことを論じた。

α 係数にカットオフ基準やゴールデンルールを設ける議論はしばしばみられる。すなわち、 α 係数について 0.8 や 0.9 といったカットオフ基準を設け、これよりも大きいことをもって、信頼性が担保されたとする議論である。しかしながら、ここまで述べてきたように α 係数は信頼性の一側面をとらえるものではあるが、その大きさが次元性を意味しないこと、項目数の単調関数となることなどを考慮すると、このようなカットオフには明確な意味は見いだしにくいとも考えられる。

近年、心理学における統計改革の流れにおいて、仮説検定と p 値への過度な依存からの脱却が広く叫ばれている (大久保, 2009)。同じように、信頼性の推定においても、Cronbach の α 係数からの脱却を考えると、 α 係数を補完・代替する方法としては、一般化可能性理論に基づくもの (いわゆる G 研究, Cronbach, 2004 ; Strube, 2000), 階層因子分析に基づくもの (McDonald, 1999), 共分散構造分析の枠組みを用いるもの (Yang & Green, 2010) など、様々な方法が提案されている。また、これらを比較検討する研究もなされている (岡田, 2011 ; Revelle & Zinbarg, 2009 ; Zinbarg et al., 2005)。こうした新しい研究のさらなる進展が望まれるところである。

信頼性の高い測定の指向

α 係数や上記のような方法を用いて信頼性を推定・確認することは、研究者にとってデータをとった後の作業である。しかし応用研究上、信頼性の評価と同じくらい、もしくはそれ以上に大切なのは、データを取得する以前の段階における、信頼性の高い測定を目指すための努力であろう。

2011 年の本学会の研究委員会企画チュートリアルセミナーは、妥当性をテーマとしたものであった。その企画趣旨の一部を、長くなるが以下に引用する。“本邦における現状はというと、型にはまった安易な方法によって作成された尺度に関して、これまた型にはまった極めて安易な検討をただで‘妥当性が確認された’という判断が下され、それで‘みそぎ’が済んだとみなしていると推察される研究が多く見られる。そして、このような甘い判断のもとに多くの構成概念とそれを測定するための (妥当性に関して多くの問題があると考えられる) 尺度が提唱・作成され、‘類似した構成概念および尺度の乱立’という望ましからざる状況も

生じてしまっていると思われる” (吉田・石井, 2012)

引用文中の言葉は強いが、こうした妥当性についての現状は、残念ながら信頼性についても同様に見られると言って過言ではないであろう。そして、Cronbach の α 係数は、信頼性についての「みそぎ」のための便利な道具として、使われてしまっている現状があるように思われる。

同チュートリアルセミナーにおいて、南風原 (2012) は村上 (2006) を引用しつつ、データ収集前の、妥当性の高い心理尺度を作成するための研究者の工夫と努力、すなわち「妥当化 (validation)」の重要性を指摘している。全く同じことを、信頼性についても指摘してよいであろう。信頼性の高い測定を目指す研究者の工夫と努力が、よい心理学研究には不可欠である。そのための方法を逐一述べるには本稿には紙幅が不足しているが、たとえば Downing & Haladyna (2006 池田訳 2008) には、(その他の問題形式よりも) 多肢選択式の問題を用いること、採点規程を注意深くかつ一貫して適用すること、項目反応理論モデルを用い、かつ事前にサンプルサイズ設計を行うこと、といった信頼性を高めることのできる方法が、先行研究に基づく知見とともに数多く紹介されている。こうした先人の研究成果の蓄積を咀嚼し活用して、信頼性の高い測定を追求していくことが、信頼性の評価・検証と同じくらい、もしくはそれ以上に、実際の心理学研究では大切であると考ええる。

謝 辞

本研究は一部に JSPS 科研費 24730544, 26285153 の助成を受けて実施しました。ここに記して謝意を表します。

引用文献

- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, **74**, 137-143.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, **3**, 296-322.
- Brownell, W. A. (1933). On the accuracy with which reliability may be measured by correlating test halves. *Journal of Experimental Education*, **1**, 204-215.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications.

- Journal of Applied Psychology*, **78**, 98-104.
- Crano, W. D., & Brewer, M. B. (1973). *Principles of research in social psychology*. New York, NY : McGraw-Hill.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**, 297-334.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, **64**, 391-418.
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of Test Development*. Mahwah, NJ : Lawrence Erlbaum Associates (ダウニング, S. M.・ハラディナ, T. M. 池田 央 (監訳) (2008). テスト作成ハンドブック 教育測定研究所)
- Flanagan, J. C. (1937). A proposed procedure for increasing the efficiency of objective tests. *Journal of Educational Psychology*, **28**, 17-21.
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics : An introduction* (2nd ed.). Thousand Oaks, CA : Sage Publications.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, **37**, 827-838.
- Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha : A cautionary tale. *Psychometrika*, **74**, 121-135.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, **10**, 255-282.
- 南風原朝和 (2012). 尺度の作成・使用と妥当性の検討 教育心理学年報, **51**, 213-217.
- 南風原朝和・市川伸一・下山晴彦 (2001). 心理学研究法入門—調査・実験から実践まで— 東京大学出版会
- Hattie, J. (1985). Methodology review : Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, **9**, 139-164.
- 堀口康太・小玉正博 (2014). 老年期の社会的活動における動機づけと well-being (生きがい感) の関連—自律性の観点から— 教育心理学研究, **62**, 101-114.
- 井原啓裕・牧 郁子 (2014). 教師の指導文化尺度作成の試み—信頼性・妥当性の検討— 日本教育心理学会第 56 回総会発表論文集, 165.
- 石井 僚 (2013). 青年期において死について考えることが時間的態度に及ぼす影響 教育心理学研究, **61**, 229-238.
- Jenkins, G. D., & Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, **62**, 392-398.
- Junker, B. W. (2012). Some aspects of classical reliability theory & classical test theory. March 1, 2012. <<http://www.stat.cmu.edu/~brian/PIER-methods/For%202013-02-26/Readings/CTT-formulas.pdf>> (December 2, 2014)
- 河村昭博・河村茂雄 (2014). 中学校教師のユーモア態度尺度の作成 日本教育心理学会第 56 回総会発表論文集, 760.
- 輕部雄輝・佐藤 純・杉江 征 (2014). 大学生における就職活動維持過程尺度作成の試み(2)—就職活動経験者の 2 時点からの回顧による妥当性の検討— 日本教育心理学会第 56 回総会発表論文集, 606.
- Kopalle, P. K., & Lehmann, D. R. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes*, **70**, 189-197.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, **2**, 151-160.
- 栗田季佳・楠見 孝 (2014). 障害者に対する潜在的態度の研究動向と展望 教育心理学研究, **62**, 64-80.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA : Addison-Wesley.
- 前川眞一 (2008). 古典的テスト理論 2008 年 9 月 6 日<<http://www.ms.hum.titech.ac.jp/mayek-awadocs/testt5.pdf>> (2014 年 12 月 2 日)
- McDonald, R. P. (1999). *Test theory : A unified treatment*. Mahwah, NJ : Lawrence Erlbaum Associates.
- 光浪睦美 (2014). レジリエンス形成要因間の関係性に関する研究(1)—尺度の作成と信頼性・妥当性の検討— 日本教育心理学会第 56 回総会発表論文集, 286.
- 村上 隆 (2006). テストの作成と妥当性検証—データ解析の視点から— 日本テスト学会第 7 回研究会報告資料 2006 年 9 月 9 日<<http://www.jar->

- test.jp/pdf/jirei7_rep2.pdf> (2014年12月2日)
- 村上達也・櫻井茂男 (2014). 児童期中・後期におけるアタッチメント・ネットワークを構成する成員の検討—児童用アタッチメント機能尺度を作成して—教育心理学研究, **62**, 24-37.
- 村木英治 (2011). シリーズ〈行動計量の科学〉項目反応理論 朝倉書店
- 村山 航 (2011). 妥当性—概念の歴史的変遷と心理測定学的観点からの考察—教育心理学年報, **51**, 118-130.
- 武蔵由佳・河村茂雄 (2014). 小学生版友人グループ尺度の作成 日本教育心理学会第56回総会発表論文集, 841.
- 鍋田智広・山本晃輔 (2014). 実践的批判的思考態度尺度作成の試み 日本教育心理学会第56回総会発表論文集, 670.
- 中間玲子 (2013). 自尊感情と心理的健康との関連再考—「恩恵享受の自己感」の概念提起—教育心理学研究, **61**, 374-386.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, **3**, 1-18.
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, **32**, 1-13.
- 岡田謙介 (2011). クロンバックの α に代わる信頼性の推定法について—構造方程式モデリングによる方法・McDonaldの ω の比較—日本テスト学会誌, **7**, 37-50.
- 大久保街亜 (2009). 日本における統計改革—基礎心理学研究を資材として—基礎心理学研究, **28**, 88-93.
- Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, **14**, 57-74.
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb : Comments on Sijtsma. *Psychometrika*, **74**, 145-154.
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, **9**, 99-103.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, **8**, 350-353.
- Shavelson, R. J. (2004). Editor's preface to Lee J. Cronbach's "My current thoughts on coefficient alpha and successor procedures" Lee J. Cronbach 1916-2001. *Educational and Psychological Measurement*, **64**, 389-390.
- Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, **74**, 107-120.
- Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, **74**, 169-173.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, **3**, 271-295.
- Strube, M. J. (2000). Reliability and generalizability theory. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding MORE multivariate statistics* (pp.23-66). Washington, DC : American Psychological Association.
- 杉本英晴 (2014). キャリア・アダプタビリティ尺度の作成 日本教育心理学会第56回総会発表論文集, 316.
- 田口禎子・橋本創一・三浦巧也・熊谷 亮 (2014). 中学生に対するメンタルヘルス尺度（こころとからだの健康気づきシート）作成の試み 日本教育心理学会第56回総会発表論文集, 925.
- 高橋 誠・森本哲介 (2014). 日本語版強みの自覚尺度の作成の試み 日本教育心理学会第56回総会発表論文集, 893.
- 高本真寛 (2014). 国内の心理尺度作成論文における信頼性係数の報告状況—2001-2012年度までに公刊された心研・教心研・社心研・実社心研・パソ研を対象に—日本教育心理学会第55回総会発表論文集, 289.
- 内田照久 (2012). 教育評価・心理測定で用いる測度の妥当性検証の機運と社会的役割を担う試験をめぐる課題解決への取り組み 教育心理学年報, **51**, 63-72.
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. : A structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
- 山田洋平・小泉令三・中山和彦・宮原紀子 (2013). 小中学生用規範行動自己評定尺度の開発と規範行動の発達的变化 教育心理学研究, **61**, 387-397.
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, **17**, 66-81.
- 吉田寿夫・石井秀宗 (2012). 企画の趣旨 研究委員会企画チュートリアルセミナー 尺度の作成・使用

- と妥当性の検討 教育心理学年報, **51**, 213.
- 湯浅英幸・大久保智生 (2014). 生徒の怠学行動に関する研究(1)—怠学行動尺度の作成— 日本教育心理学会第 56 回総会発表論文集, 205.
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, **70**, 123-133.