モデル適合度の目標適合度

—— 観測変数の数を減らすことの是非を中心に ——

南風原 朝 和

On the Relevance of Goodness of Model Fit to the Purpose of Research

—— Should We Keep the Number of Indicator Variables Small? ——

Tomokazu HAEBARA

Kano (this volume) states that it is possible to obtain an accurate estimate of disattenuated correlation even when the number of items per construct is small. This statement seems to support the common practice in the use of structural equation models (SEM) in which the number of indicator variables per construct are kept very small in order to insure a high degree of model fit. The present paper points out that latent variables with a small number of indicator variables tend to suffer from low generalizability and low validity, and thereby impair the value of the research finding. This issue exemplifies the fact that the relevance of goodness of model fit should be critically examined in reference to the purpose of research. An issue of the correction for attenuation is also examined in some detail in order to understand the nature of the correction implemented in the SEM framework.

1. 議 論 の 焦 点

構造方程式モデル (SEM) による分析の最大の特徴は、観測変数間の共分散等について、データから直接求められる値と、モデルに基づく値 (モデルに基づいてデータから推定されるパラメタ値を、モデルから理論的に導かれる構造式に代入することによって計算される値) との近さによって、モデルの適合度が評価できることであろう。

モデルの適合度というのは、ひとくちで言えば、「つじつまが合っている程度」ということである。「つじつまが合う」というのは、ある閉じた世界の中での整合性を意味する。その閉じた世界の外側からの要請、たとえば測定の妥当性を確保したいという目標に対しては、その達成度を別途評価する視点が必要になってくる。本コメント論文では、SEM の適用において、モデル適合度を研究の目的に照らして相対化してとらえることの必要性を強調したい。

本論文では、「複数の項目による尺度構成」に注目して、この問題を考える。尺度構成にあたっては従来、「実用的な範囲内でできるだけ多くの項目で尺度を構成する」という“教え”があった。これに対し、SEM を適用した研究では「ひとつの構成概念あたりの観測変数の数を少なく抑える」という実践が多くみられる。後者の実践は、そうすることによってモデル適合度が高められる (逆に言えば項目を多くすると、つじつまが合わなくなる) という事実を根拠としている。これで本当に問題がないのか、というのが本論文の具体的な問いである。この問いについて、狩野論文に言及しながら考えてみたい。

SEM を用いる研究においてひとつの構成概念あたりの観測変数の数が少ないことは、SEM ではそれでも相関の希薄化を修正した正しい推定が可能であると主張によって正当化されているように思われる。そこで本論文では、はじめに、相関の希薄化とその修正の問題をとりあげ、従来の知見と SEM における希薄化修正の意味について検討する。

2. 項目の数・尺度の信頼性・相関の希薄化に関する従来の知見

2.1. 項目の数と尺度の信頼性

一般に尺度の信頼性は、尺度を構成する項目数が多いほど高くなる。「尺度を構成する項目を多く」という従来の「教え」は、基本的にこのことが根拠となっている。たとえば、すべての項目が互いに平行測定となる特殊な場合については、有名な Spearman-Brown の公式があり、尺度の信頼性が、各項目の信頼性を項目数に応じて高めたものとなることが知られている。また、項目が平行測定でない場合でも、かなり緩い条件のもとで、項目数に応じて尺度の信頼性が高くなることが知られている。

2.2. 相関の希薄化とその修正

尺度の信頼性を高くするのは、個人差測定の文脈では測定の精度を高めるためであるが、相関研究の文脈では、尺度の信頼性が低いと尺度得点間の相関がその程度に応じて低下（希薄化）してしまうという点が重要である。狩野論文の (3.1) 式は、真の尺度得点間の相関 $Corr(x, y)$ が、尺度の信頼性 ρ_x, ρ_y が低い分だけ希薄化されて、尺度得点間の相関 $Corr(X, Y)$ の値が得られることを示している。その式を変形すると、

$$Corr(x, y) = \frac{Corr(X, Y)}{\sqrt{\rho_x \rho_y}} \quad (1)$$

という式が得られる。これは、尺度得点間の希薄化された相関 $Corr(X, Y)$ から、真の尺度得点間の相関 $Corr(x, y)$ を推定するための式で、「希薄化の修正公式」とよばれている。これによって、原理的には希薄化の問題が解決可能ということになる。

2.3. 希薄化の修正公式の適用上の問題

しかしながら、希薄化の修正公式は、心理学などの研究論文においてルーチン的に適用され、修正後の値が相関係数のより良い推定値として報告されることはなかった。その理由は、「測定誤差」、「真の得点」、そして「信頼性」という、希薄化の修正公式を構成する要素自体の多義性と関係している。つまり、測定誤差とは何か、どのような要因による変動を誤差変動とみなすのか、という基本的なことが測定の文脈に依存するため、観測値から測定誤差を差し引いたものである真の得点もそれに伴って意味内容が異なってきた、「測

定誤差による希薄化を修正して真の得点間の相関を推定する」という作業が一筋縄ではいなくなるのである。

このあと検討する SEM による希薄化修正の意味を、従来の方法と対比させて明確に理解するために、やや長くなるが、いま述べた従来の方法の適用上の問題について少し詳しく考察しておきたい。

測定誤差の要因としてよく挙げられる主なものは、以下の3つのタイプのものである。

- (1) 体調や動機づけのレベルなど、被験者の内的状態の日々の変動
- (2) 回答の際の気の迷いや読み違いなど、瞬間的な揺れ
- (3) 項目プールから尺度項目として選ばれる項目の違い

いま、相関を求める2つの尺度をひとつのセッションで連続して（あるいは2つの尺度に含まれる項目を混ぜて同時に）実施するという典型的な状況を考えてみよう。ここで、尺度に含まれる項目は固定して考えるものとする、上記の1と2のタイプの要因による変動が測定誤差とみなされることになる。

そこで、いまの状況において、2つの尺度の得点 X, Y について以下のようなモデルを考える。

$$X = x + e_x + e'_x, \quad Y = y + e_y + e'_y \quad (2)$$

ここで、 e_x および e_y は、上記のタイプ1の要因による誤差をあらわし、 e'_x および e'_y はタイプ2の要因による誤差をあらわすものとする。 x および y は、観測される尺度得点からこれら2種類の測定誤差を差し引いた真の得点である。なお、各尺度得点を構成する3つの成分は互いに無相関であることを仮定しておく。

このとき、尺度得点間の共分散はそれぞれの得点の成分間の $3 \times 3 = 9$ 個の共分散の和となるが、そのうち、 $Cov(x, e_y), Cov(x, e'_y), Cov(e_x, y), Cov(e_x, e'_y), Cov(e'_x, y), Cov(e'_x, e_y), Cov(e'_x, e'_y)$ の7個の共分散は、母集団においてはゼロと仮定するのが自然である¹⁾。しかし、タイプ1の要因による誤差間の共分散 $Cov(e_x, e_y)$ については、分析対象となる尺度の組合せにもよるが、たとえば、2つの教科の学力テストであれば、その日の動機づけがたまたま高ければ両方のテストの得点が上昇し、たまたま低ければ両方のテスト

¹⁾ 以下の議論は母集団レベルでのものであるが、煩雑さを避けるため「母集団において」という言葉は繰り返さないこととする。

の得点が低下するだろうから、一般的には $Cov(e_x, e_y) \neq 0$ と考えるべきだろう。したがって、2つの尺度得点間の共分散は、

$$Cov(X, Y) = Cov(x, y) + Cov(e_x, e_y) \quad (3)$$

とあらわされる。このとき、尺度得点間の相関 $Corr(X, Y)$ は、尺度得点の分散 $V(X)$, $V(Y)$ を用いて、

$$Corr(X, Y) = \frac{Cov(x, y) + Cov(e_x, e_y)}{\sqrt{V(X) V(Y)}} \quad (4)$$

によって与えられる。この式に基づいて、希薄化を修正して真の得点間の相関 $Corr(x, y)$ を推定することを考えてみる。

上記のタイプ1およびタイプ2の要因による変動を測定誤差とみなす場合、信頼性

$$\rho_x = \frac{V(x)}{V(X)}, \quad \rho_y = \frac{V(y)}{V(Y)} \quad (5)$$

を推定するには、それぞれの尺度を日をおいて2度ずつ実施し、その2度の間の相関をとる再テスト法が適している。仮に、再テスト法によって2つの尺度の信頼性が正確に推定できたものとし、これを(4)式に代入すると、

$$Corr(X, Y) = \frac{Cov(x, y) + Cov(e_x, e_y)}{\sqrt{V(x) V(y)}} \sqrt{\rho_x \rho_y} \quad (6)$$

となる。この式は、狩野論文の(3.1)式と類似しているが、分子にタイプ1の要因による測定誤差間の共分散が残っている点が異なる。先に述べた説明から、この誤差間の共分散は真の得点間の共分散 $Cov(x, y)$ と同じ符号をもつと考え、(6)式より、

$$|Corr(X, Y)| \geq |Corr(x, y)| \sqrt{\rho_x \rho_y} \quad (7)$$

という関係が導け、これからさらに、

$$|Corr(x, y)| \leq \frac{|Corr(X, Y)|}{\sqrt{\rho_x \rho_y}} \quad (8)$$

という関係が導ける。この式は、いま想定している状況において再テスト信頼性(の正確な推定値)を希薄化の修正公式に代入した場合、真の得点間の相関を過大に推定してしまうことを意味している。言い換えれば、再テスト信頼性は、希薄化の修正公式に適用すべき信頼性を過小評価しているということである。

このことは以下のように解釈することができる。たとえば、ある被験者がたまたま動機づけが低く、その

ために2つの尺度の得点とその被験者の真の得点より低いほうにぶれたとしよう²⁾。このような変動は、その被験者の測定という観点からは確かに測定誤差とみなすべきものであるが、集団における相関関係を推定するという観点からは、必ずしもそうではない。この後者の観点からは、その被験者のその日の動機づけが低かったという事実を、その被験者よりも(動機づけが低い分だけ)真の得点が低い被験者が代わりにサンプリングされたとみなすことができるからである³⁾。

このような見方からは、真の得点を新たに、タイプ1の要因による変動を含めて

$$x' = x + e_x, \quad y' = y + e_y \quad (9)$$

と定義し、タイプ2の要因による変動のみを測定誤差とみなすモデルが導かれる。このとき、この新たな真の得点間の共分散は、(3)式を導いたときと同じ仮定のもとで、

$$\begin{aligned} Cov(x', y') &= Cov(x, y) + Cov(e_x, e_y) \\ &= Cov(X, Y) \end{aligned} \quad (10)$$

となり、尺度得点間の共分散に等しくなる((3)式参照)。また、この場合の尺度得点の信頼性は、

$$\rho'_x = \frac{V(x')}{V(X)}, \quad \rho'_y = \frac{V(y')}{V(Y)} \quad (11)$$

と定義できるから、これらを(4)式に代入すると、

$$Corr(X, Y) = \frac{Cov(x', y')}{\sqrt{V(x') V(y')}} \sqrt{\rho'_x \rho'_y} \quad (12)$$

となる。そして、この式から、再定義された信頼性を用いて希薄化を修正し、再定義された真の得点間の相関を推定する式

$$Corr(x', y') = \frac{Cov(X, Y)}{\sqrt{\rho'_x \rho'_y}} \quad (13)$$

が導かれる。

しかし、ここでもまだ問題が残る。上記のタイプ2の要因による変動のように、その瞬間瞬間に生じる測定誤差のみを対象とする信頼性 ρ'_x , ρ'_y はしばしば“co-

²⁾ 議論を簡単にするために、ここでは学力テストのような尺度を想定しておく。

³⁾ 集団を対象とした測定は、第1段階で被験者ないしは真の得点をサンプリングし、第2段階で観測得点をサンプリングする2段階サンプリングとみなすことができる。何を真の得点とし何を測定誤差とするかという問題は、どこまでを第1段階のサンプリングとみなすかという問題であると言える。

efficient of precision” とよばれるが (Lord & Novick, 1968), この種の信頼性を適切に評価するための方法がないのである。α 係数は、古典的テスト理論の文脈ではこの信頼性の下限値を与えるものとして解釈されているが、尺度に含まれる項目の性質によっては、かなり低い下限値を与えることがあり、coefficient of precision の推定値としてルーチンの利用することはできない。

以上のように、希薄化の修正公式の適用については、理論的および現実的な問題が残されており、それがこの公式のルーチ的な利用を妨げている。これに対し、SEM による希薄化の修正は分析法そのものに組み込まれ、その意味でルーチ的に利用されている。SEM では、従来の方法における上記の問題は解決されているのだろうか。もし解決されているとしたら、どのように解決されているのだろうか。次にこのことについて考察する。

3. SEM における希薄化の修正

関連の希薄化の問題に対する SEM における対処について、狩野は、「十分多くの尺度項目を構成できず信頼性が低くとも、CFA でモデル化を行えば、誤差をコントロールして希薄化されない正しい推定値を求めることができる」と述べている (3.1 節)⁴⁾。つまり、SEM を用いれば、(1) 式のような修正式を用いなくても、いわば自動的に修正がなされるということである。

このことは SEM による分析の特長としてしばしば挙げられる。たとえば、狩野自身によるテキスト (狩野・三浦, 2002) では、「検証的因子分析モデルは、希薄化修正モデルとよばれる」ことがあること (p. 92)、そして、SEM の利用によって「考慮中の因果関係以外の変動を誤差として捨てることにより、希薄化されない正しいパス係数の推定値が得られ」ることが、述べられている (p. 278)。

SEM においては、一組の観測変数 $X_i (i=1, \dots, p_x)$ は、それらの変数が測定する潜在変数 (共通因子) f によって、

$$X_i = \mu_i + \lambda_i f + u_i \quad (14)$$

とあらわされる (狩野論文の (2.2) 式)。同様に、別の

組の観測変数 $Y_i (i=1, \dots, p_y)$ が潜在変数 g を測定しているとする。このとき、2つの潜在変数 f と g とがパスで結ばれたモデルを設定すると、そのパス係数 (双方向のパスであれば相関係数) の推定値は、誤差 u_i によって希薄化されないというのが上記の主張である。

前節における従来の枠組みでの議論をふまえて、(14) 式における誤差 u_i がどのような要因による誤差変数なのかを考えてみる。その際に重要なことは、SEM においては通常、誤差変数 u_i が互いにすべて無相関であることがモデル上仮定されることである。

先に挙げた3つの主要な誤差要因のうち、タイプ1の体調や動機づけレベルなどの日々の変動は、同一の構成概念を測るための観測変数に対しては一定の方向で作用すると考えられる。したがって、この要因による変動は互いに相関することになるため、誤差変数 u_i はこの変動を含まないものであることがわかる。ということは、こうした日々の変動は、潜在変数 f に吸収されるということである。

これに対し、タイプ2の瞬間的な揺れは、同一の構成概念を測るための観測変数に対しても独立に作用すると考えられるから、この要因による変動は誤差変数 u_i に含まれることになる。

一方、狩野論文でも言及されているように、(14) 式における誤差変数 u_i には、観測変数 X_i 独自の測定内容による成分である特殊因子 s_i が含まれている。潜在変数 (共通因子) f の値が同じ被験者でも、その観測変数独自の測定部分に関する成分 s_i が異なれば、観測変数の値が系統的に異なるということである。この特殊因子は、構成概念を測定するのにどの変数が選ばれるかという一種の偶然的要因によって生じる変動、すなわち先に挙げたタイプ3の要因によって変動する成分とみなすことができる。

したがって、SEM における希薄化の修正は、どのような具体的内容をもった観測変数が選ばれるかという要因 (タイプ3) と、瞬間的な変動要因 (タイプ2) による誤差変動を除去するものであると言える⁵⁾。従来の枠組みでは、タイプ2の要因による誤差変動の除去が、それに相応しい信頼性推定値が得られないという理由で困難であったが、SEM ではこの問題をそのままの形で解決するのではなく、タイプ3の要因を新た

⁴⁾ この中で“CFA”は検証的 (確認的) 因子分析のことであるが、ここでは SEM と読み替えても差し支えない。

⁵⁾ 被験者の内的状態の日々の変動 (タイプ1) による得点変動が誤差ではなく、潜在変数 f に含まれている点はユーザによって必ずしも正しく認識されていないかもしれない。

に誤差要因に加え、それとタイプ2の要因による誤差とを合わせて除去するという、問題設定自体を変更した上での解決を与えていると言えよう。

このように、SEMにおける希薄化修正の特徴は、観測変数の選択による変動が誤差として扱われるという点であり、このことが、「観測変数の数を少なく抑えることの是非」の問題を考える上で重要な意味をもってくる。

4. 観測変数の数を少なく抑えることの問題点

4.1. 一般化可能性の問題

一般的に言えば、観測変数の選択による変動を測定誤差とみなすのは望ましいことである。心理学などの研究においては、特定の観測変数群それ自体に興味があるというよりは、それらの背後に仮定される潜在変数（の間の関係）に興味があるから、その潜在変数とは直接関係しない、個々の観測変数に特殊な内容 (McDonald (1999) の言葉では “idiosyncratic property”) の関与はできるだけ抑えるほうが良いからである。

しかし、個々の観測変数に特殊な内容というのは、同じ構成概念を測るために他にどのような観測変数が用意されているかに依存する相対的なものである。したがって、その影響を除いた共通因子としての潜在変数の内容も、具体的に用意した観測変数群に依存して変化することになる。

狩野論文では、このことに関連して、「CFAの大きな特徴は、モデルが適合しているならば、変数の削除によって分析結果が大きな変化を受けないということである」と述べている(3.1節)。ここで、「モデルが適合している」ということの意味は、ある観測変数群の間の相関関係が1つの共通因子で説明でき、その共通因子の影響を除いたときの偏相関がすべてゼロになるということである。確かに、モデルが完全に適合していれば、その観測変数群から少数個の変数だけを選んでも、共通因子の内容自体は変化せず、したがって分析結果もほとんど変化しないだろう。

しかし、心理学などの研究で用いられる尺度の多くは、明確な2因子以上の構造はもたないまでも、ただ1つの共通因子のみを共有し、あとはどの項目も互いにまったくバラバラの内容という1因子構造を示すとは考えにくい。実際、項目作成の段階から、当該の構成概念を反映するようないくつかの状況を設定した

り、行動のカテゴリをリストアップしたりした上で、それぞれの状況や行動カテゴリごとに複数の項目を作成して尺度とすることが多い。その場合は、当然、それらの状況や行動カテゴリごとにある種のまとまりが生じることが予想される。それでもどの項目も第1因子への負荷が高く、第1因子への寄与が十分大きければ、便宜的に1因子構造をもつものとみなしているのがほとんどであろう。このような尺度項目から少数個の項目を観測変数として選択するとしたら、どの状況あるいは行動カテゴリからどの項目が選ばれるかによって、潜在変数の内容は変化してくると考えざるを得ない。したがって、たまたま選ばれた項目群を用いて得られる結果から、測定したい内容を広くカバーした項目ユニバースを用いたときに得られるであろう結果への一般化可能性が低いという問題をかかえることになる。

4.2. 妥当性の問題

いま述べたように、従来は1因子モデルの適合度ではなく、第1因子の寄与が大きいことを重視した尺度構成がなされてきた。この2つの指針はかなり異質のものであり、1因子モデルが完全に適合しても寄与は小さいこともあり、逆に1因子モデルの適合が悪くても寄与は十分大きいこともある(狩野・三浦, 2002; 南風原, 2003)。寄与を重視して尺度構成をした場合、1因子モデルの当てはまりが悪くて、上記のようにその中のどの項目が観測変数として選ばれるかによって潜在変数の内容が変化するという問題が生じるが、逆に1因子モデルの適合度を追求した場合、結果的に特定の状況や行動カテゴリに含まれる項目だけが残る可能性がある。この場合は、測定すべき内容領域に対して狭く偏った内容の潜在変数となり、その分、妥当性が低くなるという問題が生じる。

なお、狩野論文では、「尺度において最も重要な性質は(信頼性ではなく)内容的妥当性である」としたうえで、「内容的妥当性を強調すればするほど信頼性を高めることが難しくなり、この事実は、CFAの重要性をさらに強調することになる」と述べている(3.1節)。これもモデルの適合を前提とした議論である。つまり、「構成概念の意味する様々な側面をもらさずに測定」するようにしてもなお、1因子モデルが適合するという理想的なケースが想定されている。

5. まとめと討論

5.1. 観測変数の数はどうするか

観測変数の数を少なく抑えることによって、結果の一般化可能性が制限され、測定の妥当性も損なわれるとしたら、従来の尺度構成と同様に、良質の項目を十分多く用いるべきだという結論になる。おそらくはそれによってモデルの適合度は低下する（つじつまが合わない部分が出てくる）だろうが、一般化可能性および妥当性の向上に伴って低下するようなモデル適合度であれば、それは低下しても構わないのではないか。

なお、この疑問に対して、「多くの観測変数をバラバラに用いるのではなく、少数個の“下位尺度”にまとめて新たに観測変数とすれば、全体として観測変数の数を少なく（自由度を小さく）することができ、モデル適合度が低下しないようにすることができる」という“回答”を耳にしたことがある。こうすることによって、結果として多くの項目によって得られる情報が利用できるのは良いことだが、下位尺度へのまとめ方によって結果が異なってくるという任意性が残る。また、こういう手段を用いて適合度を上げることによってどのような本質的意味があるのかという疑問が残る。

5.2. モデル適合度の目標適合度

いま述べたことは結局のところ、研究においては、モデル適合度を犠牲にしても追求すべきこと（結果の一般化可能性および測定の妥当性など）があるのではないか、モデル適合度を、研究の目的に照らして相対化してとらえる必要があるのではないかということである。言い換えれば、モデル適合度を高めることが、研究の目標・目的にどれだけ適合したことなのか、どういう意味で重要なのか、という冷静な検討が必要ではないかということである。

従来の心理学などの研究において、検定で有意差を出すことを追求しすぎて、とるに足りないような相関や効果量でも、有意であるというだけで過大解釈してしまうという問題が指摘されてきたが、同じような問題がSEMでも繰り返されかねないように思う。有意性検定における“5%水準”のように、モデルが適合しているか否かについて、“ $GFI \geq .90$ （あるいは $.95$ ）”とか“ $RMSEA \leq .05$ ”といった基準が提案されそれが共有されつつある。そしてそれが、「有意か否か」、「適合か否か」といった2値判断を求めがちな研究者にとって非常に便利で魅力的なものであるがゆえに、モ

デルが適合する（あるいはしない）ということの意味を冷静に吟味することなく、結果的に適合度に振り回されてしまいかねない危険を感じるのである。

5.3. 測定部分の適合度と構造部分の適合度の分離

狩野論文では、「SEMの大きな特徴は、構成概念を規定する測定モデルと研究対象である構造モデルの分離にある」と述べている（3.2節）。これはその通りであるが、モデルの適合度については、測定部分と構造部分を分離することなく、総合的・全体的に評価される。したがって、先に述べたように、結果の一般化可能性と測定の妥当性のために、測定部分の適合を多少犠牲にするというを仮に実行した場合、それによって全体としての適合度が低下してしまうことは避けられない。このとき、仮に構造部分の適合に問題がないとしても、モデル適合度だけからそのことを正しく把握するのは困難である。

測定部分と構造部分の適合度を分離して評価するには、ひとまず構造モデルとしてすべての潜在変数間にパスを引いた飽和モデルを想定して適合度を求め、次に、検討したい構造モデルを導入してもう一度適合度を求める、といった方策が考えられる。第一の適合度によって測定部分のみの評価を行い、第二の適合度がそれよりどれだけ低下するかによって構造部分の評価を行うという発想である（Anderson & Gerbing, 1988）。

5.4. 一般化可能性および妥当性の検証

本論文では、ひとつの構成概念あたりの観測変数の数が少ない場合、結果の一般化可能性と測定の妥当性が損なわれるのではないかという懸念を表明した。個々の研究状況において、この懸念が現実のものとなっているのか、それとも杞憂にすぎないのかは、実際に確認することが可能である。

まず、結果の一般化可能性については、利用可能な項目群（すでにデータが得られているもの）から観測変数として少数の項目を選ぶ場合、その選び方によってパス係数などの推定値がどの程度変動するかを実際に調べてみればよい。どの項目セットを用いても、また観測変数の数を大幅に増やしても結果が変動しないのであれば安心である。もし結果が大きく変動するようであれば、やはり、良質の項目を十分多く用いるべきであろう。

次に、観測変数群によって測定される潜在変数につ

いて、外的変数を導入して直接的に妥当性検証をすることができる(南風原, 2002, 10.5 節)。この場合は、妥当性検証のために、当該の潜在変数と外的変数とをパスで結んだモデルを SEM で分析することになる。もしいろいろな外的変数との間に、妥当性の収束的証拠および弁別的証拠が十分に得られたら安心である。ただし、この場合、外的変数として何をもってくるかによって、パラメタ推定における最適化の結果として、当該の潜在変数から各観測変数へのパス係数(因子負荷)が多少とも変化することが考えられる。極端に言えば、外的変数を変えて妥当性を検証しようとするたびに、その対象となる潜在変数の内容が変化するということである。これは、SEM のおそらくあまり好ましくない性質であろう。

参 考 文 献

- Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, **103**, 411-423.
- 南風原朝和 (2002). 心理統計学の基礎——統合的理解のために. 有斐閣.
- 南風原朝和 (2003). 統計法の適用と結果の解釈をめぐって. 日本教育心理学会(編) 教育心理学ハンドブック. 有斐閣, pp. 170-181.
- 狩野 裕 (2002). 構造方程式モデリングは、因子分析、分散分析、パス解析のすべてにとって代わるのか? 行動計量学, **29**, 138-159.
- 狩野 裕・三浦麻子 (2002). AMOS, EQS, CALIS によるグラフィカル多変量解析——目で見る共分散構造分析 増補版. 現代数学社.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

(2002 年 9 月 19 日受付, 2003 年 1 月 31 日最終修正)