

構造方程式モデリングによる 一対比較型尺度構成法の開発

-妥当性への脅威とその統計的対処-

川端一光

目 次

第Ⅰ部 問題・目的	4
第1章 テスト妥当性の脅威・本稿の構成	5
第2章 社会的望ましさによる反応歪曲と従来の対処法	7
2.1 社会的望ましさ尺度による対処	7
2.2 リックート法を利用しない対処	7
2.3 統計的対処法	8
2.4 本稿の目的	8
第3章 モデルの制約による尺度得点の歪曲とその対処	10
3.1 非線形ロジットIRTモデル	10
3.2 結果・考察	11
第4章 不適切な標本抽出法による尺度得点の歪曲とその対処	12
4.1 傾向スコア重み付け法	12
4.2 ニューラルネットワークの援用	13
4.3 結果・結論	14
第Ⅱ部 構造方程式モデリングによる一対比較型尺度構成法の開発	15
第5章 構造方程式モデリングによる一対比較法の表現	16
5.1 SEMによる被験者内計画の為の一対比較法	16
第6章 不完全一対比較データへのSEMによる対処	18
6.1 不完全データの多母集団解析・交互作用の制約	18
6.2 結果・考察	19
第7章 一対比較型尺度構成法の基礎モデル	20
7.1 測定するEQの領域・尺度	20
7.2 一対比較型テストモデル	21

7.3 標準化・妥当性研究の結果	22
第8章 社会的望ましさの尊重度を考慮したテストモデル	23
8.1 テストモデル	23
8.2 コンピテンシー尺度	24
8.3 標準化・妥当性研究の結果	25
第9章 社会的望ましさの個人差を考慮したテストモデル	26
9.1 テストモデル	26
9.2 標準化と妥当性研究	28
引用文献	29

第I部

問題・目的

第1章 テスト妥当性の脅威・本稿の構成

知能や学力といった各種心理特性を測定するテストの作成において、第一に留意されるべき概念とはテストの信頼性と妥当性である。妥当性はテストが測定対象である心理特性を適切に捉える程度であり、信頼性(テスト得点に対する測定誤差の影響の程度)に先立って確認される必要がある(南風原, 2002)。構成概念を扱うことの多い心理学研究において、テストの妥当性はその根幹を成す極めて重要な概念である。

テストに含まれる項目は、通常、当該領域の専門家によって選定された項目から構成されるため、内容的妥当性は保証される。しかし妥当性への脅威は多く、専門家による質的な妥当性の確認だけでは不十分である場合が多い。

テストの妥当性に対する脅威としては、1)社会的望ましさ、2)テストモデルの制約、3)不適切な標本抽出法の3点を挙げることができる。テストの妥当性の確保は、心理研究を根幹を成す極めて重要な問題であり、分析者は上述した3点に関して十分配慮する必要がある。

本稿では、特に、古くからその対処が試みられて来た、1)社会的望ましさによるテストの妥当性の損失に関して、これを大幅に抑止する心理評定モデル並びに、尺度構成の手続きを開発した。また、2)テストモデルの制約による妥当性の損失、3)不適切な標本抽出による妥当性の損失についても、統計的な対処によって、事後的に妥当性を向上させる方法を開発した。

本稿の構成を次に示す。

第Ⅰ部：第Ⅰ部では本稿の問題と目的が論じられる。特に社会的望ましさによる被験者データの反応歪曲に関する、先行研究における統計対処法の成果と課題を概観した。第Ⅰ部3章では、テストモデルの制約によってテストの妥当性が損失する問題に関して、非線形ロジットIRTモデルによる対処を示した。また第Ⅰ部4章では、標本抽出法として有意抽出を利用した場合の、反応歪曲、そして妥当性の損失に抗する為、ニューラルネットワークを併用した傾向スコア重み付け法を開発し、その有効性を示した。

第Ⅱ部：第Ⅱ部では本稿の主旨である、社会的望ましさによる反応歪曲に耐性を持つ、一対比較型尺度構成法について論じている。本手法は構造方程式

モデリング (Structural Equation Modeling, SEM) によって表現されるテストモデルであり、その有効性が反応歪曲に脆弱とされるリッカート法との比較によって検討された。

第Ⅱ部5章では本稿の数理的基礎を与える、SEMによる一対比較法について論じられている。また第Ⅱ部6章では、項目数が爆発的に増加する一対比較測定の問題に対して、比較回数が少ないにも関わらず妥当な解を得るための統計的手法を、SEMによる多母集団同時分析によって表現する方法を示した。

本稿で開発された一対比較型尺度構成法は3モデルに大別される。第Ⅱ部7章では他モデルの基礎を与える、一対比較型テストモデルが論じられる。第Ⅱ部8章では基礎モデルを発展させた、社会的望ましさへの個人毎の尊重度を表現するテストモデルが論じられる。最後に第Ⅱ部9章では社会的望ましさへの個人差からの影響を評価・排除する一対比較型テストモデルについて論じられる。

第Ⅲ部：第Ⅲ部では本稿の全体的考察と限界が述べられる。

第2章 社会的望ましさによる反応歪曲と従来の対処法

現在までに行われてきた社会的望ましさへの対処法は以下の3手法に大別することができる。

- 1) 社会的望ましさ尺度・虚偽尺度を用意し、性格検査の妥当性を推測する方法
- 2) 反応歪曲が容易な評定尺度であるリッカート法以外の心理測定尺度を利用する方法
- 3) 社会的望ましさ尺度を利用して、性格検査に含まれる反応歪曲を統計的に修正する方法

2.1 社会的望ましさ尺度による対処

社会的望ましさ尺度とは、性格検査に含まれる反応歪曲を検出することを目的として作成される尺度の総称であり、Edwards (1957) で考案された39項目からなる社会的望ましさ尺度をその嚆矢として、現在までに多数作成されている。

Edwards(1957) 以降では、例えば Crowne & Marlowe (1960) が、33項目からなる Marlowe-Crowne 社会的望ましさ尺度を、Wiggins (1959) が 40項目からなる社会的望ましさ尺度を作成している。

2.2 リッカート法を利用しない対処

性格検査に含まれる項目は自己記述型の評定尺度であることが多く、具体的にはリッカート法が採用されている。そしてリッカート法を含めた評定尺度では、意図的な反応歪曲が容易であることが報告されている (Alliger & Dwight, 2000; Caldwell-Andrews, Baer & Berry, 2000; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Visweswaran & Ones, 1999)。

また Kanning & Kuhne (2006) は、社会的望ましさによる反応歪曲に対して耐性の高い心理測定法の検討の中で、リッカート法で代表される評定尺度が、他の心理測定法に比較して反応歪曲に脆弱であることを報告している。

Edwards (1959) で考案された一対比較型の典型値検査である EPPS では、回答への社会的望ましさの影響を排除する為に、同程度の社会的望ましさをもった質問文の比較判断を被験者に求めることで、社会的望ましさの影響を統制している。一対比較法を利用すること自体に、意図的な反応歪曲を抑制する効果が期待されるが、それと同時に社会的望ましさの程度が等しい比較試料を対にして、実験計画的に反応歪曲に対処している。

多次元の性格特性を測定する一対比較検査 (multidimensional forced-choice, MFC) も考案されており (例えば Heggestad, Morrison, Reeve & Christiansen, 2006; Christiansen, Edelstein & Fleming, 1998; Jackson, Wroblewski & Ashton, 2000), やはりリッカート型の評定尺度よりも反応歪曲に耐性を持つことが報告されている。ただしこれらの方法は統計モデルとして精緻に表現されているわけではないため、その一般性については十分に考察ができないという欠点がある。

2.3 統計的対処法

登張 (2007) によれば社会的望ましさに対する従来の統計的対処法は、1) 既存の社会的望ましさ尺度を利用し、尺度得点における反応歪曲を修正する方法と、2) 興味のある尺度に対する外的基準との相関係数 (基準関連妥当性) を、社会的望ましさ尺度を抑制変数として利用することで向上させる方法の 2 つに分類される。

しかし社会的望ましさに対する上記 2 つのアプローチは、社会的望ましさによる反応歪曲の修正という意味で十分有効な方法であるとは言い難い (登張, 2007)。例えば社会的望ましさ尺度からの回帰係数を利用した修正法では、平均的な補整効果はみられるものの、個人レベルで十分な補整が確認されるわけではない。また抑制変数として社会的望ましさ尺度を導入し、尺度の基準関連妥当性を向上させるという方法に関しては、抑制変数を導入しても基準関連妥当性は殆ど変わらないか、むしろ低くなることが知られている。

2.4 本稿の目的

社会的望ましさに対する従来の対処法についてその成果を要約すると次のようにになる。

- 1) リッカート法は社会的望ましさによる反応歪曲に対して脆弱であるのに対して、一対比較法は比較的頑健である。ただし統計モデルとしての表現が難しく、理論的に整備されていない
- 2) 統計的修正法を利用しても、尺度得点 (合計得点) レベルでの反応歪曲の補正効果は少ない

上述の問題を受け、本稿では構造方程式モデリングによる一対比較法を利用した統計的尺度構成法を提案した。本手法と従来の統計的対処法との大きな相違点は、評定尺度としてその頑健性が確認されている(Edwards, 1959, 1970; Kanning & Kuhne, 2006)一対比較法を利用する点である。

前節で論じた従来の統計的対処法の本質は、興味の対象となる尺度の他に、社会的望ましさ尺度を用意することで事後的に反応歪曲を統制するというものである。そして興味の対象となる尺度が基本的にリッカート項目で構成されているという前提がある。この点に対して、本稿では興味の対象となる尺度を構成する項目の心理評定モデルを、意図的な反応歪曲が難しい一対比較モデルに変更し、被験者の反応に社会的望ましさの影響が極力含まれないようにする。またSEMを利用することによって、反応に混入した社会的望ましさの影響も同時に評価し、尺度得点から分離する方法を提示する。

一対比較法を利用する点では本稿で開発するテストモデルはEdwardsのEPPSと類似している。しかしEPPSが社会的望ましさの実験計画的相殺のために一対比較法を利用するのに対して、本稿ではそのような相殺を意図していない。EPPSのように同程度の社会的望ましさを持つ刺激文を作成することは必ずしも容易ではない。事前調査において作成された刺激文に関する社会的望ましさを計量した上で、完全に社会的望ましさが同じ刺激文の対を探すことは不可能であり、EPPSは社会的望ましさの程度に関して、粗い近似において刺激対を作成しているに過ぎない。

この点に関して後に詳述するように、本稿における方法ではEPPSのように事前に社会的望ましさの程度をそろえた項目対を作成する必要はない。また構造方程式モデリングの枠組みで表現された一対比較型テストモデルを利用するため、社会的望ましさによる妥当性の低下が修正された尺度得点の算出だけでなく、社会的望ましさの影響の高い項目や、社会的望ましさへの個人毎の尊重度、社会的望ましさの個人差からの影響をそれぞれ定量的に評価できる。

第3章 モデルの制約による尺度得点の歪曲とその対処

本章では現代テスト理論における尺度構成法の1つ、項目反応理論 (Item Response Theory, IRT) で議論されることの多い、名義反応モデルについて、テストモデルの妥当性の向上を試みた。

3.1 非線形ロジットIRTモデル

IRTを用いた尺度構成においては、被験者の反応パターンの確率モデルを構築し、その母数を個人の尺度得点として定義している。通常被験者の反応パターンの確率モデルは、統計的処理が扱いやすい性質のよく知られた確率分布であることが多く、あくまで心的事象の便宜的な近似である。

例えばIRTにおける名義反応モデルでは、能力 θ_j を持つ被験者 j が、ある項目のカテゴリ k に回答する確率を次の式で表現する。

$$p_k(\theta_j) = \frac{\exp[\zeta_k + \lambda_k \theta_j]}{\sum_s^K \exp[\zeta_k + \lambda_k \theta_j]} \quad (3.1)$$

ここで指数内の $\zeta_k + \lambda_k \theta_j$ はカテゴリ k のロジット(計量経済学では効用)と呼ばれ、各個人ごとに定義されるカテゴリ k への選好度と解釈される。

ロジットは式からも明らかのように、単純な線形和で表現されているが、この仮定はデータの実態に対して制約的かもしれない。仮に制約的であるならば、本式で定義される能力母数 θ の妥当性は損なわれていることになる。そこで本研究では、データの実態に関してより柔軟に適合するロジットを選択し、能力母数 θ の妥当性の向上を試みた。具体的には、例えば

$$p_k(\theta_j) = \frac{\exp[\zeta_k + \lambda_k \theta_j + \delta_k \theta_j^2]}{\sum_s^K \exp[\zeta_k + \lambda_k \theta_j + \delta_k \theta_j^2]} \quad (3.2)$$

のようにロジットを θ に関する多項回帰式として表現した。

本章ではMayer, Salovey & Caruso (2004)によるEI(Emotional Intelligence)に関するEI尺度構成の試みの一貫として、非線形ロジットIRTモデルの適用を行

なった。また尺度選択の試みについても論じられた。EI 尺度とはフードサービス業界において、接客担当者に求められる接客マナー・基本ビジネス技能検査を測定する単一次元の尺度である。

3.2 結果・考察

信頼性が確認された 14 項目に関して、93 名のデータに対して非線形ロジット IRT モデルを適用した。

算出されたモデルの推定値を元に、項目特性曲線を描画し項目分析を行なったところ、ロジットを非線形多項回帰式で表現するモデルにおいて、比較基準に対する妥当性の向上が強く示唆された。ロジットに多項回帰式のような非線形関数を導入する可能性が開けたといえる。

また非線形ロジット IRT モデルで定義された能力母数 θ は、通常の名義反応モデルに対して高い収束的妥当性を持つ一方で、特に高特性者に対して高い識別力をを持つことが示唆された。非線形ロジット IRT モデルを利用することによって、尺度の妥当性の向上と、目的に合わせた尺度選択がある程度可能になると結論付けられた。

第4章 不適切な標本抽出法による尺度得点の歪曲とその対処

尺度の標準化の際に利用されるデータは、尺度が対象とする母集団から適切に抽出されている必要がある。標本抽出が不適切であった場合には、そのデータによって標準化された尺度は、対象とする母集団に対して妥当性を欠くことになる。標準化データの収集には、母集団の性質を尺度が満遍なく表現することを意図して、広く無作為抽出すべきである。母集団の規模が大きければ、多段抽出法といった階層的標本抽出法を利用する必要性も生じるだろう。従って偏った標本抽出を行う可能性の高い有意抽出法を利用することは、構成された尺度の妥当性の観点から大いに問題がある。

本章では、不適切な標本抽出法による尺度の妥当性低下に対処するため、有意抽出調査であるWeb調査の結果を、無作為抽出調査である郵送調査の結果に近似する傾向スコア Rosenbaum & Rubin (1983) による検討を述べる。またこの研究は傾向スコアの算出にあたり、よりデータに対して柔軟に適合するニューラルネットワークを利用し尺度の妥当性を高める試みである。

4.1 傾向スコア重み付け法

実験法を利用した因果仮説の検討には、外的妥当性を向上させるため、標本はなるべく多くかつ偏り無く母集団から抽出され、各実験条件に無作為に割り当てられる必要がある。しかし実際には、不本意にも有意抽出法を利用するしかない状況で、実験条件の割り当てが必ずしも無作為でないような場合もある。

ここで、無作為割り当て群を0、有意割り当て群を1とする確率変数 x を定義する。自ら進んで実験群に参加する人には、共通のプロファイル(共変量) z があると考えるのが自然なので、 x と z には何らかの統計的従属性が仮定される。ここで次のような統計量を定義する。

$$\epsilon(z) = p(x = 1|z) \quad (4.1)$$

すなわち共変量が与えられた下での、実験条件の割り当て確率である。この条件付き確率を傾向スコアと呼ぶ(星野, 2003; 星野・繁樹, 2004; 狩野, 2002; 佐藤, 2002)。

傾向スコアは、実験条件への割り当てを2値のカテゴリカルな基準変数としたロジスティック回帰分析における、共変量による条件付き予測確率と表現されることもある。

傾向スコアを利用してWeb調査などの有為抽出調査の結果を無作為調査の結果に近似させる方法が、例えば(鈴木・星野, 2004)によって提案されている。

今、有意抽出法の結果を無作為抽出法の結果へ近似させることを考える。このとき y は調査の測定対象となる変数、 z はフェイス項目などの共変量を表すとする。そして無作為調査による標本数・測定値・共変量・傾向スコアを、それぞれ N_0 , y_0 , z_0 , e_0 とし、Web調査によるそれらを、それぞれ N_1 , y_1 , z_1 , e_1 とする。

このときWeb調査と無作為調査を合わせた母集団における、Web調査のデータから計算される補正された y の平均推定量には、例えば

$$\bar{y} = \frac{1}{N_*} \sum_{i=1}^{N_1} \frac{y_{1i}}{e_{1i}}, \quad N_* = \sum_{i=1}^{N_1} \frac{1}{e_{1i}} \quad (4.2)$$

が利用される。ただし i は被験者の添え字である。これは傾向スコア重み付け法(Propensity score weighting)におけるHorvitz-Thompson型推定量(Horvitz & Thompson, 1952; 星野・繁樹, 2004)の一種である。

以上から、有意抽出法における結果を無作為抽出法における結果に近似するためには、有意抽出された標本の共変量と、基準となる無作為抽出標本におけるデータが手元にあればよいことがわかる。

4.2 ニューラルネットワークの援用

(4.1)式で定義される傾向スコアは、ロジスティック回帰分析によって近似されることが多い。しかしロジスティック回帰分析モデルにおけるロジットは前章で議論した名義反応モデル同様、共変量の単純な(1次の)線形結合である。共変量と割り付け変数 x 間の関数関係を表現するには簡潔すぎる。

この点に関して、事前に特定の統計モデルを仮定しないという特徴をもつニューラルネットワークのデータに対する適合は、通常の統計モデルよりも良くなることが多い(豊田, 1996)。2値型の従属変数を予測するロジスティック回帰モデルは、反応確率モデルとしてロジスティック関数を仮定しているが、前章でも議論したようにこのような仮定はデータの実態に関して制約的に作用している可能性も十分ある。しかしながらニューラルネットワークモデルは、事前にモデルを仮定しないから、データの過学習(過適応)に配慮すれば、ロジスティック回帰モデルにおける結果よりも傾向スコアの予測精度が向上する可能性が極めて高く、かつ有意抽出データの補正も向上することが推測される。

4.3 結果・結論

本研究では、傾向スコア推定法の代替案を提案するにあたって、補正精度の向上を確認する為にWeb、郵送の両調査の下で得られた大規模なデータを必要とする。そこで経年の分析において高い信頼性が確認されており、かつ大標本が確保されているブランドジャパン(日経BPコンサルティング, 2002)のデータ(郵送調査・Web調査)を分析に利用した。

ニューラルネットワークによって傾向スコアを算出し、ブランドジャパンのWeb調査データを、無作為抽出である郵送調査の結果に近似するよう補正した。また補正効果がどれだけ向上するかについて、従来用いられることが多かったロジスティック回帰分析での結果を基準として考察した。

その結果、3層で中間層のユニット数が20のニューラルネットワークモデルの補正効果が認められ、ロジスティック回帰分析の代替案として有効である可能性が示された。また3層以上の複雑なモデルでは予測が不安定になる傾向がみられており、ロジスティック回帰分析の代替案として用いる際に、この点に留意する必要性が示された。

また秘匿にされることが多い共変量を公開し、かつ決定木によって被験者の調査形態への割り当てに対する影響を検討した。被験者の割り当てに強い主効果を持っているのはWeb利用頻度であり、更に結婚の有無・性別・役職という3つの共変量との交互作用が被験者の割り当てに関して影響を持っていることが明らかになった。

第II部

構造方程式モデリングによる一対比較
型尺度構成法の開発

第5章 構造方程式モデリングによる 一対比較法の表現

本稿では構造方程式モデリングによる一対比較法を利用することで、社会的望ましさによる反応歪曲に抗する尺度構成法を表現する。特に第II部・5章では(豊田・室橋・尾崎・芳賀, 2004a; 豊田・尾崎・室橋・芳賀, 2004b)によって示されたシェフェー一対比較法のSEMによる表現法について概観されている。

5.1 SEMによる被験者内計画の為の一対比較法

本稿のモデルを基礎付ける被験者内計画の為のシェフェー一対比較モデルは次式で定義される。

$$x_{ijk} = (\alpha_j - \alpha_k) + \tau_{jk} + \delta_{jk} + e_{ijk} \quad (5.1)$$

ここで x_{ijk} は被験者 $i (= 1, \dots, I)$ の、比較試料 (j, k) の一対比較判断であり、例えば5件法、7件法で測定される。

α_j (α_k) は試料 j (k) に対する集団全体で定義される平均的嗜好度であり、 τ_{jk} は比較試料 (j, k) の組み合わせ効果を、そして δ_{jk} は試料提示に関する順序効果をそれぞれ意味している。最後の e_{ijk} は x_{ijk} に対応する測定誤差である。

またモデル識別の為に、平均的嗜好度と組み合わせ効果に関して以下の制約を入れる。

$$\sum_j \alpha_j = 0, \quad \sum_j \tau_{jk} = \sum_k \tau_{jk} = 0, \quad \tau_{jk} = -\tau_{kj} \quad (5.2)$$

以上がシェフェー一対比較モデルの概要である。(5.1)式において推定対象となる母数は、平均的嗜好度 α_j 、組み合わせ効果 τ_{jk} 、順序効果 δ_{jk} 、そして e_{ijk} の分散 $\sigma_{e_{jk}}^2$ である。また全ての項目について $\sigma_{e_{jk}}^2$ は等しいという仮定が置かれる。

このシェフェーの一対比較モデルに対して(豊田他, 2004a,b)は次のようにシェフェーモデルを拡張している。

$$x_{ijk} = (\alpha_j - \alpha_k) + (f_{ij} - f_{ik}) + \tau_{jk} + \delta_{jk} + \epsilon_{ijk} \quad (5.3)$$

(5.1) 式との本質的な差異は、試料 $j(k)$ に対する個人の嗜好度 $f_{ij}(f_{ik})$ の導入である。 $f_{ij}(f_{ik})$ は潜在変数における因子得点であり、SEM を利用したことによって初めて表現が可能となる母数である。項目反応 x_{ijk} には、全体で定義される嗜好度 α だけではなく、個人で定義される嗜好度も少なからず反映されるというアイデアは実質科学的に妥当である。

第6章 不完全一対比較データへの SEMによる対処

本章では、SEMによる一対比較法による尺度構成法を論じる前に、心的負担が大きくなりやすい一対比較法の欠点と、それに対する統計的対処法について論じた。

一対比較法を利用した調査では、比較試料数が多い場合には回答における被験者の心的負担が甚大になるという問題点がある(豊田他, 2004a,b)。本章ではこの問題に関して、SEMによる多母集団同時分析を用いた対処を行なった。具体的には混合計画の為のSEMによるシェフェー一対比較法を提示した。

混合計画に対応するシェフェー一対比較法は、SEMにおける平均構造を導入した多母集団共分散構造モデルによって表現する。具体的には母集団の区別を、被験者間要因である折半された項目群の区別とする。そして各母集団毎でそれぞれ異なる一対比較項目が被験者内要因として表現される。各集団では(豊田他, 2004a,b)で表現された被験者内計画の為の構造方程式モデリングによるシェフェー一対比較法が適用される。

6.1 不完全データの多母集団解析・交互作用の制約

折半された項目群をそれぞれ一母集団とみなした。各母集団には下記の一対比較モデルが適用された。

$$x_{ij}^{(1)} = \tau_{ij}^{(1)} + (\alpha_i^{(1)} - \alpha_j^{(1)}) + (f_i^{(1)} - f_j^{(1)}) + e_{ij}^{(1)} \quad (6.1)$$

$$x_{ij}^{(2)} = \tau_{ij}^{(2)} + (\alpha_i^{(2)} - \alpha_j^{(2)}) + (f_i^{(2)} - f_j^{(2)}) + e_{ij}^{(2)} \quad (6.2)$$

記号の右肩に記載されている括弧付きの数字は、それぞれ母集団の区別を表現している。

2つの群は異なる一対比較項目群を評価する(モデルが異なる)という点で区別されるが、本質的に同一母集団から抽出された被験者で構成される。従って下式

の制約を入れることは妥当である.

$$\begin{aligned} E[f_i^{(1)}(f_j^{(1)})] &= E[f_i^{(2)}(f_j^{(2)})] \\ V[f_i^{(1)}(f_j^{(1)})] &= V[f_i^{(2)}(f_j^{(2)})] \\ \alpha_i^{(1)}(\alpha_j^{(1)}) &= \alpha_i^{(2)}(\alpha_j^{(2)}) \end{aligned}$$

ここで σ_e^2 に関しては、シェフェモデルの根本的制約として、群を超えて全ての項目で等値制約が仮定される.

交互作用の推定の為には上述したように $\sum_{i \neq j} \tau_{ij} = 0$, $\tau_{ij} = -\tau_{ji}$ という制約が必須である. また交互作用 τ_{ij} に、群の識別子 k を加えて,

$$\sum_{k=1}^2 \sum_{i \neq j} \tau_{ijk} = 0, \quad \tau_{ijk} = -\tau_{jik} \quad (6.3)$$

のように母集団を超えて制約を置く.

6.2 結果・考察

混合計画の為のシェフェ一対比較法の開発に関しては、SEMにおける多母集団分析を利用することで、シミュレーションで得られた完全データにおける解がほぼ再現できることを確認した. モデル適合度の観点からも安定しており、実用的な手法であることが示された.

第7章 一対比較型尺度構成法の基礎 モデル

前章までの議論を受け、本章では、本稿の趣旨である社会的望ましさに抗する一対比較型尺度構成モデルについて詳細に論じた。

選抜的評価場面という状況において、リッカート法による典型値検査は社会的望ましさによる意図的な反応歪曲に脆弱である。この問題に対して本章では、意図的な反応歪曲の難しい一対比較法を利用した尺度構成法を提案する。測定対象とする構成概念には EQ(Emotional Intelligence quotient) を選定した。

7.1 測定するEQの領域・尺度

Salovey & Mayer (1990), Mayer & Salovey (1993), Goleman (2000) は、EQ の基本的な領域として、1) 自分自身の情動を知る、2) 自分の感情を制御する、3) 自分を動機付ける、4) 他人の感情を認識する、5) 他者との人間関係をうまく処理する、という5つを挙げている。1), 2), 3) は自分に関する領域、4), 5) は他者に関する領域、1), 4) は知ることに関する領域、2), 3), 5) は働きかけに関する領域である。そこで本研究では、自己か他者か、知ることか、働きかけに注目し、2)と3)を1つにまとめて、

- a. ”自己認識”：自己感情の識別、上述(1)
- b. ”自己制御”：自己感情の利用、上述(2), (3)
- c. ”対人理解”：対人感情の認知、上述(4)
- d. ”対人技能”：対人感情の調整、上述(5)

の4つとする。

次にこの4つの領域内で、それぞれ3つの下位尺度 x, y, z を構成する。

”自己認識”は、自分の気持ちや感情を読み取り、自覚する能力とし、下位尺度 ax ”モニタリング”， ay ”適切な自己評価”， az ”肯定的自己像”を設定する。

”自己制御”は、自分の気持ちや感情を適切に統制し、活用する能力とし、下位尺度 bx ”自己抑制”， by ”達成動機・統制感”， bz ”安定感・柔軟性”を設定する。

”対人認知”は、相手の気持ちや感情を読み取り、共感する能力とし、下位尺度 cx ”共感性”， cy ”センシブル”， cz ”サービス精神”を設定する。

”対人技能”は、相手の気持ちや感情を適切に活用する能力とし、下位尺度 dx ”リーダーシップ”， dy ”アサーション”， dz ”チームワーク”を設定する。

12の尺度で、それぞれ6項目（テスト全体で72項目）を作成する。具体的な評定方法は、2つの試料 [mik] と試料 [njl] を比較して、どちらが被験者に当てはまっているかを考えさせ、当てはまり具合を5段階評定尺度上に1個所だけ○印を付けることで表現させた。

m と n は領域を示す添え字であり、 a, b, c, d の値をとる ($m \neq n$)。 i と j は下位尺度を示す添え字であり、 x, y, z の値をとる。下位尺度は領域の入れ子であるから、 i と j は、それぞれ m と n にネストし、正確に書くなら $i(m)$ と $j(n)$ である。 k と l は、下位尺度を構成する項目番号であり、1から当該下位尺度の項目数まで動く。

7.2 一対比較型テストモデル

観測変数は以下の測定方程式によって

$$x_{miknjl} = \delta_{miknjl} + (f_{mik} - f_{njl}) + \tau_{mkinjl} + e_{miknjl} \quad (7.1)$$

$$f_{mik} = \alpha_{mik} + f_{mi} + d_{mik} \quad (7.2)$$

$$f_{njl} = \alpha_{njl} + f_{nj} + d_{njl} \quad (7.3)$$

のように分解する。(7.1)式～(7.3)式で定義される一対比較型テストモデルは、基本的に豊田ら(2004)に基づいてはいるが、複数の潜在的嗜好度 f_{mik} (f_{njl}) の背後に、これを統合する2次因子 f_{mi} (f_{nj}) を仮定している点で大きく異なる。

α_{mik} (α_{njl}) と τ_{minj} が母数(定数)であり、残りが確率変数(外生変数)である。右辺の確率変数の期待値は全て0であり、分散は自由母数として全て推定する。

共分散は f_{mi} と f_{nj} の間に仮定される以外は、全て0に固定する。

右辺の各項の解釈的意味を説明する。まず α_{mik} は (α_{njl} は)，個人の特性とは独立の項目 mik の重要度 (項目 njl の重要度) であり、ここに社会的望ましさが含まれる。

$$E[x_{miknjl}] = \alpha_{mik} - \alpha_{njl} \quad (7.4)$$

であるから、 α_{njl} が一定で、 α_{mik} が大きくなれば、当該観測変数の期待値も大きくなる。またここでは社会的望ましさを、EPPSにおける定義と同様に、集団における値として定義している。よって(7.2)式における α は、各被験者ではなく被験者集団の代表値として算出される。

f_{mi} は (f_{nj} は)，尺度 mi における (尺度 nj における) 個人の尺度値である。平均共分散構造モデルでは、因子パターンが1か-1に固定された構成概念として表現される。

δ_{miknjl} は観測変数の切片を, τ_{miknjl} は試料 mik と試料 njl の組み合わせ効果を表現する項である. e_{miknjl} は測定誤差を表現している.

7.3 標準化・妥当性研究の結果

本章では上述したテストモデルに関して、妥当性研究と標準化研究を行った。

研究1では一対比較型EQ検査の反応歪曲への耐性を検討するために、140名の被験者に対して一対比較型EQ検査と、同一項目で構成されたリッカート測定法によるEQ検査を、選抜的評価場面、非選抜的評価場面の2場面を想定して実施した。その結果、一対比較型EQ検査は測定場面にかかわらず、標準偏差、効果量、収束的妥当性の観点からリッカート測定法よりも反応歪曲に耐性を持っていることが確認された。

研究2では一対比較型EQ検査のスコアリングを安定させる為、新たに用意された383名の被験者のデータを研究1の被験者のデータに加えて計算を行った。また代替検査の開発の為にテストを20パターンに折半し、10パターンそれぞれで尺度間相関を求めた。その結果代替検査信頼性の観点から支持される代替テストが得られた。

第8章 社会的望ましさの尊重度を考慮したテストモデル

本章では、前章で提案されたテストモデルに、社会的望ましさの尊重度を導入することで、データに対してより妥当なテストモデルを提示した。

8.1 テストモデル

前章で提示した一対比較型テストモデルは、4つのEQ領域、12の下位尺度とそれに含まれる7項目という高次の構造を表現していた。しかし心理検査は通常、尺度と項目という構成で作成されるから、前章(7.1)式～(7.3)式は

$$x_{ikjl} = \delta_{ikjl} + (f_{ik} - f_{jl}) + \tau_{ikjl} + e_{ikjl} \quad (8.1)$$

$$f_{ik} = \alpha_{ik} + f_i + d_{ik} \quad (8.2)$$

$$f_{jl} = \alpha_{jl} + f_j + d_{jl} \quad (8.3)$$

と領域の為の添え字 $m(n)$ を省略して表現できる。

ただ(8.1)式～(8.3)式には更なる2つの改良の余地がある。1つは、項目の社会的望ましさが全ての受験者に同様に影響していると仮定している点である。社会的規範の一部ともいえる項目の社会的望ましさの体系を比較的強固に受け入れている個人と、全く気にしない個人の違いを表現することは、個人特性(あるいは個人の状態)の記述という観点からは有効かも知れない。

もう1つは、(8.2)式、(8.3)式は因子パタンが1か-1かに固定された平均構造のある確認的因子分析の形式で表現されている点である。一対比較モデルは、特性の差を表現する際に1か-1の係数を利用することが多い。しかし因子パタンの等しい因子分析モデルは、全ての項目の信頼性が等しいというタウ等価測定のモデル化である。現実には、必ずしもタウ等価測定が成り立つとは限らないので因子パタンを項目の信頼性にあわせて推定することは有効かもしれない。

また(7.2)式、(7.3)式における f_i は、モデルとしてはイプサティブスコア¹とし

¹ イプサティブスコアとは被験者毎に尺度を考え、変数を尺度上に位置づける得点である(豊田, 2000)。イプサティブ変数はその和が一定に調節されており、例えば5件法の一対比較法ならば“どちらでもない”への反応は両対象に3点を、“AよりBがやや当てはまる”への反応に、Aに4点、Bに2点を与えるならば総和は6で一定となっている。

て導入されていないが、タウ等価の仮定が入っているために、通常の SEM のソフトで構成概念スコアを推定すると推定量がイプサティブスコアになってしまうという弊害がある（これを回避するようにテスト得点を定義することは可能であるが、通常のソフトの範囲でできないのは不便である）。因子パタンを自由母数にすることにより、モデルばかりでなく、構成概念スコアの推定値もイプサティブスコアでなくなる。

そこで本章では、前章でその有効性が確認された(7.1)式～(7.3)の一対比較型テストモデルを

$$x_{ikjl} = \delta_{ikjl} + f_{ik} - f_{jl} + \tau_{ikjl} + e_{ikjl} \quad (8.4)$$

$$f_{ik} = \alpha_{ik} f_0 + a_{ik} f_i + d_{ik} \quad (8.5)$$

$$f_{jl} = \alpha_{jl} f_0 + a_{jl} f_j + d_{jl} \quad (8.6)$$

と拡張し、それに従ってオリジナルのテストを作成しその効果を調べる。

x_{ikjl} は項目 (i, k) と項目 (j, l) の比較を表現する観測変数である。 f_0 は、平均 1 で正の値をとることが期待される潜在変数であり、値が大きいほど α_{ik} による社会的望ましさの体系を尊重し、0 に近いほど社会的望ましさの体系を気にしないという個人傾向である。また「自分に当てはまっていると素直に言える」という別の解釈も可能であるが、この表現は社会的望ましさの体系を内面化した被験者の状態を表現していると考えれば、 f_0 は社会的望ましさの体系を表現していると考えることができる。 f_0 の導入によって前章(7.1)式～(7.3)式から増加する母数は、その分散だけであり、適合度の観点からは大きな影響はない予想される。

次に $a_{ik}(a_{jl})$ は、項目 ik (項目 jl) の因子パタンであり、その値が高いほど当該項目の信頼性は高いと解釈する。また前述の通り $f_i(f_j)$ は因子を表現している。この改良は導入した時点でイプサティブでないテスト得点を構成できるという長所を有するので比較成績が前章で論じた方法と同程度か、それ以上であれば導入の効果があると判断できる。 τ_{ij} は組み合わせ効果を、 δ_{ik} は切片を、 d_{ik}, e_{ikjl} は測定誤差をそれぞれ示している。

8.2 コンピテンシー尺度

テストモデルが適用される心理尺度は、専門家によって選定された 8 つのコンピテンシーである。本研究では第 1 尺度：“現状認識力”，第 2 尺度：“ビジョン創出力”，第 3 尺度：“達成行動力”，第 4 尺度：“自己管理力”，第 5 尺度：“メンタルタフネス”，第 6 尺度：“人間関係構築力”，第 7 尺度：“コミュニケーション力”，第 8 尺度：“組織適応力”という 8 つのコンピテンシーを測定対象とした。

8.3 標準化・妥当性研究の結果

上述したモデルに関して、本章では標準化研究、妥当性研究の2研究を行った。標準化研究において、予備調査で作成された8つのコンピテンシー尺度を用いて、一対比較検査を構成し、これをA・Bの2検査に折半し標準検査を作成した。400名の被験者に対して両検査を実施した結果、検査の信頼性・妥当性が確認された。

妥当性研究では標準化検査の反応歪曲への耐性を検討するために、新測定法とリックアート法による検査が、選抜的評価場面、非選抜的評価場面の2場面を想定して、420名の被験者に対して実施された。

選抜的評価場面・非選抜的評価場面におけるリックアート測定法、一対比較法のそれぞれの尺度得点の平均値・標準偏差・効果量の比較の結果、評価場面の変化に対して一対比較法における尺度得点は効果量、標準偏差の両方で、リックアート法に比較して不变性を持つことが明らかになった。つまり評価場面における反応歪曲に対して頑健であることが明らかになった。

また選抜的評価場面と非選抜的評価場面の統計量における大きな差異により新検査の妥当性を評価することは、リックアート測定法による検査と、一対比較法による検査において全く同一の項目を用いているという事実からも支持され、従って反応歪曲に抗する一対比較モデルとしての妥当性を本検査は保持している可能性が示された。

また非選抜場面において、自由推定された f_0 の期待値が極めて1に近かったことで、モデルの交差妥当性が示唆された。ただモデルが複雑になるため因子スコアの推定値が不安定になる傾向認められた。しかし選抜的評価場面における f_0 の期待値が1よりも低かったことで、 f_0 を利用する可能性が開けたといえ、モデルの安定性を含めて今後詳細に検討する余地があるといえる。

第9章 社会的望ましさの個人差を考慮したテストモデル

9.1 テストモデル

前章で論じた社会的望ましさへの尊重度を考慮した一対比較モデルについて、以下の利点を挙げることができる。

- 1) 一対比較法を利用したことによって意図的な歪曲が抑止される。
- 2) 各試料に対する平均的な社会的望ましさが、潜在的嗜好度の切片 $\alpha_{ik}(\alpha_{jl})$ に含まれ、尺度得点 $f_i(f_j)$ から分離される。
- 3) 潜在的嗜好度の切片 $\alpha_{ik}(\alpha_{jl})$ との積で、社会的望ましさの尊重度 f_0 が定義され、尺度得点 $f_i(f_j)$ から分離される。

一方で前章の方法には解決されるべき 2 点の問題があると思われる。

先ず前章の議論から、切片 α との積で定義される f_0 の導入はモデルの適合という観点からは不利であることが分かっている。

また先行研究では問題になることが多い、社会的望ましさの個人差が回答に及ぼす影響の評価・修正について考慮されていないという問題もある。

本章では第 7 章で構成された一対比較型 EQ 尺度を利用して、上記の 2 点の問題を克服した新たなテストモデルを提案した。

テストモデルが適用される尺度は、第 7 章で標準化が試みられた EQ 尺度である。以下の議論では第 7 章におけるモデル表記を利用するので留意されたい。

試料に対する潜在的な嗜好度と一対比較型の観測変数との関係を以下の測定方程式によって表現する。

$$x_{miknjl} = \delta_{miknjl} + f_{mi} - f_{nj} + \tau_{miknjl} + \epsilon_{miknjl} \quad (9.1)$$

$$f_{mik} = \alpha_{mik} + \beta_{mik} f_{(s)mik} + a_{mik} f_{mi} + d_{mik} \quad (9.2)$$

$$f_{njl} = \alpha_{njl} + \beta_{njl} f_{(s)njl} + a_{njl} f_{nj} + d_{njl} \quad (9.3)$$

$m(n)$ は尺度の領域の識別子であり a, b, c, d の 4 状態を取る。次に $i(j)$ は各領域に含まれる下位尺度の識別子であり x, y, z の 3 状態を取る。また $k(l)$ は各尺度に含まれる試料の識別子であり 1 から 6 の値をとる。

x_{miknjl} は一対比較型の観測変数であり、領域 m の尺度 i を測定する第 k 番目の試料 (m, i, k と領域 n の尺度 j を測定する第 l 番目の試料との一対比較を表現している。ただし $m \neq n$ であり同一領域内での比較は行わない。

δ_{miknjl} と τ_{miknjl} は観測変数の切片と試料間の交互作用をそれぞれ表現している。両母数はモデルの識別を考慮して 0 に固定された。 ϵ_{miknjl} は観測変数の測定誤差である。

f_{mik} は試料 (m, i, k) の潜在的な嗜好度である。また $f_{mi}(f_{nj})$ はその相関行列を説明する下位尺度因子であり、標準正規分布に従うことが仮定された。 $a_{mik}(a_{njl})$ は $f_{mi}(f_{nj})$ から潜在的嗜好度への因子パタンである。 $f_{(s)mik}(f_{(s)njl})$ と $\beta_{mik}(\beta_{njl})$ は社会的望ましさの個人差と、その潜在的嗜好度への影響を示しており、以下に示される (9.4) 式で定義される。

$\alpha_{mik}(\alpha_{njl})$ は因子の切片であり社会的望ましさの個人差と、下位尺度得点が 0 である時の潜在的嗜好度の値と解釈される。この母数には受験者集団全体で定義される社会的望ましさが含まれると仮定される。

(9.2) 式に明らかのように、 $\beta_{mik}(\beta_{njl})$ の絶対値が大きな試料ほど、社会的望ましさの個人差から強い影響を受けると解釈することができる。しかし提案モデルでは、要因として社会的望ましさの個人差を取り上げているので、仮に絶対値の大きな β が得られたとしても、その影響を排除した上で、各試料に対する嗜好の個人差が得られる。

(9.4) 式は社会的望ましさの個人差 $f_{(s)mik}(f_{(s)njl})$ を定義する測定方程式である。

$$x_{(s)miknjl} = \delta_{(s)miknjl} + f_{(s)mik} - f_{(s)njl} + \tau_{(s)miknjl} + \epsilon_{(s)miknjl} \quad (9.4)$$

$x_{(s)miknjl}$ は採用試験官の立場から (9.1) 式で用いたものと同一の試料を一対比較評定させた結果である。

この測定方程式では、 $\delta_{(s)miknjl} = 0$, $\tau_{(s)miknjl} = 0$ というモデル識別の為の仮定の他に、

$$\begin{aligned} E[f_{(s)mki}(f_{(s)njl})] &= 0, \quad V[f_{(s)mki}(f_{(s)njl})] = 1, \\ E[\epsilon_{(s)miknjl}] &= 0, \quad Cov[f_{(s)mki}, f_{(s)njl}] = 0 \end{aligned}$$

を仮定する。特に $E[f_{(s)mki}(f_{(s)njl})] = 0$ という仮定は、(9.1) 式の期待値が

$$E[x_{miknjl}] = \alpha_{mik} - \alpha_{njl} \quad (9.5)$$

に一致する為の便宜である。(9.5) 式は切片に含まれる社会的望ましさを評価する上で必要とされる仮定である(豊田・川端・松下, 2005)。また $Cov[f_{(s)mki}, f_{(s)njl}] = 0$ という制約は、推定結果の安定性を考慮したものである。

9.2 標準化と妥当性研究

本章では提案テストモデルの標準化と妥当性が試みられた。

標準化研究では正直回答条件と、社会的望ましさ条件において被験者内で回答した 199 名のデータを完全データ群、第 7 章で利用した 530 名の正直回答のデータを欠測データ群として、テストモデルに関して SEM による多母集団同時分析を適用した。代替検査の作成も行った。

標準化の結果、算出された β 推定値を固定母数として、(9.2) 式、(9.3) 式のテストモデルに与えることで、社会的望ましさの個人差にまつわる評定を行なわなくとも、その影響を取り除いた尺度得点(因子スコア)を求めることが可能になった。

実際の採用試験場面では、社会的望ましさの評定を併せて行なうことは、被験者の疲労・実施時間・運用費用の観点から必ずしも現実的な選択ではない。 β_{mik} の推定値が得られているならばこの問題は解消される。また β_{mik} を含んだモデルの外的妥当性については、後に測定されたトレーニングデータによって向上させることができある。

標準テストの信頼性については、代替検査信頼性の観点から許容範囲であることが示された。

妥当性研究は以下の 3 つの観点から検討された。先ず評価場面間での平均値に関する効果量の観点から、新手法は特定の尺度においてリッカート型尺度よりも反応歪曲の補整に対して高い効果を持っていることが示された。

また潜在プロファイルモデル分析により、新尺度は選抜的評価場面での回答に含まれる社会的望ましさの個人差の影響を適切に評価することが示された。

更に、各試料の α_{mik} と β_{mik} によるプロファイル分析により、 α_{mik} と β_{mik} にそれぞれ集団で定義される社会的望ましさ、個人で定義される社会的望ましさが反映されることが明らかとなった。

引用文献(概要内)

- Alliger, G. & Dwight, S. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching, *Educational and Psychological Measurement*, **60**, 59–72.
- Caldwell-Andrews, A., Baer, R. A. & Berry, D. T. R. (2000). Effects of response sets on neo-pi-r scores and their relations to external criteria, *Journal of Personality Assessment*, **74**, 472–488.
- Christiansen, N. D., Edelstein, S. & Fleming, B. (1998). Reconsidering forced-choice formats for applicant personality assessment, in *Paper presented at the meeting of the Society of Industrial and Organizational Psychology, Dallas, TX*.
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology, *Journal of Consulting Psychology*, **24**, 349–354.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*: New York: Dryden Press.
- Edwards, A. L. (1959). *Manual, Edwards personal preference schedule*: New York: The psychological Corporation.
- Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*: New York: Holt, Rinehart, & Winston.
- Goleman, D. (2000). *Working with emotional intelligence*: Bantam Dell Pub Group, (梅津祐良訳, ビジネス EQ — 感情コンピテンスを仕事に生かす—, 東洋経済新報社, 2000年) .
- 南風原朝和 (2002). 心理統計学の基礎, 有斐閣.
- Heggestad, E. D., Morrison, M., Reeve, C. L. & Christiansen, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance, *Journal of Applied Psychology*, **91**, 9–24.

- Horvitz, D. & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, **47**, 663–685.
- 星野崇宏 (2003). 調査データに対する傾向スコアの適用, *品質*, **33**, 44–51.
- 星野崇宏・繁樹算男 (2004). 傾向スコア解析法による因果効果の推定と調査データの調整について, *行動計量学*, **31**, 43–61.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D. & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities, *Journal of Applied Psychology*, **75**, 581–595.
- Jackson, D. N., Wroblewski, V. R. & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance*, **13**, 371–388.
- Kanning, U. & Kuhne, S. (2006). Social desirability in a multimodal personnel selection test battery, *European Journal of Work and Organization Psychology*, **15**, 241–261.
- 狩野裕 (2002). 構造方程式モデリング, —因果推論そして非正規性. 多変量解析の展開—甘利俊一・狩野裕・佐藤俊哉・松山裕・竹内啓・石黒真木夫(編), 岩波書店, 64–130.
- Mayer, J. D. & Salovey, P. (1993). The intelligence of emotional intelligence, *Intelligence*, **17**, 433–442.
- Mayer, J. D., Salovey, P. & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications, *Psychological Inquiry*, **15**, 197–215.
- 日経BPコンサルティング (2002). ブランドジャパン2002 総合報告書[解説書], 日経BPコンサルティング.
- Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**, 41–55.
- Salovey, P. & Mayer, J. (1990). Emotional intelligence, *Imagination, cognition, and personality*, **9**, 185–211.
- 佐藤俊哉 (2002). 傾向スコアを用いた因果効果の推定, 多変量解析実例ハンドブック 柳井晴夫・繁樹算男・高木寛文・岩崎学(編), 朝倉書店, 240–250.

- 鈴木督久・星野崇宏 (2004). 傾向スコアを巡る対話, マーケティング・リサーチャー, **24**, 32–38.
- 登張真穂 (2007). 社会的望ましさ尺度を用いた社会的望ましさ修正法——その妥当性と有効性——, パーソナリティ研究, **15**, 228–239.
- 豊田秀樹 (1996). 非線形多変量解析 —— ニューラルネットによるアプローチ——, 朝倉書店.
- 豊田秀樹 (2000). 共分散構造分析 [応用編] —— 構造方程式モデリング——, 朝倉書店.
- 豊田秀樹・川端一光・松下信武 (2005). 採用場面におけるEQ検査の改善, 教育心理学研究, **53**, 456–466.
- 豊田秀樹・室橋弘人・尾崎幸謙・芳賀麻薺美 (2004a). 実験デザインに基づく一対比較データの解析 —— 構造方程式モデリングによる表現——, 心理学研究, **75**, 33–40.
- 豊田秀樹・尾崎幸謙・室橋弘人・芳賀麻薺美 (2004b). 構造方程式モデリングによる一対比較法の分析 —— シェッフェの方法とその改良——, 心理学研究, **75**, 299–307.
- Visweswaran, C. & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement, *Educational and Psychological Measurement*, **59**, 197–210.
- Wiggins, J. S. (1959). Interrelationships among mmri measures of dissimulation under standard and social desirability instructions, *Journal of Consulting Psychology*, **23**, 419–427.