博士学位論文概要

多特性多方法データを用いて測定の信頼性と妥当性を定量的に評価するための方法に関する研究

久保 沙織

第1章　問題および目的

一般に，心理学的測定は信頼性と妥当性という 2 つの観点から評価される。とりわけ，構成概念という目に見えないものを測定の対象とすることが多い心理学においては，妥当性の検討が重要な問題となる。しかしながら，妥当性はその概念を集約的に解釈することは難しく，様々な側面からの検討を要する。妥当性の証拠を集める側面として，Messick (1995) は，内容的側面，本質的側面，構造的側面，一般化可能性の側面，外的側面，結果的側面の 6 つを挙げている。これら 6 つのうち，外的側面は，収束的妥当性と弁別的妥当性という 2 つの観点を含む。収束的妥当性と弁別的妥当性の検証においては，多特性多方法 (multitrait-multimethod; MTMM) 行列 (Campbell & Fiske, 1959) が有用である。本論文では，MTMM 行列を分析するための確認的因子分析モデルに焦点を当てる。

t 個の特性を m 個の方法によって測定した MTMM データに関する確認的因子分析モデルは，以下のように表現される。

$$\boldsymbol{x} = \boldsymbol{\Lambda}_T \boldsymbol{f}_T + \boldsymbol{\Lambda}_M \boldsymbol{f}_M + \boldsymbol{e} \tag{1}$$

$\boldsymbol{x}$ は $tm \times 1$ の観測変数ベクトルであり，$\boldsymbol{f}_T$ は特性因子，$\boldsymbol{f}_M$ は方法因子をそれぞれ縦に並べたベクトル，$\boldsymbol{e}$ は誤差変数ベクトルである。

観測変数間の共分散構造行列 $\boldsymbol{\Sigma}$ は，以下となる。

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_T \boldsymbol{\Phi}_T \boldsymbol{\Lambda}_T' + \boldsymbol{\Lambda}_M \boldsymbol{\Phi}_M \boldsymbol{\Lambda}_M' + \boldsymbol{\Psi} \tag{2}$$

ただし，$\boldsymbol{\Lambda}_T$ と $\boldsymbol{\Lambda}_M$ はそれぞれ特性因子と方法因子の因子負荷行列である。$t \times t$ 行列 $\boldsymbol{\Phi}_T$ と $m \times m$ 行列 $\boldsymbol{\Phi}_M$ は，それぞれ特性因子間，方法因子間の相関行列であり，$E[\boldsymbol{e}\boldsymbol{e}'] = \boldsymbol{\Psi}$ は対角要素に誤差分散 $\sigma^2_{e_{ij}}$ を配した $tm \times tm$ 対角行列である。

(1) 式は，測定において用いられた t 個の特性と m 個の方法をそれぞれ因子として仮定したモデルであり，それらが独立に観測変数を説明するモデルは CT-CM (correlated trait-correlated method) モデル (Marsh, 1989; Widaman, 1985) と呼ばれる。CT-CM モデルは，特性と方法をそのまま因子として扱うため，モデルの仮定が明快で，解釈が容易である。一方で，しばしばモデルが識別されない場合があること，不適解を生じやすいことの 2 点が大きな問題として指摘されてきた。CT-CM モデルにおける不適解と識別の問題を回避するためのモデルとして，CT-CU (correlated trait-correlated uniqueness) モデル (Kenny, 1976) や CT-UM (correlated trait-uncorrelated method) モデル (Grayson & Marsh, 1994; Marsh & Grayson, 1995)，そして CT-C(M−1) (correlated trait correlated method minus one) モデル (Eid, 2000) などがある。しかしながら，これら 3 つのモデルも，それぞれ解釈面での問題があり，MTMM 行列に対する確認的因子分析モデルには，汎用性に優れたモデルがいまだに存在しない。

また，MTMM データにおける “方法” は，一般的に 3 つの種類に分けられる (Kenny, 1994) 。1 つ目は評価者 (rater) ，2 つ目は測定方法 (instrument-based) ，3 つ目は測定機会 (temporally-based) である。このうち評価者を “方法” として扱う MTMM データは特に，多特性多評価者 (multitrait-multirater; MTMR) データ (Conway, 1996) と称され，主として，企業における 360 度フィードバックの結果として得られる。MTMR データの分析においては，MTMR データ特有の問題を考慮して信頼性と妥当性の検討を行う必要がある。

以上を踏まえ，本論文の目的は，MTMM データを用いた信頼性および妥当性の検証に利用される確認的因子分析モデルについて，その問題点を改善し，より実用的な分析方法を提案することとする。まず第 2 章で論じる研究 I では，信頼性と妥当性の解釈が必ず一意に定まり，かつ識別可能性の高いモデルを開発した。続く研究 II（第 3 章）および研究 III（第 5 章）では，MTMR データの分析に焦点を絞り，データの持つ特徴を考慮して，信頼性と妥当性の推定および解釈をより精緻に行うための方法について論じた。研究 II は，各特性の合計得点の信頼性と収束的妥当性，弁別的妥当性の検討を定量的に行い，その結果をもとに評価者ごとに適切な項目配分を決定するための方法に関する研究である。研究 III では，同一立場内に複数の評価者が存在する場合に，その影響を適切に評価した上で信頼性と妥当性を検討するための方法を提案した。

第 2 章　研究 I: 信頼性と妥当性の解釈が一意に定まるモデル

目的　研究 I では，信頼性および収束的妥当性・弁別的妥当性の解釈が 1 つのデータに対して一通りに定まるモデルを提案することを目的とする。これまでに提案されてきたモデルのうち，CT-CM モデルは，1 つのデータに対して必ず一通りのパラメータセットが得られ，信頼性や妥当性の解釈が一通りに定まるというメリットがある。しかしながら，モデルの識別不定および不適解の発生が，実用における深刻な課題である。これに対して，CT-C(M−1) モデルは，必ず識別されることが証明されている。しかし，同一データに対して m（方法の数）通りのパラメータセットが得られるため，信頼性や妥当性に関する解釈が一意に定まらないという欠点がある。そこで本研究では，実質科学的に解釈可能な仮定の下で，汎用的かつ実用的なモデルを提案する。

方法　提案モデルでは，方法因子の因子得点の和を 0 とする制約を導入する。この方法では，データと推定結果が必ず一対一対応で定められる。また，特性因子と方法因子とは無相関とすることで，分散成分の分解を利用して収束的妥当性・弁別的妥当性の解釈を容易に行うことができる。

方法因子の因子得点の和が 0 という制約を $m \times (m-1)$ の計画行列 $\boldsymbol{k}$ を用い

て表現すると，(1) 式は (3) 式に，(2) 式は (4) 式に書き換えられる。

$$\boldsymbol{x} = \boldsymbol{\Lambda}_T \boldsymbol{f}_T + \boldsymbol{\Lambda}_M \boldsymbol{k} \boldsymbol{f}_{M-1} + \boldsymbol{e} \tag{3}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_T \boldsymbol{\Phi}_T \boldsymbol{\Lambda}'_T + \boldsymbol{\Lambda}_M \boldsymbol{k} \boldsymbol{\Phi}_{M-1} \boldsymbol{k}' \boldsymbol{\Lambda}'_M + \boldsymbol{\Psi} \tag{4}$$

計画行列 $\boldsymbol{k}$ は，サイズ $m-1$ の単位行列の下にすべての要素が -1 の行ベクトルを結合した $m \times (m-1)$ の行列である。$\boldsymbol{f}_{M-1}$ は，計画行列 $\boldsymbol{k}$ に対応して方法因子 $\boldsymbol{f}_M$ の最後の 1 つを除いた因子得点ベクトルであり，同様に，$\boldsymbol{\Phi}_{M-1}$ は $(m-1) \times (m-1)$ とした方法因子間相関行列である。

(4) 式は $(m-1)$ 個の因子を仮定したモデルで $\boldsymbol{\Lambda}_M$ および $\boldsymbol{\Phi}_{M-1}$ を推定することを意図しているが，$\boldsymbol{k}\boldsymbol{\Phi}_{M-1}\boldsymbol{k}'$ を展開することで，それは m 個の方法因子を仮定したモデルで因子間の分散と共分散に制約を課すことと同値であることが確認できる。したがって，提案モデルでは，すべての方法因子について因子負荷量と分散・共分散が自由母数として推定される。結果として，信頼性と妥当性の解釈を必ず一通りに定めることができる。

適用例では，過去の文献から引用した 12 の相関行列を分析対象とした。それぞれのデータに対して提案モデル，CT-CM モデル，CT-C(M−1) モデルの 3 種類のモデルを適用し，推定結果について，モデルの識別と不適解の有無という 2 つの観点から比較検討を行った。さらに，推定値を利用して，信頼性係数と，妥当性に関する一貫性係数や方法特異性係数 (Eid, 2000) を計算し，各モデルの特徴を考察した。また，提案モデルの識別の可能性について，シミュレーションデータによる検証を行った。

結果と考察　まず，シミュレーション研究の結果，19683 通りすべての相関行列において収束が確認された。したがって，提案モデルは識別の可能性が非常に高いと考えられる。実データに対する適用例の結果から，CT-CM モデルは，すでに多くの文献で指摘されている通り，識別不定や不適解に陥りやすいことが確認された。一方で CT-C(M−1) モデルでは，どの方法を基準とするかによって，m 通りの信頼性と収束的妥当性，弁別的妥当性に関する解釈が可能となってしまい，特に，信頼性の低いデータに対してはこの影響が深刻な問題となる可能性が示唆された。

これらのモデルに比較して，方法因子の因子得点の和を 0 とする制約を導入した提案モデルは，信頼性および妥当性の解釈を必ず一通りに行うことができ，より汎用性の高いモデルである。さらに，本論文で用いた 12 の MTMM 行列に対して提案モデルが識別されない例はなく，シミュレーション研究の結果と併せても，提案モデルは識別の可能性が高い有望なモデルであることが確認された。

第 3 章　研究 II: 特性ごとの合計得点における信頼性と妥当性の検討

目的　360 度フィードバックデータをはじめとする MTMR データの分析では，それぞれの特性に関する異なる評価者からの評定結果を合計した得点（特性値）に興味がある場合も多いだろう。そこで研究 II では，評価者によって評価項目数を変化させた場合の，合計得点における信頼性係数と妥当性係数の一般式を導出することを目的とした。適用例においては 360 度フィードバックの実データを利用し，各評価者に対する項目数の配分と，特性ごとの合計得点の信頼性および妥当性との関係について考察した。

方法　本研究では，CT-UM モデルを利用して分析を行った。研究 I で用いた一貫性係数や方法特異性係数は，係数の意味付けが複雑で解釈が難しかった。そこで研究 II では，解釈のしやすさを優先し，研究 I で定義した信頼性や妥当性に関する係数を再定義した。

信頼性係数は，研究 I の場合と同様に，

$$\mathrm{Rel}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 \mathrm{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \mathrm{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \mathrm{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \mathrm{var}(f_{M_j}) + \mathrm{var}(e_{ij})} \tag{5}$$

とする。収束的妥当性は，

$$\mathrm{Con}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 \mathrm{var}(f_{T_i})}{\lambda_{T_{ij}}^2 \mathrm{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \mathrm{var}(f_{M_j}) + \mathrm{var}(e_{ij})} \tag{6}$$

によって検討することが可能である。(6) 式の値が大きいほど，収束的妥当性が高いと解釈できる。一方で，非弁別的妥当性係数を，

$$\mathrm{Dis}(x_{ij}) = \frac{\lambda_{M_{ij}}^2 \mathrm{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \mathrm{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \mathrm{var}(f_{M_j}) + \mathrm{var}(e_{ij})} \tag{7}$$

で定義し，(7) 式の値が大きいほど弁別的妥当性が低いと解釈する。

ここで，特性ごとの合計得点の信頼性と妥当性について考える。いま，360 度フィードバックの実施により，自己評価，上司評価，同僚評価が得られているとする。3 つの異なる評価者による評定結果を，それぞれ $\{\alpha_1, \alpha_2, \alpha_3\}$ 倍して合計すると，その和得点 x_{sum3} の分散は

$$\begin{aligned}\mathrm{var}(x_{\mathrm{sum3}}) = &(\alpha_1\lambda_{T_{i1}} + \alpha_2\lambda_{T_{i2}} + \alpha_3\lambda_{T_{i3}})^2 \mathrm{var}(f_{T_i}) \\ &+ (\alpha_1\lambda_{M_{i1}})^2 \mathrm{var}(f_{M_1}) + (\alpha_2\lambda_{M_{i2}})^2 \mathrm{var}(f_{M_2}) + (\alpha_3\lambda_{M_{i3}})^2 \mathrm{var}(f_{M_3}) \\ &+ \alpha_1^2 \mathrm{var}(e_{i1}) + \alpha_2^2 \mathrm{var}(e_{i2}) + \alpha_3^2 \mathrm{var}(e_{i3})\end{aligned} \tag{8}$$

となる。自己，上司，同僚のそれぞれで項目数が増減しても，各評価者の立場ごとに，もとの調査票による測定に対してタウ等価測定が成り立っているものとし，かつ，誤差分散の平均は変化しないと仮定する。ここで，x_{sum3} を，本論文では再構成スコアと呼ぶ。再構成スコアの分散である (8) 式を分母として，(5) 式〜(7) 式を再表現すると，合計得点における信頼性係数と収束的妥当性係数，非弁別的妥当性係数はそれぞれ以下のように導出される。

$$\mathrm{Rel}(x_{\mathrm{sum3}}) = \frac{(\alpha_1\lambda_{T_{i1}}+\alpha_2\lambda_{T_{i2}}+\alpha_3\lambda_{T_{i3}})^2\mathrm{var}(f_{T_i})+(\alpha_1\lambda_{M_{i1}})^2\mathrm{var}(f_{M_1})+(\alpha_2\lambda_{M_{i2}})^2\mathrm{var}(f_{M_2})+(\alpha_3\lambda_{M_{i3}})^2\mathrm{var}(f_{M_3})}{\mathrm{var}(x_{\mathrm{sum3}})} \tag{9}$$

$$\mathrm{Con}(x_{\mathrm{sum3}}) = \frac{(\alpha_1\lambda_{T_{i1}}+\alpha_2\lambda_{T_{i2}}+\alpha_3\lambda_{T_{i3}})^2\mathrm{var}(f_{T_i})}{\mathrm{var}(x_{\mathrm{sum3}})} \tag{10}$$

$$\mathrm{Dis}(x_{\mathrm{sum3}}) = \frac{(\alpha_1\lambda_{M_{i1}})^2\mathrm{var}(f_{M_1})+(\alpha_2\lambda_{M_{i2}})^2\mathrm{var}(f_{M_2})+(\alpha_3\lambda_{M_{i3}})^2\mathrm{var}(f_{M_3})}{\mathrm{var}(x_{\mathrm{sum3}})} \tag{11}$$

適用例では，産業能率大学の人材アセスメントツールの 1 つである，“ビジネス基礎力診断 S-BASE” による 360 度フィードバックのデータを用いた。特性は “自己確立”，“仕事確立”，“関係性確立” の 3 つであり，方法は “自己評価”，“上司評価”，“同僚評価” の 3 つであった。“同僚評価” は，1 人の被評価者に対して，1 人〜最大 6 人までの同僚による評定が行われており，分析においては，あらかじめ被評価者ごとに同僚の人数に応じて平均値を計算した。欠測値を除いた有効回答数は 401 となった。このデータに CT-UM モデルを適用し，その結果として得られる推定値を利用して，(9)，(10)，(11) 式に従い，評価者ごとに項目数を変化させて合計した再構成スコアの信頼性係数と妥当性に関する係数を算出した。

結果と考察　分析の結果，“自己評価” の項目数を増やすと信頼性と妥当性は低下し，一方で，“上司評価” と “同僚評価” の項目数を増やすことは信頼性と妥当性の向上に貢献することが示された。しかし，現実の 360 度フィードバック実施場面では，項目数を減らすことは容易でも，増やすことは困難な場合が多いだろう。評価者への負担，あるいは実施に要する時間やコスト等の制約により，現状よりも項目数を増やすことが難しい場合には，“自己評価” の項目数を減らすことによってもまた，信頼性と妥当性の改善が見込まれる。再構成スコアの信頼性係数と (非) 妥当性係数は，目標とする信頼性と妥当性を設定した上で，現状から項目数をどれくらい減らしてもその目標を達成できるか検討するためにも有用である。本研究の結果から，(9)，(10)，(11) 式を利用することで，測定全体の信頼性と妥当性という観点から，実現可能な範囲内で，評価者ごとに最適な項目数の配分を検討できる可能性が示唆された。

第4章　ベイズ統計学とマルコフ連鎖モンテカルロ法

第 4 章では，研究 III で提案するマルコフ連鎖モンテカルロ (Markov chain Monte Carlo method; MCMC) 法によるアプローチについて理解を深めるため，ベイズ統計学および MCMC 法とその推定アルゴリズムについて概説した。

第5章　研究 III: 同一立場内に複数の評価者がいる多特性多評価者データにおける信頼性と妥当性の検討

目的　MTMR データにおける，測定方法や測定機会を “方法” として扱う MTMM データとの大きな違いは，同一 “方法” 内で複数の測定値が得られる可能性があるという点である。例えば 360 度フィードバックでは，1 人の被評価者につき，複数の同僚，あるいは複数の部下から評定結果が得られる場合がある。さらに，得られる他者評価の数は，被評価者によって異なることが多い。このように複数人から他者評価が得られた MTMR データに対して，MTMM データのための確認的因子分析モデルを適用する際には，これまで平均値を観測変数として用いて MTMM 行列が計算されてきた。

本来，評価者の人数は測定の信頼性に直接的に貢献する要因であると考えられる。しかし，単純に平均値を計算し，それをそのまま分析に用いるという従来の方法では，評価者の人数に関する情報を分析結果に反映することができていない。また，評価者の人数という要因を無視して求められる分散成分は，本来得られるべき値とは異なったものになるだろう。このとき，分散成分の分解によって定義される信頼性係数と妥当性に関する係数もまた，誤ったものになることが懸念される。そこで研究 III の目的は，評価者の人数を適切に評価し，測定の信頼性と妥当性を検討するための方法を提案することとする。

方法　研究 III においても，研究 II と同様に，CT-UM モデルによる推定を行う。本研究では，まず，同一立場内に複数の評価者が存在する場合に，その平均値を用いて確認的因子分析モデルを実行するとき，評価者の人数に応じて分散成分がどのように変化するかを数学的に導出した。その結果，同一立場内に評価者が複数いる場合，その平均値における分散は

$$V[x^*_{ij}] = \lambda^2_{T_{ij}} V[f_{T_i}] + \lambda^2_{M_{ij}} V[f_{M_j}] + \frac{1}{A^{(n)}} V[e_{ij}] \tag{12}$$

となり，誤差分散が $1/A^{(n)}$ 倍となることが導かれた。ただし，n 番目のオブザベーションについて，i 番目の特性の j 番目の評価者の立場による測定値が $A^{(n)}(=1,\ldots,A)$ 人の評価者から得られているとする。

そこで，同一立場内の評価者の人数に応じて被評価者ごとに異なる誤差分散を CT-UM モデルに反映させるためのアプローチとして，MCMC 法を利用した

モデルの推定を提案する。なお，以下では，自己・上司・同僚の 3 つの立場からの評価を行い，そのうち同僚評価のみ複数の回答が得られている状況を考える。各パラメータの事前分布は以下のように設定した。

$$
\begin{aligned}
x_{nij} &\sim Multi\text{-}Normal(\mathbf{0}, \mathbf{\Lambda}_T \mathbf{\Phi}_T \mathbf{\Lambda}'_T + \mathbf{\Lambda}_M \mathbf{\Lambda}'_M + \mathbf{\Psi}_n) \\
\sigma_e^2 &\sim Inv\text{-}Gamma(0.001, 0.001), \quad \mathbf{\Phi}_T \sim \boldsymbol{U}(-1, 1) \\
\lambda_{T_{ij}} &\sim N(0, 10^6), \quad \lambda_{M_{ij}} \sim N(0, 10^6)
\end{aligned} \tag{13}
$$

ここで，(13) 式は誤差分散がオブザベーションごとに異なることを表しており，同僚評価に関する誤差分散に関して σ_e^2 を $A^{(n)}$ で除したものを $\mathbf{\Psi}_n$ の対角要素として用いる。

適用例では，研究 II に引き続き，“ビジネス基礎力診断 S-BASE” による 360 度フィードバックの結果を用いた。ビジネス基礎力診断データでは，同僚評価は 1 人の被評価者につき，1 人～6 人からの評定が得られていた。分析においては，変数ごとに平均による中心化を施した。MCMC 法により推定を行った結果をもとに，研究 II で定義した (5) 式から (7) 式を計算することで，信頼性および収束的妥当性，非弁別的妥当性について定量的な解釈を行う。なお，母数の推定においては，ハミルトニアンモンテカルロ (Hamiltonian Monte Carlo; HMC; Duane, Kennedy, Pendleton & Roweth, 1987, Neal, 1996) 法を実装した Stan を用いた。

また，同一立場内の評価者の人数に応じて異なる誤差分散を CT-UM モデルに反映させるためのもう 1 つの方法として，多母集団同時解析によるアプローチが考えられる。そのため，適用例においては，MCMC 法によるアプローチの有効性を確認することを目的として，多母集団同時解析モデルの適用結果との比較検討を行った。

結果と考察　適用例では，同一立場の評価者から複数の評定結果が得られた場合に，平均値を利用した相関行列に従来のモデルをそのまま適用すると，誤差分散が過小評価されることが明らかとなった。つまり，差分散の値を過小評価する従来の方法によって，測定の信頼性と妥当性を検討する場合には，誤った解釈につながる恐れがある。誤差分散の過小評価は，提案モデルの適用により改善される。したがって，MCMC 法による推定を適用することで，被評価者ごとに異なる他者評価の人数を統計的に適切に処理した上で，より正確な信頼性と妥当性を求めることができる。

多母集団同時解析でもまた，誤差分散の過小評価の問題を改善できる可能性が示唆された。しかしながら，多母集団同時解析を行うためには，事前に評価者の人数によって群分けを行う必要があり，もし構成員が 1 人しかいない群がある場合には，その群のオブザベーションはデータから除外するという対処をとらざるを得ない。一方で，MCMC 法によるアプローチでは，事前に群分けし

たデータを作成する必要もなく，特定の人数の評価者を有するオブザベーションが1人しかいない場合でも，そのデータを含めたすべてのデータを用いて推定を行うことができるというメリットがある。さらに，複数の評価者からの評定が得られている評価者の立場が2つ以上となった場合でも，MCMC法によるアプローチではデータを群分けすることなく，従来の方法とまったく同じデータセットを利用して分析を実行することが可能である。

第6章　総合考察

本論文では，確認的因子分析モデルの枠組みでMTMMデータを分析し，測定の信頼性および収束的妥当性，弁別的妥当性を定量的に検討するための方法について論じてきた。まず，研究Iでは，汎用的かつより実用的なMTMMデータのための確認的因子分析モデルの開発を目的として，方法因子の因子得点の和を0に制約することで，信頼性および妥当性の解釈が1つのデータに対して必ず一意に定まるモデルを提案した。先行研究によって提案されてきたこれまでのモデルは，識別不定や不適解に陥りやすいといった分析面での問題や，1つのデータに対して解が一意に定まらないという解釈面での問題から，実用性に欠けると考えられる。これに対して，提案モデルは識別の可能性も高く，分析結果を必ず一意に定めることができる。ただし，提案モデルの識別に関しては，シミュレーション研究による確認にとどまっており，必ず識別されることが証明されたわけではない。今後，提案モデルの有用性の証左を示すためにも，モデルが識別可能であることを数学的に証明する必要がある。

研究IIおよび研究IIIでは，現実場面でもっとも扱う機会が多いと考えられるMTMRデータの分析に焦点を絞り，議論を進めた。人事アセスメントの現場では，360度フィードバックの結果として，“方法”を評価者の違いとするMTMRデータの収集が盛んに行われている。360度フィードバックの結果として得られるMTMRデータを分析対象とした信頼性と妥当性の検討は，人事アセスメントの質を吟味するという意味で非常に重要であり，現場からの需要も高いと考えられる。

研究IIでは，立場の異なる複数の評価者からの評定結果を特性ごとに合計し，各被評価者の得点(特性値)を計算するという場面に着目した。この合計点の信頼性と妥当性という観点から，評価者ごとに適切な項目配分を決定する方法について提案した。合計得点の信頼性と妥当性という観点から得られる項目配分に関する示唆は，これまで蓄積されてきた人事アセスメントの知見に照らしても納得できる結果となった。しかしながら，実際に項目を増やしたり減らしたりする際に，どの項目を除くのか，あるいはどのような項目を加えるべきかという判断までを，本研究の範囲で行うことはできない。研究IIの方法によって得られた示唆に基づき，実際に現状の調査項目から項目数を変化させるために

は，内的整合性の観点からの項目分析や，項目プールの用意など，付加的な検討事項を要する。

最後に研究 III では，同一立場内に複数の評価者が存在する MTMR データの分析において，評価者の人数を適切に考慮し，信頼性と妥当性を検討するための方法を提案した。推定に MCMC 法を用いることで，評価者の人数に応じて被評価者ごとに異なる測定誤差の推定を可能にした。HMC 法による MCMC 法を用いた推定は，Stan を利用して容易に実行することができる。標本サイズが小さくても安定的な推定が可能であり，複雑なモデルも柔軟に表現することができる MCMC 法による推定は，今後広く利用されることが期待される。また，本研究により，誤差分散が正しく推定されることで，分散成分の分解によって定義される信頼性係数および，収束的妥当性と弁別的妥当性に関する係数もまた，評価者の数の影響を反映したより正確な値となるはずである。したがって，これらの係数を利用することで，測定の信頼性と妥当性に関してより精緻な解釈を行うことが可能となる。

Campbell & Fiske (1959) を嚆矢として，心理学的測定においても弁別的妥当性と収束的妥当性を検討するために MTMM 行列が利用されるようになり，現在，その分析には確認的因子分析モデルが広く用いられている。しかしながらこれまで，モデルの適用から，信頼性と妥当性の解釈までを手続き化して示した先行研究はほとんどなく，確認的因子分析モデルの推定結果を用いて定量的に信頼性と妥当性に関する係数を定義する試みは Eid (2000) が初めてであろう。本論文では，3 つの研究を通して，モデルの適用から，信頼性と妥当性を解釈するための係数の算出および，それらの解釈までを一貫して行った。MTMM データを用いて測定の信頼性と収束的妥当性，弁別的妥当性を検討するための手続きを体系化して示したことで，本研究の成果が人事アセスメントの領域ばかりではなく，心理学の研究においても実際に活用される日が来ることを願う。

引用文献

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-mutimethod matrix. ıPsychological Bulletin, **56**, 81-105.

Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, **22**, 139-162.

Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, **195**(2), 216-222.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, **65**(2), 241-261.

Grayson, D., & Marsh, H. W. (1994). Identification with deficient rank loading matrices in confirmatory factor analysis: multitrait-mutimethod matrix. *Psychometrika*, **59**, 121-134.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, **12**, 247-252.

Kenny, D. A. (1994). The multitrait-multimethod matrix: Design, analysis and Conceptual Issues. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, mehods, and Theory* pp. 111-124. Hillsdale, NJ: Lawrence Erlbaum.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, **112**(1), 165-172.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: many problems and a few solutions. *Applied Psychological Measurement*, **13**(4), 335-361.

Marsh, H. W., & Grayson, D. (1995). Latant variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issue, and applications*. Thousands Oals, CA: Sage. pp.177-198.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Amertican Psychologist*, **50**, 741-749.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics 118, NY: Springer–Verlag.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, **9**(1), 1-526.