

多面観察評価における信頼性係数と妥当性係数の導出 -評定者の違いを考慮した項目数決定のために-†

Coefficients of Reliability and of Convergent and Discriminant Validity for Multisource Assessment:
To Determine the Number of Items Taking into Account Differences between Rating Source Groups

久保 沙織*¹ 豊田 秀樹*² 福中 公輔*³

Abstract

Reliability and validity of measurement influence the availability of human resource assessment. In multisource assessment, because it is not always possible to ensure sufficient number of raters from each rating source groups, defining the relationship between the number of items for each source and the reliability and the validity of the rating has great significance. In order to determine the appropriate number of items to provide to each rating source group, it would be important to take into account the contribution of each source to reliability and validity of measurement.

This article presents a multitrait-multimethod confirmatory factor analytic approach to multisource assessment data and develops the expressions for coefficients of reliability and of convergent and discriminant validity of the sum score of multiple sources for a trait. These coefficients are calculated in actual multisource assessment data, changing the number of items for each rating source groups, and the practicality of this approach is demonstrated. The result of the analysis showed that increasing the items for supervisor or peer ratings contributed to the improvement of reliability and the convergent and discriminant validity of sum score.

Key words:

Multisource assessment, 360-degree feedback, Reliability, Convergent validity, Discriminant validity

1. はじめに

職場での具体的な行動傾向を把握する方法に、多面観察評価手法がある（田中，2011）。多面観察評価は360度評価、あるいは360度フィードバック（360-degree feedback）とも呼ばれ（Edwards & Ewen, 1996; Lepsinger & Lucia, 2009）、一人の被評定者に対して上司や部下、同僚など周囲の人たちから仕事ぶりについてフィードバックしてもらい、本人の気付きを促し、人材育成や組織の活性化に役立てようというものである（相原・南雲，

2009）。

日本では、人事考課の諸問題を克服するため1965年に関本昌秀が導入して以来、多面観察評価手法はいくつかの企業で評価と育成を目的として独自に実施されてきた（馬場・馬場，2005）。近年では、大企業を中心に導入する企業も増え、リクルートマネジメントソリューションズによるMOA、日本能率協会マネジメントセンターによるNAVI360やRoundReview、日本経営協会総合研究所による多面観察＜マネジメント版・リーダーシップ版＞、そして産業能率大学総合研究所

論文受理日：2014.2.6 論文受稿日：2013.7.12

† 本稿は、早稲田大学特定課題研究助成費（課題番号 2013C - 619）による研究成果の一部である。

*¹KUBO, Saori：早稲田大学グローバルエデュケーションセンター
(Global Education Center, Waseda University)

*²TOYODA, Hideki：早稲田大学文学学術院
(School of Humanities and Social Sciences, Waseda University)

*³FUKUNAKA, Kosuke：産業能率大学総合研究所
(Research and Assessment Center, The SANNO Institute of Management)

によるビジネス基礎力診断 S-BASE、プロ人材特性診断 S-ProII、マネージャー力診断 S-ATOM など、若手社員から中間管理職、経営者レベルまで様々な立場を対象とした多面観察評価ツールが数多く開発され、各社から提供されている。

一般に、人事アセスメントにおいては測定の精度と正確さ、すなわち信頼性と妥当性が、アセスメントの有効性を左右することになる(大沢, 2000)。人事アセスメントツールの1つである多面観察評価の実施においても、このことは例外ではない。多面観察評価の信頼性に関する日本国内の研究としては、入江・鷺坂・舛田・二村(2004, 2005)が挙げられる。入江ら(2004)では、一般化可能性理論(Cronbach, Gleser, Nanda & Rajaratnam, 1972; Brennan, 1992; Shavelson & Webb, 1991)を用いることでMOA(M型)の分析を行い、対象者、評定者、項目の各要因が測定の信頼性に与える影響の大きさを推定した。信頼性向上のためには、項目数よりも評定者数を増加させることが有効であるというその結果を受けて、入江ら(2005)は項目の相をモデルから除外し、多変量一般化可能性理論(Brennan, 2001)の適用により、上司・同僚・部下のそれぞれで、評定者数の変化による一般化可能性の変動を確認している。

入江ら(2004)では、項目数が16という条件の下で0.7以上の信頼性係数を実現するためには上司3~6名、同僚7~8名、部下4~6名の評定者を確保することが望ましいと結論付けていた。しかしながら、現実場面では複数の上司や同僚による評定が確保できるとは限らない。通常、上司からの評価は第一考課者としての直属管理者一名のみによる。得られる同僚評定の数はその部署の規模に応じて決まるため、小さな部署では同僚が2名以上いないということがあり得る。入江ら(2004)で、上司、同僚、部下いずれにおいても「評定者×項目、対象者×評定者×項目、その他誤差」に起因する分散成分の割合がもっとも大きくなっていった点からも、複数の評定者が確保できない場合を想定して、各評定者に回答してもらう項目数と、測定の信頼性および妥当性との関連を検討することの重要性が示唆されている。

では、どのような観点から評定者ごとの項目数の配分を決定すれば、より有効なアセスメントを実施できるだろうか。評定者の立場の違いが、測定の信頼性と妥当性に与える影響について着目する。例えば、Conway & Huffcutt(1997)が過去の文献から収集した177の事例をもとにメタ分析を行った結果では、部下評定や同僚評定と比較して、上司評定の信頼性は非常に高いことが示されている。Greguras & Robie(1998)でも、人数と項目数が評定者間で等しいという条件のもとでは、同僚評定や部下評定と比較して上司評定の信頼性が高いと述べられている。一方で、評定者の立場によって重要視する特性が異なるため、上司評定、同僚評定、部下評定はそれぞれ独自の情報を提供し、妥当性の向上に寄与しているという結果がConway, Lombardo, & Sanders(2001)によって明らかになっている。したがって、各評定者に回答してもらう項目数の配分を決定する際に、自己、上司、部下といった評定者の立場の違いによって、特性ごとに測定の精度と正確さへの寄与が異なるという人事アセスメントの視点を活かすことで、より精緻な信頼性と妥当性の検証が可能になると考えられる。

本研究では、一般化可能性理論とは別の枠組みから、多面観察評価データの信頼性と妥当性へのアプローチを試みる。多面観察評価は、英語では multisource rating、あるいは multirater assessment と表現され、得られたデータは多特性多方法(multitrait-multimethod; MTMM)データとして扱うことができる。MTMMデータとは、複数の構成概念(特性)を複数の方法によって測定したときに得られるデータのことであり、多面観察評価では複数の立場の評定者を通して評価を行う点で、評定者が測定方法に相当する。本論文では、多面観察評価により得られたデータをMTMMデータとして扱い、確認的因子分析モデルを適用することで、信頼性のみならず、一般化可能性理論の枠組みでは焦点を当てることのできなかつた収束的妥当性と弁別的妥当性も同時に検証する方法を提案する。

多面観察評価データをMTMMデータとして扱うアプローチの歴史は古く、Lawler(1967)では、

多特性多方法行列 (Cambell & Fiske, 1959) と呼ばれる相関行列を計算し、信頼性と収束的妥当性、弁別的妥当性の検討が行われている。Conway (1996) は、多面観察評価データから計算される多特性多方法行列を多特性多評定者行列 (multitrait-multiratermatrix; MTMR 行列) と呼び、20の MTMR 行列を分析対象として3種類のモデルを適用した結果から、CT-CU (correlated trait-correlated uniqueness) モデル¹⁾が、多面観察評価データの分析にはもっとも適していると述べている。また、Mount, Judge, Scullen, Sytsma, & Hezlett (1998) は、2350人の管理者を対象として実施された多面観察評価の結果から得られた MTMR データを確認的因子分析によって分析し、3つの特性因子と7つの評定者因子 (自己評価、上司評価1、上司評価2、同僚評価1、同僚評価2、部下評価1、部下評価2) を仮定したモデルが最も当てはまりがよいことを示した。これらアメリカの先行研究の結果を受けて、日本における多面観察評価から得られたデータをもとに MTMR 行列を構成し、確認的因子分析を適用して収束的妥当性および弁別的妥当性を検証した例として、高橋 (2010) が挙げられる。

先行研究で報告されている信頼性や妥当性についての考察は、特性と方法との組み合わせで規定される変数間で計算される相関行列の利用が主流であり、信頼性係数や収束的・弁別的妥当性に関する係数もまた、変数ごとに算出される。ところが、多面観察評価データの分析では、各特性について異なる評定者からの評価を合計して、被評定者ごとに得点 (特性値) を計算することとなる。そこで本研究では、評定者によってテストの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第一の目的とする。さらに、第二の目的は、これらの係数を実データに対して計算し、評定者ごとの項目数の配分と、測定の信頼性、収束的・弁別的妥当性との関係について考察することとする。なお、分析には産業能率大学総合研究所による“ビジネス基礎力診断 S-BASE”の実施により得られた多面観察評価データを用いた。

2. 方法

2.1 確認的因子分析モデル

t 個の特性と m 個の方法を用いた MTMM データで、 $tm \times 1$ の観測変数ベクトル x に対する確認的因子分析のモデル式は

$$x = \Lambda_T f_T + \Lambda_M f_M + e \quad (1)$$

と表される。 f_T は特性因子、 f_M は方法因子をそれぞれ縦に並べたベクトル、 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 Σ は、以下となる。

$$\Sigma = \Lambda_T \Phi_T \Lambda_T' + \Lambda_M \Phi_M \Lambda_M' + \Theta \quad (2)$$

ただし、 Λ_T と Λ_M はそれぞれ特性因子と方法因子の因子負荷行列であり、以下のように構成される。

$$\Lambda_T = \begin{bmatrix} \Lambda_{T1} \\ \vdots \\ \Lambda_{Tj} \\ \vdots \\ \Lambda_{Tm} \end{bmatrix}, \Lambda_M = \begin{bmatrix} \lambda_{M1} & 0 & 0 & \dots & 0 \\ 0 & \lambda_{M2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_{Mm} \end{bmatrix} \quad (3)$$

Λ_T は $t \times t$ の対角行列 $\Lambda_T = \text{diag}(\lambda_{Ti})$ で、 λ_{Ti} は j 番目の方法で測定された i 番目の特性の観測変数が i 番目の特性因子 f_{Ti} から受ける影響の強さである。一方 λ_{Mj} は、 i 番目の特性を j 番目の方法で測定した観測変数が j 番目の方法因子 f_{Mj} から受ける影響の強さ λ_{Mij} を要素としてもつ $t \times 1$ ベクトルである。 $t \times t$ 行列 Φ_T と $m \times m$ 行列 Φ_M は、それぞれ特性因子間、方法因子間の相関行列であり、 $E[ee'] = \Theta$ は対角要素に誤差分散を配した $tm \times tm$ 対角行列である。

本論文では、方法因子と特性因子とは無相関とし、特性因子間には互いに相関を仮定するが、方法因子間は無相関とする CT-UM モデル (correlated trait uncorrelated method model; Grayson & Marsh, 1994; Marsh & Grayson, 1995) を採用する。すなわち、 Φ_M の対角要素のみ自由母数とし、非対角要素はすべて0に固定する。因子間相関に対する上述の制約は、評定者の立場の違いを意味する方法因子を系統誤差として扱うという仮定を表現したものであり、評定者ごとの誤差は、同じ立場に属

する評定者による測定にのみ寄与し、異なる立場の評定者による測定には寄与しないことを表している (McDonald,1999)。この仮定により、古典的テスト理論の手法を応用して、分散成分の分解による信頼性係数や収束的・弁別的妥当性の指標となる係数を定式化することが可能となる。

2.2 信頼性係数および妥当性に関する係数

以下では、Eid (2000) に基づき、古典的テスト理論における分散成分の分解を利用して信頼性係数および収束的妥当性係数、非弁別的妥当性係数を表現し、それらの解釈方法について述べる。まず、特性因子と方法因子が無相関であるという仮定により、(1) 式から観測変数の分散は

$$V[x_{ij}] = \lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + V[e_{ij}] \quad (4)$$

と導かれる。なお、 $V[\]$ によって分散を表す。(4) 式右辺の各項は順に、特性因子の影響による変動、方法因子の影響による変動、そしての誤差による変動を表している。

測定の信頼性は、一般的に、測定値の分散のうち真値の割合、すなわち誤差以外の部分であり、因子分析モデルの文脈では、観測変数全体の分散に占める因子の分散として表現することができる。

$$\begin{aligned} \text{Rel}(x_{ij}) &= \frac{\lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}]}{\lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + V[e_{ij}]} \\ &= \frac{\text{真値の分散}}{\text{測定値の分散}} \end{aligned} \quad (5)$$

ただし、何度測定しても安定して出てくる傾向である系統誤差は、真値に含まれてしまっている (南風原, 2011)。そこでさらに、(5) 式の分子を、純粋に特性を反映する部分と、評定者の立場の違いによる系統誤差とに分離する。このとき、

$$\text{Con}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 V[f_{T_i}]}{\lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + V[e_{ij}]} \quad (6)$$

によって、収束的妥当性を検討することが可能である。収束的妥当性は、測定方法が異なっても同一特性の測定間で相関が高いことによって示される。(6) 式の収束的妥当性係数は、観測変数全体

の変動に対する特性の因子負荷および分散の相対的な大きさを表しているため、その値が大きいほど、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は、

$$\text{Dis}(x_{ij}) = \frac{\lambda_{M_{ij}}^2 V[f_{M_j}]}{\lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + V[e_{ij}]} \quad (7)$$

と表される。MTMM データにおいては、測定方法が同じであっても、異なる構成概念を測定していれば、同一構成概念を測定した変数同士の場合と比較して相関が低いことで弁別的妥当性の証拠となる。(7) 式は、測定値全体の分散に占める系統誤差の影響を表現している。もし系統誤差が大きければ方法の違いが測定値に与える一定の影響が大きいかを意味し、本当に測りたい構成概念と交絡して、特性が違うのに同一方法間での相関が高くなる可能性がある。したがって、系統誤差が小さいと、方法の影響には左右されずに特性の違いが弁別できている状態であり、(7) 式で算出される値は小さいほど弁別的妥当性が高いと解釈できる。反対に、(7) 式の値が大きい場合には、特性による変動よりも方法の違いによる変動の方が相対的に大きいことを意味し、弁別的妥当性が低いと結論づけられる。

2.3 和得点の信頼性と妥当性

いま、多面観察評価の実施により、自己評定、上司評定、同僚評定という3つの異なる方法による評定結果が得られているとする。評定者の立場の違いによって測定全体の信頼性と妥当性に与える影響が異なることを考慮して、評定者ごとにテストの項目数を変化させた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の一般式を導出したい。そこで、自己、上司、同僚に回答してもらう項目数をそれぞれ現状から α_1 倍、 α_2 倍、 α_3 倍して、各評定結果を足し上げた和得点について考える。ただし、自己、上司、同僚のそれぞれで項目数の増減により下位テストの長さが変わっても、下位テスト内では変更前のテスト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ものと仮定する。

評定者の異なる3つの下位テストを、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

$$\begin{aligned} V[x_{sum3}] &= V[\alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3}] \\ &= (\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 V[f_T] \\ &\quad + (\alpha_1 \lambda_{M_{i1}})^2 V[f_{M_1}] + (\alpha_2 \lambda_{M_{i2}})^2 V[f_{M_2}] + (\alpha_3 \lambda_{M_{i3}})^2 V[f_{M_3}] \\ &\quad + \alpha_1 V[e_{i1}] + \alpha_2 V[e_{i2}] + \alpha_3 V[e_{i3}] \end{aligned} \quad (8)$$

となる。 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、(6)式、(7)式を再表現すると、再構成スコアの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。

$$\begin{aligned} Rel(x_{sum3}) &= \frac{(\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 V[f_T]}{(\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 V[f_T] + (\alpha_1 \lambda_{M_{i1}})^2 V[f_{M_1}] \\ &\quad + (\alpha_2 \lambda_{M_{i2}})^2 V[f_{M_2}] + (\alpha_3 \lambda_{M_{i3}})^2 V[f_{M_3}] + \alpha_1 V[e_{i1}] + \alpha_2 V[e_{i2}] + \alpha_3 V[e_{i3}]} \end{aligned} \quad (9)$$

$$Con(x_{sum3}) = \frac{(\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 V[f_T]}{V[x_{sum3}]} \quad (10)$$

$$Dis(x_{sum3}) = \frac{(\alpha_1 \lambda_{M_{i1}})^2 V[f_{M_1}] + (\alpha_2 \lambda_{M_{i2}})^2 V[f_{M_2}] + (\alpha_3 \lambda_{M_{i3}})^2 V[f_{M_3}]}{V[x_{sum3}]} \quad (11)$$

なお、(8)式から(11)式までの方法に関する添字について、1は自己評定、2は上司評定、3は同僚評定に対応している。

3. 適用例

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”による多面観察評価データを用いて、適用例を示す。“ビジネス基礎力診断 S-BASE”の実施概要は以下の通りであった。

内容 ビジネス基礎力診断では、企業における若手社員から中堅社員を対象者として、以下の3つの領域でビジネスに必要な基礎的能力²⁾を設定し、測定している(学校法人産業能率大学総合研究所, 2012)。各領域の内容と、該当する項目数を以下に示した。

- ・“自己確立”領域：自分自身の確固とした想いや考えを持って、常に成長を目指す力。(34項目)
- ・“仕事確立”領域：仕事のサイクルに従って堅実に業務を進め、目標を達成する力(33項目)。
- ・“関係性確立”領域：影響力を発揮し、周囲と

協働を図りながら、組織としての成長を出す力。(31項目)

調査 内容および個人情報秘匿されることを説明した上で、計98項目から構成される調査票(質問紙)を用いて、“自己評定”、“上司評定”、“同僚評定”による回答を求めた。“上司評定”は各被評定者に対して上司一人による評定であり、“同僚評定”は、一人の被評定者に対して、1人～最大6人までの同僚による評定が行われた。

質問項目への反応は“あてはまらない”、“ややあてはまらない”、“ややあてはまる”、“あてはまる”の4件法とした。また他者(上司・同僚)評定調査票の項目は、自己評定調査票と同一内容としたが、文章表現を“その人は～”で始まる形式にした。

期間 2010年12月から2011年2月までの3ヶ月間。

回収方法 調査票の受け渡しは郵送で行った。

対象 日本に本社を置く31社の企業。企業の所在地は関東地方21社、中部地方4社、近畿地方3社、中国地方3社であり、業種は農林、住宅、食品、紙・パルプ、化学・薬品、機械・電気、精密、商社、小売・飲食、運輸、サービス、教育、医療など多岐にわたった。また、調査対象者は入社2年目から10年目までとし、自己評定については855人(男性641人、女性211人、不明3人)、“上司評定”については446人、同僚評定については1369人の回答を得た。

分析にあたって、領域ごとに自己評定、上司評定、同僚評定の合計点を算出した。この際、同僚評定については、あらかじめ被評定者ごとに同僚の人数に応じて平均値を計算した。その結果、すべての領域で欠損値のなかった有効回答数は403(平均25.8歳、SD3.55)となった。

4. 結果と考察

ビジネス基礎力診断データの相関行列を表1に示した。特性の異同に関わらず同一評定者内で相関が高い傾向があり、収束的妥当性および弁別的妥当性の十分な証拠は得られなかった。しかし、自己評定と他者評定との相関は総じて低いこと、自己評定に比較して他者評定の方が妥当性が高い

表1 ビジネス基礎力診断データの相関行列

	自己評定			上司評定			同僚評定			
	自己	仕事	関係性	自己	仕事	関係性	自己	仕事	関係性	
自己評定	自己	1.00								
	仕事	0.79	1.00							
	関係性	0.72	0.73	1.00						
上司評定	自己	0.27	0.26	0.22	1.00					
	仕事	0.21	0.29	0.19	0.89	1.00				
	関係性	0.21	0.24	0.28	0.86	0.84	1.00			
同僚評定	自己	0.28	0.22	0.20	0.47	0.46	0.43	1.00		
	仕事	0.27	0.28	0.24	0.46	0.51	0.46	0.92	1.00	
	関係性	0.22	0.18	0.26	0.42	0.41	0.47	0.88	0.86	1.00

傾向にあることなど、二村 (2001) で紹介されている多面観察評価データの一般的な特徴と一致している。

前節で述べたように、“ビジネス基礎力診断 S-BASE” では自己、上司、同僚のすべての評定を通して、特性ごとに決まった項目数で実施されていた。以下では、(9)、(10)、(11) 式を用いて、他者評定項目の項目数と、再構成スコアの信頼性・妥当性との関係を考察する。

まず、特性因子間にのみ相関を仮定した CT-UM モデルを適用し、確認的因子分析を実行した結果、適合度指標の値は、CFI = 0.994, TLI = 0.984, RMSEA = 0.063であった ($\chi^2 = 39.111, df = 15$)。推定された CT-UM モデルのパラメータの非標準化解は、表2の通りである³⁾。表2には、推定された因子負荷および因子の分散、誤差分散

の値を (5) 式から (7) 式に当てはめて、変数ごとに信頼性係数と収束的妥当性係数、非弁別的妥当性係数を算出した結果も併せて示した。表1の相関係数の観察結果と同様に、自己評定に比較して、上司評定と同僚評定では信頼性、収束的・弁別的妥当性ともに高いことが確認できた。

4.1 係数の算出

表2の推定値を用いて、(9)、(10)、(11) 式に従って、評定者ごとに項目数を変化させ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。“自己評定”、“上司評定”、“同僚評定”それぞれの項目配分 $\{\alpha_1, \alpha_2, \alpha_3\}$ を変化させたときの、再構成スコアの信頼性係数 $Rel(x_{sum3})$ および収束的妥当性係数 $Con(x_{sum3})$ 、非弁別的妥当性係数 $Dis(x_{sum3})$ の変動を表3に示した。

表2 CT-UM モデルによる推定結果 (非標準化解) と信頼性係数および妥当性係数

変数	λ_T	λ_M	誤差分散	Rel	Con	Dis
x_{11}	1.000	1.000	34.627	0.813	0.252	0.561
x_{21}	1.000	0.934	23.147	0.850	0.261	0.589
x_{31}	1.000	0.839	36.510	0.752	0.254	0.498
x_{12}	1.332	1.000	12.496	0.936	0.422	0.515
x_{22}	1.529	1.016	21.224	0.903	0.428	0.475
x_{32}	1.101	0.865	12.509	0.895	0.382	0.513
x_{13}	1.128	1.000	4.400	0.970	0.409	0.561
x_{23}	1.376	0.950	9.627	0.940	0.479	0.461
x_{33}	1.462	0.862	18.305	0.894	0.461	0.433
特性因子間の分散共分散行列			方法因子間の分散共分散行列			
	46.702		<i>sym.</i>	103.929	0	0
	39.709	40.236		0	101.166	0
	36.087	33.265	37.392	0	0	81.372

配分が $\{\alpha_1: \alpha_2: \alpha_3\} = \{1:1:1\}$ のときは、項目数は現状維持で手元のデータにおいて特性ごとの単純な和得点を考えたときの信頼性と妥当性を意味するため、ベースラインとして参照することができる。 $\{\alpha_1: \alpha_2: \alpha_3\} = \{1:1:1\}$ のとき、3つの特性ともに信頼性係数は0.9以上と十分に大きな値であり、妥当性係数の値も特性間でほとんど違いはなかった。

S-BASE は、“自己確立” 34項目、“仕事確立” 33項目、“関係性確立” 31項目の計98項目から構成されており、回答者への負担が大きくないとは言いがたい現状である。多面評価を現場で実施する場合には、実施時間の制約もあるだろう。そこで、実施時間を短縮せざるを得ない場合に、 $\{\alpha_1: \alpha_2: \alpha_3\}$ を変化させ、係数を計算することで、自己、上司、同僚のいずれの項目をどれくらい減らすと信頼性や妥当性がどの程度低下するのか、あるいは低下しないのかを事前に予測できることは、実践的意義が大きいと考えられる。

以下では、回答者への負担軽減および実施時間短縮を目的とする実践場面を想定して、(1) 項目数の合計がテスト全体で変わらない場合、と (2) 項目数を現状維持または減らす場合、の2パターンに分けて、 $\{\alpha_1: \alpha_2: \alpha_3\}$ を変化させ、再構成スコアの信頼性と妥当性がどのように変化するか観察する。項目配分 α を検討することで、3つの評定者の立場による測定全体（合計点）としての信頼性や妥当性をできるだけ高く維持した状態で、効率よく項目数を減らすためには、どの立場からの評定をより重視すべきかという示唆を得ることができる。

(1) 項目数の合計がテスト全体で変わらない場合

表3の一段目と二段目はいずれも、3つの方法の重みの和が $\alpha_1 + \alpha_2 + \alpha_3 = 3$ となるような α の組み合わせである。これは、評定者の立場ごとにテストの項目数を変化させるけれども、全体としては、項目数の合計が変化しない状況を意味する。

$\alpha_1: \alpha_2: \alpha_3 = \{1:1.5:0.5\}$ と $\alpha_1: \alpha_2: \alpha_3 = \{1:$

表3 重みの違いによる係数の変動

$\alpha_1:\alpha_2:\alpha_3$	1:1:1	1:1.5:0.5	1:0.5:1.5	1.5:1:0.5	1.5:0.5:1	
特性1の Rel(x_{sum3})	0.943	0.944	0.946	0.931	0.931	
特性1の Con(x_{sum3})	0.623	0.593	0.594	0.561	0.557	
特性1の Dis(x_{sum3})	0.319	0.352	0.352	0.370	0.374	
特性2の Rel(x_{sum3})	0.942	0.943	0.948	0.936	0.938	
特性2の Con(x_{sum3})	0.656	0.612	0.641	0.589	0.598	
特性2の Dis(x_{sum3})	0.287	0.330	0.306	0.346	0.340	
特性3の Rel(x_{sum3})	0.910	0.918	0.911	0.900	0.896	
特性3の Con(x_{sum3})	0.632	0.615	0.593	0.580	0.564	
特性3の Dis(x_{sum3})	0.279	0.302	0.318	0.321	0.332	
$\alpha_1:\alpha_2:\alpha_3$	0.5:1:1.5	0.5:1.5:1	1.25:1.25:0.5	1.25:0.5:1.25	0.5:1.25:1.25	
特性1の Rel(x_{sum3})	0.961	0.959	0.937	0.938	0.959	
特性1の Con(x_{sum3})	0.626	0.621	0.584	0.583	0.631	
特性1の Dis(x_{sum3})	0.335	0.339	0.352	0.355	0.329	
特性2の Rel(x_{sum3})	0.953	0.951	0.938	0.942	0.952	
特性2の Con(x_{sum3})	0.665	0.646	0.609	0.627	0.662	
特性2の Dis(x_{sum3})	0.288	0.305	0.330	0.315	0.290	
特性3の Rel(x_{sum3})	0.929	0.931	0.908	0.902	0.929	
特性3の Con(x_{sum3})	0.631	0.637	0.605	0.585	0.641	
特性3の Dis(x_{sum3})	0.298	0.294	0.304	0.317	0.288	
$\alpha_1:\alpha_2:\alpha_3$	0.5:1:1	1:0.5:1	1:1:0.5	1:0.5:0.5	0.5:1:0.5	0.5:0.5:1
特性1の Rel(x_{sum3})	0.948	0.927	0.926	0.899	0.928	0.931
特性1の Con(x_{sum3})	0.628	0.588	0.588	0.547	0.599	0.605
特性1の Dis(x_{sum3})	0.320	0.339	0.338	0.352	0.329	0.326
特性2の Rel(x_{sum3})	0.940	0.931	0.928	0.907	0.922	0.927
特性2の Cin(x_{sum3})	0.658	0.630	0.613	0.582	0.619	0.646
特性2の Dis(x_{sum3})	0.283	0.302	0.314	0.325	0.303	0.281
特性3の Rel(x_{sum3})	0.911	0.886	0.892	0.853	0.889	0.883
特性3の Con(x_{sum3})	0.633	0.587	0.602	0.550	0.609	0.596
特性3の Dis(x_{sum3})	0.279	0.299	0.290	0.303	0.280	0.287

0.5:1.5)の結果から、自己評定の項目数は現状のままでも、“上司評定”か“同僚評定”のいずれかを減らすと、ベースラインに比較して信頼性係数の値にはあまり変化がないものの、妥当性係数に与える影響は大きいことがわかる。

$\alpha_1 : \alpha_2 : \alpha_3 = \{1.5 : 1 : 0.5\}$, $\{1.5 : 0.5 : 1\}$, $\{1.25 : 1.25 : 0.5\}$, $\{1.25 : 0.5 : 1.25\}$ のように、自己評定の項目数を増やすと、妥当性ばかりでなく信頼性係数の低下にもつながる。反対に、 $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1 : 1.5\}$, $\{0.5 : 1.5 : 1\}$, $\{0.5 : 1.25 : 1.25\}$ のように“自己評定”の項目数を減らして、“上司評定”と“同僚評定”の一方、あるいは両方の項目数を増やすと、信頼性、妥当性ともにベースラインよりも向上した。

また、相対的に信頼性と妥当性に対する項目数 $\{\alpha_1 : \alpha_2 : \alpha_3\}$ の変化の影響をもっとも受けにくいのは特性2“仕事確立”であった。これに対して、“自己確立”と“関係性確立”の信頼性や妥当性は、評定者ごとの項目数の変化に敏感である。“ビジネス基礎力診断 S-BASE”で測定の対象としている3つの特性のなかでは、“仕事確立”がもっとも客観的な判断がしやすい領域であることから、項目数の増減に左右されず、評定者の立場が違って安定かつ正確な測定がなされやすかったのかもしれない。

項目数の合計がテスト全体で変わらない場合という条件の下で、最適な項目数の配分について検討すると、信頼性係数をもっとも大きくなるのは $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{0.5 : 1 : 1.5\}$ の場合である。ただし、収束的妥当性係数および非弁別的妥当性係数の値も含め総合的に判断すると、信頼性と妥当性の両方が改善されるのは $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1.25 : 1.25\}$ の場合であることがわかる。

(2) 項目数を現状維持または減らす場合

表3の最下段は、 α の値をいずれも1あるいは0.5とし、項目数を現状維持、あるいは減らす場合のみを考えている。これは、回答時間や回答者の負担、同一特性を測定すると見なせる項目数が限られているといった理由から、現状より項目を増やすことが難しい状況を想定している。

この場合でも、“上司評定”と“同僚評定”のいずれかでも項目数を少なくすると、信頼性、妥

当性ともにベースラインよりも低くなることがわかる。一方で、“自己評定”のみ項目数を半分にして、 $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1 : 1\}$ とした場合には、ベースラインに比較して信頼性と妥当性がわずかに改善する傾向が見られた。

(1) で見た $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{1.5 : 0.5 : 1\}$ と(2)の $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{0.5 : 0.5 : 1\}$ を比較すると、他者評価の項目配分は両者で同じであるにもかかわらず、“自己評定”の項目数を半分にした(2)の場合の方が、収束的妥当性と弁別的妥当性が向上している。

(1)、(2)両方の場合を通して、“自己評定”は測定の信頼性と妥当性の維持および改善にはあまり貢献しないことが示唆された。(2)の結果より、項目数を増やすことができない場合には、“自己評定”の項目数を減らすことで、測定の信頼性と妥当性を考慮しつつ、効率的な多面観察評価の実施が可能となる。

一般に、自己評定の項目を減らすと、本人の自己認識が不確かになってしまい、行動変容が起きにくくなるという懸念も考え得る。しかし、自己評定を減らしても、測定の信頼性・妥当性が低下しないということは、測定したい特性に関して、他者評価の方がより正確かつ安定的に測定できているということを意味する。これは、自己評定の項目数を減らしても情報損失は最小限に抑えることができると保証されているということである。したがって、自己評定の項目数を減らしても、他者からのフィードバックが充実していれば、その結果を考察することで自己の気づきを促し、行動変容に結びつけることができると考えられる。

ただし、実際に評定者間で回答してもらう項目数を変えて多面観察評価を実施する場合には、項目選定および被評定者に結果をフィードバックする際に工夫が必要となる。例えば、(1)の結果に従い、自己評定の項目数を減らして上司評定と同僚評定の項目を増やしたとき、自己評定に使用した項目がすべての立場で共通した項目となる。自己評定の項目を減らすとき、項目内容を十分に吟味した上で、経験的に重要と考えられる項目を残すように、かつ下位領域をまんべんなく網羅するように選定できていれば、それらの共通項目のみ

を利用することによっても有意義なフィードバックが可能となるであろう。

4.2 係数ごとの特徴

以下では、 α_1 の値を固定した上で、 α_2 と α_3 をそれぞれ0.25から2.0までの範囲で変化させたときの各係数の変動について考察する。

特性3の信頼性に関する考察 まず、ベースラインでもっとも信頼性係数が低かった特性3“人間関係確立”に注目して、信頼性係数 $Rel(x_{sum3})$ の変動を等高線によって表したのが、図1の上段である。左から順に、 $\alpha_1=0.5, 1.0, 1.5$ の場合を示している。これら3つの図から、信頼性の向上に対する“上司評定”と“同僚評定”の影響力はほぼ同じであり、いずれも項目数を現状から2倍にすることで信頼性係数を0.94以上まで改善できる

ことがわかった。

特性1の収束的妥当性に関する考察 次に、ベースライン時にもっとも $Con(x_{sum3})$ の値が小さかった特性1“自己確立”について、重みの変化に伴う収束的妥当性係数の変動を考察する。図1の中段では、それぞれ $\alpha_1=0.5, 1.0, 1.5$ としたときの和得点の収束的妥当性係数を、左側から順に図示した。左側から右側へ“自己評定”の重み α_1 を大きくするにしたがって、等高線の間隔は広く、カーブは緩やかになっている。

図1の中段左図より、 $\alpha_1=0.5$ と“自己評定”の項目を現状より減らした場合でも、“上司評定”と“同僚評定”が $\alpha_2=\alpha_3=1$ と現状維持であれば、収束的妥当性係数がベースラインの0.623を下回ることはないことがわかる。それどころか、 $\alpha_1=0.5$ としても α_2 と α_3 をわずかに増やして、そ

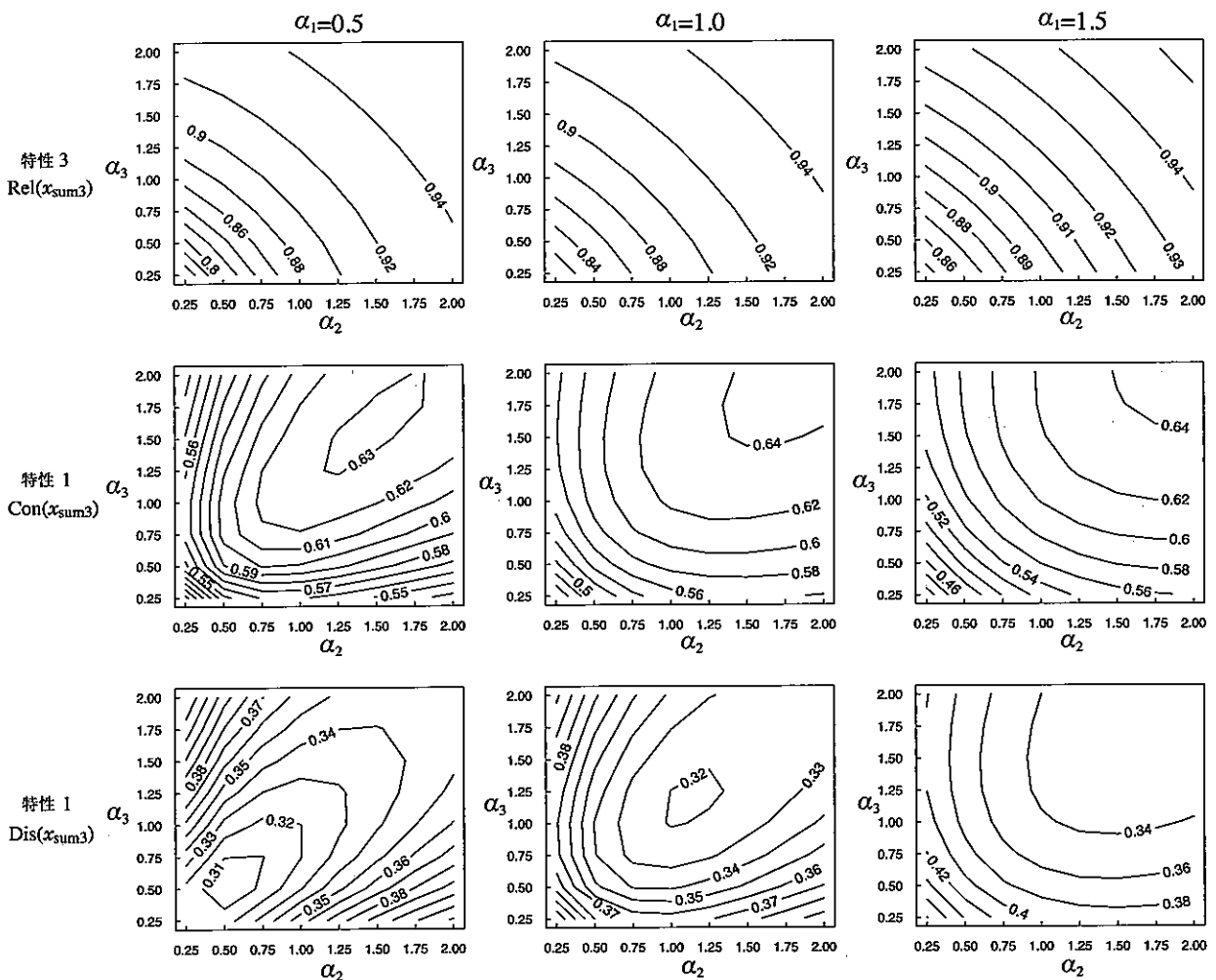


図1 項目数の変化に伴う再構成スコアの信頼性係数と妥当性係数の変動

それぞれ1.25とすれば収束的妥当性係数はベースラインより大きくなり、 $Con(x_{sum3}) = 0.63$ の領域に入る。

α_2 の値を固定して、 α_3 の値を大きくしていくと、ある一定の値までは収束的妥当性係数も大きくなるものの、それ以降は α_3 をいくら大きくしても収束的妥当性係数が再び小さくなってしまふ様子を図1の中段左図は表している。これは α_3 を固定して、 α_2 を大きくする状況でも同じである。例えば、 $\alpha_2=1$ とすると、 α_3 が約0.8以上で $Con(x_{sum3})$ の値は0.62を上回るが、さらに α_3 を1.5より大きくすると、収束的妥当性係数は再び小さくなってしまふ。これに対して中図と右図では等高線のカーブが緩やかになっており、 α_1 が1以上のときには、 α_2 と α_3 のいずれか一方のみを増加させた場合でも収束的妥当性係数の増加は頭打ちにならずに、妥当性が改善され続けることを示している。

また、図1中段右図から、 α_1 を増やしても、 α_2 および α_3 が約1.3を下回ると、表3に示したベースラインと比較して収束的妥当性が低下することが示唆された。例えば、 $\alpha_1=1.5$ と“自己評定”の項目数のみ増やし、“上司評定”と“同僚評定”については $\alpha_2=\alpha_3=0.5$ とすると、 $\{\alpha_1:\alpha_2:\alpha_3\}=1:1:1$ のときに0.632だった収束的妥当性係数は約0.50まで低下することがわかる。

特性1の弁別的妥当性に関する考察 最後に、非弁別的妥当性係数の変動については、 $Dis(x_{sum3})$ の値がベースラインで0.3を上回っていた特性1を考察の対象とする。図1下段の3つの図の比較から、 α_1 の値を大きくするにつれて非弁別的妥当性係数が大きくなり、弁別的妥当性が低くなることが示された。

図1下段中図より、“自己評定”の項目数は変えず $\alpha_1=1$ としたときに、もっとも弁別的妥当性を高くするには、 $\alpha_2=\alpha_3=1.0\sim 1.3$ とすればよいことがわかる。この領域から α_2 と α_3 を増やしても減らしても、非弁別的妥当性係数の値は大きくなってしまふ。

$\alpha_1=1.5$ のときについて図示した図1右図は、“上司評定”の項目配分 α_2 と“同僚評定”の項目配分 α_3 をそれぞれ大きくするにしたがって非弁別

的妥当性係数が小さな値となることを表しているが、 $\alpha_2=\alpha_3=2$ としてもなお0.34であり、ベースラインの0.319を上回っている。すなわち、もともと測定精度の低い“自己評定”を増やしたことで、他者評価の項目数を増やしても弁別的妥当性が改善されにくくなっていると解釈できる。

5. 結論

本論文の目的は、評定者によってテストの項目数を変化させた場合の、再構成スコアにおける信頼性係数と妥当性係数の一般式を導出し、多面観察評価における項目数と、測定の信頼性および収束的・弁別的妥当性との関係を考察することであった。多面観察評価データの信頼性と妥当性を改善するためには、もちろん他者評定に関して評定者の人数を増やすことも有効である。しかしながら、日本企業においては、上司評定は一次人事考課者に該当する一人のみが担当することが多く、収集可能な同僚評定の数も部署の規模に応じて決まってくる。したがって、評定者間での項目数の配分と、測定の信頼性、妥当性との関係を明確にすることは、多面観察評価の適正な運用のための有益な示唆となり得る。

多面観察評価データをMTMMデータとして扱い、確認的因子分析モデルを適用することで、再構成スコアの信頼性係数と収束的妥当性係数、非弁別的妥当性係数を表現した。これらの係数を、多面観察評価ツール“ビジネス基礎力診断S-BASE”の実施結果に対して計算した結果から、評定者ごとに質問紙の項目数を変化させ、適切に配分することでより効率的に測定の信頼性と妥当性を高められることが確認できた。

分析の結果、“自己評定”の項目数を増やすと信頼性と妥当性は低下し、一方で、“上司評定”と“同僚評定”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。これらの傾向は、本データにおいては特に収束的妥当性と弁別的妥当性に関して顕著であった。今回用いた“ビジネス基礎力診断S-BASE”で、現実的に実行可能な範囲内で項目数を変更するとすれば、“自己評定”については半分に減らして約17項目

とし、上司と同僚に回答してもらう項目数は現状の1.25倍の40項目前後まで増やすことで、いずれの特性においても全体的に信頼性と妥当性が向上することが示唆された。

人事考課に携わるのは上司であるという事実を鑑みても、上司に回答してもらう項目数を増やして測定の信頼性と妥当性を高めようという試みは合理的である。また、今回の分析結果において、上司評定と同様に、同僚評定の項目数の増加もまた、信頼性と妥当性の向上につながる事が明らかとなった。この結果は、自己評定と人事評価との相関はほとんどないのに対して、同僚評定を含む他者評定は人事評価と比較的高い相関があるという二村 (1998) の報告を支持するものである。多面観察評価を利用してより適切な人事アセスメントを実施するためには、人事考課に直接関わる上司だけではなく、同僚評定も疎かにしてはならない。

現実の多面観察評価実施場面では、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。回答者への負担、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、“自己評定”の重み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、目標とする信頼性係数と収束的妥当性係数、非弁別的妥当性係数を設定した上で、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について検討するためにも有用である。測定の質は高い水準で保ちつつ、各回答者への負担を必要最小限に抑えた理想的な人事アセスメントツールの作成に役立てることができる。

以上より、 $\{\alpha_1: \alpha_2: \alpha_3\}$ を変化させ、再構成スコアの信頼性係数と妥当性係数を計算することで、測定全体の信頼性と妥当性を高めるために最適な項目数の配分を検討できることが示された。一方で、本研究の結果から得られる項目配分に関する示唆は、信頼性と妥当性という心理測定論的観点に依拠するものであり、項目配分 α を変化させることの実践的意義を直接明らかにすることは難しいという課題も残る。信頼性や妥当性が極

端に損なわれない範囲で項目数を減らすことが可能なのであれば、フィードバックの際の情報損失も最小限に抑えられると見なすことができ、実施時間短縮や回答者への負担軽減といったメリットの方が大きいと考えられる。しかしながら、評定者の立場によって回答してもらう項目数を変更することが、実践場面でどのような意味を持つのかについては、今後慎重に検討する必要がある。

引用文献

- ・相原孝夫・南雲道朋 (2009) 『チームを活性化し人材を育てる360度フィードバック』 日本経済新聞出版社。
- ・馬場昌雄・馬場房子監修、岡村一成・小野公一編 (2005) 『産業・組織心理学』 白桃書房。
- ・Brennan, R. L. (1992). *Elements of generalizability theory. Revised ed.* Iowa City, IA: American College Testing Program.
- ・Brennan, R. L. (2001). *Generalizability theory.* New York: Springer.
- ・Cambell, D. T., & Fiske, D. W. (1959). "Convergent and discriminant validation by the multitrait-multimethod matrix." *Psychological Bulletin*, 56, 81-105.
- ・Conway, J. M. (1996). "Analysis and design of multitrait-multirater performance appraisal studies." *Journal of Management*, 22, 139-162.
- ・Conway, J. M. & Huffcutt, A. I. (1997). "Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings." *Human Performance*, 10, 331-360.
- ・Conway, J. M., Lombardo, K., & Sanders, K. C. (2001). "A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings." *Human Performance*, 14, 267-303.
- ・Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: *Theory of generalizability of scores and profiles.* New York: Wiley.
- ・Edwards, M. R., & Ewen, I. J. (1996). *360° feedback: The powerful new model for employee assessment & performance improvement.* New

- York: AMACOM American Management Association.
- Eid, M. (2000). "A multitrait-multimethod model with minimal assumptions." *Psychometrika*, 65, 241-261.
 - 学校法人産業能率大学総合研究所『ビジネス基礎力診断 S-BASE (エス・ベース)』 <http://www.hj.sanno.ac.jp/cp/page/7638> (2014年2月)
 - 学校法人産業能率大学総合研究所『プロ人材特性診断 S-Pro (エスプロ) II』 <http://www.hj.sanno.ac.jp/cp/page/5714> (2014年2月)
 - 学校法人産業能率大学総合研究所『マネジャー力診断 S-ATOM (エスアトム)』 <http://www.hj.sanno.ac.jp/cp/page/5716> (2014年2月)
 - 学校法人産業能率大学総合研究所 (2012). 『～最適人事と能力開発のための～アセスメントツール総合ガイド2012-2013』.
 - Grayson, D., & Marsh, H. W. (1994). "Identification with deficient rank loading matrices in confirmatory factor analysis: multitrait-multimethod matrix." *Psychometrika*, 59, 121-134.
 - Greguras, G. J., & Robie, C. (1998). "A new look at within-source reliability of 360-degree feedback ratings." *Journal of Applied Psychology*, 83, 960-968.
 - 南風原朝和 (2011)『臨床心理学をまなぶ7 量的研究法』 東京大学出版.
 - 入江崇介・鷺坂由紀子・舛田博之・二村英幸 (2004)「多面観察評価手法の特性に関する検討～一般化可能性理論によるアプローチ～」『日本テスト学会第2回大会発表論文抄録集』, 68-69.
 - 入江崇介・鷺坂由紀子・舛田博之・二村英幸 (2005). 「多面観察評価における上司・同僚・部下の評定結果の統合について -多変量一般化可能性理論による一般化可能性の検討-」『日本テスト学会第3回大会発表論文抄録集』, 60-63.
 - Lawler, E. E., III. (1967). "The multitrait-multirater approach to measuring managerial job performance." *Journal of Applied Psychology*, 51, 369-381.
 - Lepsinger, R. & Lucia, A. D. (2009). *The art and science of 360-degree feedback. 2nd Ed.* San Francisco: Jossey-Bass A Wiley Import.
 - Marsh, H. W., & Grayson, D. (1995). "Latent variable models of multitrait-multimethod data." In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issue, and applications.* Thousands Oaks, CA: Sage. pp.177-198.
 - McDonald, R. P. (1999) *Test theory: A unified treatment.* New Jersey: Lawrence Erlbaum.
 - Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R. & Hezlett, S. A. (1998). "Trait, rater and level effects in 360-degree performance ratings." *Personnel Psychology*, 51, 557-576.
 - 日本経営協会総合研究所『多面観察<マネジメント版・リーダーシップ版>』 <http://www.noma.co.jp/service/sas/sas0101.html> (2014年2月)
 - 日本能率協会マネジメントセンター『NAVI 360』 <http://www.jmam.co.jp/productservice/assess/item/navi360.html> (2014年2月)
 - 日本能率協会マネジメントセンター『Round Review』 <http://www.jmam.co.jp/productservice/assess/item/roundreview.html> (2014年2月)
 - 二村英幸 (1998)『人事アセスメントの科学』 産能大学出版部.
 - 二村英幸 (2001)『人事アセスメント入門』 日本経済新聞社.
 - 大沢武志 (2000)「経営人事における人事アセスメント」大沢武志・芝祐順・二村英幸編『人事アセスメントハンドブック』 金子書房, 3-22.
 - リクルートマネジメントソリューションズ『MOA | 360° 評価システム』 http://www.recruit-ms.co.jp/service/service_detail/orgkey/A005/ (2014年2月)
 - 関本昌秀 (1965)『わが国における人事考課の反省-管理者層の人事考課を中心として』 慶應ビジネスフォーラム.
 - Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* London: Sage Publications.
 - 高橋潔 (2010)『人事評価の総合科学-努力と能

力と行動の評価』 白桃書房。

- ・田中堅一郎編 (2011)『産業・組織心理学エッセンシャルズ【改訂三版】』 ナカニシヤ出版。

○ ○ ○

- 1) 方法因子を仮定せず独自因子に相関を持たせたモデル。
- 2) 本調査では、“ビジネスに必要な基礎的能力”が測定したい構成概念であり、“自己確立”、“仕事確立”、“関係性確立”という3つの領域がMTMM データにおける特性として扱われる。また方法は“自己評定”、“上司評定”、“同僚評定”の3つであり、 $t=3, m=3$ である。

- 3) 表1では Campbell & Fiske (1959) に従い、信頼性および収束的妥当性、弁別的妥当性の解釈を行うために相関行列を示したが、信頼性係数と妥当性に関する係数を計算する際には、共分散行列をもとに推定された CT-CU モデルの非標準化解を用いた。特に、再構成スコアの信頼性と妥当性を考えるとき、分析に相関行列を用いたり、係数の計算時に標準化解を用いたりすると、データを基準化した上で項目数を変化させることとなるため、本手法の適用にあたっては、共分散行列（生データ）から計算される非標準化解を利用する必要がある。

