

管理者層を対象とした性格検査・知的能力検査 の妥当性のメタ分析と一般化

人事測定研究所 二 村 英 幸・今 城 志 保・内 藤 淳

A meta-analysis and validity generalization study of a personality test and a general cognitive ability test for measuring managerial aptitude.

Hideyuki Nimura, Shiho Imashiro, & Jun Naito
(Human Resource Research Institute)

A meta-analysis was conducted based on 24 studies reporting the validity of a personality test and a general cognitive ability test, both of which are subtests of a managerial aptitude test, NMAT(New Managerial Aptitude Test). Criteria used in the studies were typically based on performance appraisals by superiors.

Regarding the personality test, the estimated correlation coefficients of Compliance- Leadership (similar to Potency claimed to be a subdimension of the Big5 Extraversion) and of Conservative -Innovative (similar to Achievement claimed to be a subdimension of the Big5 Conscientiousness) were $r=0.311$ and $r=0.259$ respectively, after correction of range restriction and imperfect reliabilities. As to the cognitive ability test, the estimated validity coefficient of Cognitive Ability was $r=0.257$ after correction of range restriction and imperfect reliabilities. In general, the validity of the personality test is higher and the validity of the cognitive ability test is lower than those generally reported in the U.S.

The possibility of validity generalization was also explored. For the personality test, 45 to 100% of total variance was accounted for by artificial factors, whereas only 25% was so for the general cognitive ability test. Thus, it was concluded that the validity of the personality test is more stable than that of the cognitive ability test over various factors, such as differences in organizational culture or in job content.

The results were discussed in relation to the nature of job environment in Japan, where behaviors regarded as Organizational Citizenship Behavior (OCB) are more explicitly expected and appraised than in the U.S. Further, the possibility was considered that the relatively high validity of the personality test may be partly due to the fact that the predictors had been developed specifically for evaluation of managerial competencies.

The article concludes by pointing out the limitations of this study and future directions of the meta-analysis.

I. 問題の所在と研究の目的

適性検査には通常、性格的な特性を測定するものと能力的な特性を測定するものがあり、採用選考場面や昇進・昇格選考場面等において、意思決定のための参考情報として利用されている。利用目的に沿って適性検査の妥当性を確認しておくことは、経営人事に関わる実務家や研究者にとっての大きな関心事である。

米国においては性格検査、知的能力検査ともに数多くの妥当性研究がこれまで行われてきた。特に近年、過去の数多くの研究を統合するメタ分析の手法が用いられる

ようになったことで、状況の個別性を超えて得られる妥当性係数の平均的な値や、妥当性の高低に影響を与える要因の特定、またその作用の方向などについての研究が盛んに行われている。これに対して、日本においては発表された妥当性研究は数少ないのが現状である（高橋・西田, 1994 ; 二村, 1998）。

性格検査の妥当性については従来、Guion (1965) や Mischel (1968) に代表されるように、職務遂行能力 (Performance) との間に一般化できるような関連性は見られないとされ、たとえ関連性が確認されたとしてもその程度は極めて小さく、性格検査を職務遂行能力の予

測変数とするのは現実的ではないという主張が大勢であった。しかし、Big 5という性格理解のための共通の枠組みが提唱されたことが、メタ分析手法の発展とあいまって、性格検査に関するこの悲観的な見解に対して一石を投じることになった。

Barrick & Mount (1991) は、Big 5 の枠組みを用いて性格検査に関するメタ分析を行い、Conscientiousness (勤勉性もしくは誠実性) がさまざまな職種において0.20～0.23（補正值、職務遂行能力評価が基準）の妥当性係数を示すことを報告している。また、Hough (1992) が行ったメタ分析は、Conscientiousness の構成要素の一つであるとされる Achievement (達成動機) の妥当性係数が0.15（補正なし、職務遂行能力評価が基準）であることを示している。これらの研究結果は、少なくとも性格特性と行動の結びつきは単なる偶然や統計的な誤差によるものではなく、性格検査を職務遂行能力の予測に用いることには意味があるという主張に裏付けを与えたことになり、性格検査の復権といわれる近年の米国における状況をもたらす契機の一つとなつた。

しかしながら、Barrick ら (1991) の研究の中で、産業特性や企業特性あるいは職務特性などによる状況の個別性を超えて妥当性係数の真値が一つに定まるか（妥当性一般化：Validity Generalization）を検証した分析では、一般化はできないという結果が示されている。すなわち、妥当性係数の真値はさまざまな状況要因によって異なるということである。Barrick ら (1991) や Hough (1992) のこれらの研究においては、妥当性係数に影響を及ぼす要因の特定も試みられており、ほとんどの尺度について職種や基準変数の性質によって妥当性係数が異なるという結果が得られている。また、職場における役割行動の明確さの相違が妥当性係数の値に影響を及ぼすという研究報告もあり (Barrick & Mount, 1993)，性格検査と職務遂行能力との間に関連性があることは確認されたものの、その関連性の強さについては状況によってさまざまであると考えるのが妥当であろう。

他方、知的能力検査の妥当性に関しては米国では数多くのメタ分析が行われており、g 因子と呼ばれる一般知的能力因子について高い水準の妥当性係数が安定的に得られるというのが、おおむね統一された見解となっている。Hunter & Hunter (1984) は、職種別のメタ分析を行い、職務の複雑さによって値は異なるものの0.23～0.56（補正值、職務遂行能力評価が基準）を報告している。また、対象となる業種や職種を限定したメタ分析もこれまで相当数が行われている。Pearlman, Schmidt & Hunter (1980) は事務職を対象とした研究で0.52

（補正值、職務遂行能力評価が基準）を報告している。また、Schmidt, Hunter & Caplan (1981) は石油業界のオペレータを対象に0.26、メンテナンスを行う人を対象に0.30（ともに補正值、職務遂行能力評価が基準）を報告している。日本では高橋・西田 (1994) がメタ分析を行っており、0.300という妥当性係数の値（基準変数の信頼性係数を0.60とした場合の補正值）が報告されている。

Schmidt らの研究グループは、一般知的能力検査の妥当性を幅広く一般化することが可能であると主張しているが、これはあらゆる状況を通じて妥当性係数の真値が一つに決まるという意味よりは、むしろ一般知的能力検査がさまざまな状況の違いを超えて安定して高い水準の妥当性係数を示すとの主張であると考えてよい。これまで行われてきた数多くの妥当性一般化に関する研究を見ても、一般化できたものもあればそうでないものもあり、一般化できているケースでは、職種別に分析を行ったり、基準変数を職務遂行能力評価と研修評価とに分けたり、専門的な能力に特化してより細分化された能力検査ごとに分析を行うなど、何らかの形で状況が統制されていることが多い。

妥当性係数の値に影響を与える要因としてほぼコンセンサスが得られているものの一つに職務の複雑さがあり、Hunter ら (1984) の研究によれば、職務が複雑なほど妥当性が高くなることが報告されている。また、別の要因として基準変数の性質があげられ、職務遂行能力を評価したものに比べ研修評価を基準としたときのほうが妥当性は高くなることが確認されている (Hunter, 1986)。

対象を管理者に限った場合のメタ分析結果に目を向けると、性格検査については、前述の Barrick ら (1991) が職種別の分析を行っており、管理者について Conscientiousness (勤勉性、誠実性) の妥当性係数が0.22、Extraversion (外向性) が0.18（ともに補正值、職務遂行能力評価が基準）であることが報告されている。Hough (1992) の分析では、Achievement (達成動機) が0.18であるのに加えて Potency (支配性) も0.18（ともに補正なし）という妥当性係数が報告され、これらの結果は、前向きな姿勢や対人場面での積極性、押しの強さといった特徴が、管理者としての評価と関連していることを示唆するものとされている。

また、一般知的能力検査では、Hunter ら (1984) の職種別分析で管理者に対しては0.53と、営業の0.61、事務の0.54に次いで高い妥当性が報告されている（いずれも補正值、職務遂行能力評価が基準）。日本でも、高橋 ら (1994) の研究で管理者層を対象として職務内容

別に分析が行われており、0.286～0.416（基準変数の信頼性係数を0.60とした場合の補正值）が報告されている。全般的に、管理者層における知的能力検査の妥当性係数は他の職種と比べてやや高い値が得られている。

本研究では管理者層を対象に、性格検査および一般知的能力検査に関する妥当性一般化の検証を含むメタ分析を行う。職務遂行能力評価を基準変数としたときの能力検査の妥当性に関するメタ分析としては、日本にも高橋ら（1994）による先行研究があるが、職務遂行能力評価を基準変数としたときの性格検査の妥当性に関するメタ分析は、筆者が確認した範囲では見当たらず、本研究がおそらく初めての試みとなる。

米国と日本との人事管理や経営に関する風土の違いあるいは文化的な相違を考えれば、日本における性格検査の妥当性は米国とは異なる結果を示す可能性も十分にある。例えば、米国における研究で、性格検査の妥当性は、組織市民行動（Organizational Citizenship Behavior: OCB）と呼ばれる勤勉さ、同僚を支援していく姿勢、あるいは組織や職務に対する前向きな態度など、職務としては明確に定義されない行動を基準変数とした際により高くなる可能性が指摘されている（Motowidlo & Van Scotter, 1994; Schneider & Hough, 1995）。このような組織や職務に対する前向きな姿勢や態度は、日本の企業においては情意考課という形で直接的に人事評価の対象とされていることも多く、従来より従業員に求められてきた要素であるといえる。この点を考慮すれば、性格検査の妥当性係数の値が米国における以上に日本において高くなるという可能性も考えられる。

このような現状の中で、本研究では管理者層を対象に以下の2点について確認する。

- ① 職務遂行能力評価を基準変数としたときの性格検査、一般知的能力検査の妥当性の程度
- ② 上記の妥当性に関する職務内容の違いや企業の個別性を超えた妥当性一般化の可能性

II. 方 法

1. 使用するデータ（予測変数および基準変数）

本研究では、予測変数に管理者適性検査NMAT（人事測定研究所）の諸尺度を用いて、妥当性に関するメタ分析および妥当性一般化の検証を行う。NMATは、一般知的能力検査、性格検査、指向検査の3つで構成される適性検査であり、企業における管理者の選抜場面や管理者層を対象にした研修の際の参考資料として主に利用されている。本研究では、NMATに含まれる性格検査と一般知的能力検査を予測変数として用いる。

性格検査は対極的な概念の対で構成される両側形式の尺度であり、「内向－外向」、「調整－統率」、「心情－理性」、「繊細－強靭」、「維持－変革」、「思索－行動」、「慎重－大胆」、「自律－承認」¹⁾の8尺度よりなる。これらの尺度は、得点が高くなるほど尺度名に示された右側の概念で表される特性を強く持つことを意味している。また、一般知的能力検査は、言語的な能力を測定する「概念的理説」と論理・数量的な能力を測定する「論理的思考」の2つの下位尺度、およびその合成指標である「総合」より構成される。性格検査、能力検査ともに、企業における管理者層を基準母集団として標準化されている。それぞれの尺度の信頼性係数（Cronbachのα係数）を表1に示す。

表1 性格検査および一般知的能力検査の信頼性係数
(n=1,592)

性格検査	α係数	一般知的能力検査	α係数
内向－外向	0.93	概念的理説	0.84
調整－統率	0.90	論理的思考	0.83
心情－理性	0.86	総合	0.92
繊細－強靭	0.91		
維持－変革	0.89		
思索－行動	0.88		
慎重－大胆	0.87		
自律－承認	0.81		

基準変数としては、NMATを管理者選考場面あるいは研修場面で使用している企業22社（製造業12社、非製造業10社：うち14社は従業員数1,000名以上）における職務遂行能力評価を用いる。ただし、22社中2社については異なる2つの階層に対するデータが得られたため、メタ分析の対象としたのは合計24の個別妥当性研究である。職務遂行能力評価は1996年2月～1997年8月の間に収集されたものであり、その内訳は上司による人事考課が21、昇進スピードが2、研修場面における評価が1となっている。1研究あたりの被験者数の平均は184人（35人～859人）であり、この種のメタ分析としては比較的大きな標本に基づくものである。

2. メタ分析の手法

妥当性一般化の手続きについては、これまでにさまざまな手法が提案されているが（Raju, Anselmi, Goodman & Thomas, 1998），本研究ではHunter & Schmidt（1990）に説明されている手続きに従う。この手続きは、先行研究であるBarrick & Mount

(1991) や高橋ら (1994) が用いたものと同じである。

一般に、基準関連妥当性の研究において得られる妥当性係数は、個々の研究ごとにかなり異なる値をとることが多い。この妥当性係数の値の研究間におけるばらつきは、従来、産業特性、企業特性、職務特性、基準変数や予測変数の性質など、状況による個別性 (Situational Specificity) の反映であるとされてきたが (Ghiselli, 1966), この説に反対して、Schmidt & Hunter (1977) が唱えたのが妥当性一般化 (Validity Generalization) の考え方である。それによれば、妥当性係数の観測値間に見られるばらつきは人為的な誤差 (Artifacts) が混入した結果であるとされ、もし何らかの手法を用いてこれらの誤差を完全に除去することができれば、妥当性係数の真値はさまざまに異なる状況を通じてただ1つの値に固定できると考えられる。

研究間における妥当性係数のばらつきを生じさせる人為的な誤差として、Schmidt, Gast-Rosenberg & Hunter (1980) は、つぎの7種類をあげている。1) 標本誤差によるもの、2) 基準変数の測定誤差によるもの、3) 予測変数の測定誤差によるもの、4) 予測変数の範囲制限 (Range Restriction) によるもの、5) 基準変数が構成概念からずれていることによるもの (Contamination & Deficiency)、6) 研究者の計算、タイミング、転記の間違いによるもの、7) 個々の研究間で予測変数の構成概念が微妙に異なっていることによるもの。

以上の7つの人為的な誤差のうち、5) から7) までの3つの要因については、その影響を取り除くことはできないが、1) から4) までの4つの要因については、統計的な手法を用いることによってその影響を除去することが可能である。実際に、妥当性係数の観測値の分散のすべてでなくとも、その大部分が1) から4) までの要因に起因する分散によって説明できる場合には、妥当性が一般化されたものと判断することができる。その際の一つの判断基準として、Schmidt, Gast-Rosenberg & Hunter (1980) は75%ルールと呼ばれる基準を提示している。これは、1) から4) までの要因に起因する分散が、妥当性係数の観測値の分散のうちの75%以上を説明するならば、妥当性は一般化できるとするものである。ただし、妥当性一般化の判断基準は、絶対的な根拠に基づくものというよりはある意味で恣意的に決めざるを得ないものであり、この75%ルールに関しては批判もある (James, et al., 1988; Kemery, et al., 1987; Osburn, et al., 1983; Pease & Switzer III, 1988; Sackett, et al., 1986)。

また、状況個別性が完全には棄却されない場合でも、

推定された妥当性係数の真値が十分に大きいときには、その妥当性係数の値は意味を持ち、他の異なる状況に適応することが可能だと考えられる。具体的には、推定された妥当性係数の真値の分布の下方から10パーセンタイルにあたる値 (90%信頼限界下方値 : 90% Credibility Value) が0を超えていれば、その妥当性は状況を超えた適用可能性 (Transportability : Schmidt, Hunter & Caplan, 1981; Kemery, Mossholder & Roth, 1987) を持つものと解釈される。なお、上述した状況に対しても同様に妥当性一般化という言葉が用いられることがあるが、ここでは用語上の混乱を避けるため、妥当性係数の真値が異なる状況を通じてただ1つの値に固定できる場合を妥当性一般化という言葉で表し、90%信頼限界下方値が0を超えている状況に対しては適用可能性を有するという表現を用いることにする。

3. 計算手続き

本研究では、観測された個々の妥当性係数の値に以下に述べる計算手続きを用いて補正をほどこし、最後に研究ごとの標本数による加重平均をとることによって「妥当性係数の真値の推定値 ($\hat{\rho}$)」および「妥当性係数の真値の分散 ($\sigma_{\hat{\rho}}^2$)」を求める。

なお、以下の式における記号はそれぞれつぎの意味を表す (ただし表記中、添字の*i*は個別研究の別を表す)。

ρ_{xx} : 予測変数の信頼性係数

ρ_{yy} : 基準変数の信頼性係数

U_i : [範囲制限を受けていない予測変数のSD]

/ [範囲制限を受けた予測変数のSD]

r_i : 観測された妥当性係数

r_{ci} : 予測変数の範囲制限と予測変数、基準変数の信頼性の低さによる希薄化の補正後の妥当性係数

n_i : 標本数

N : 総標本数 ($\sum n_i$)

σ_{ei}^2 : 補正前の妥当性係数の標本誤差分散

σ_{cei}^2 : 補正後の妥当性係数の標本誤差分散

$\hat{\rho}$: 妥当性係数の真値の推定値

$\sigma_{\hat{\rho}}^2$: 妥当性係数の真値の分散

1) 範囲制限と信頼性の低さによる希薄化の補正

①範囲制限による希薄化の補正

$$r'_i = U_i r_i / \sqrt{(U_i^2 - 1) r_i^2 + 1}$$

②信頼性の低さによる希薄化の補正

$$r_{ci} = r'_i / \sqrt{\rho_{xx} \rho_{yy}}$$

2) 補正後の妥当性係数の標本誤差分散 (σ_{cei}^2) の算出

$$\sigma_{\text{eei}}^2 = \alpha_i A_i \sigma_{ei}^2$$

$$\text{ただし, } \alpha_i = 1 / ((U_i^2 - 1) r_i^2 + 1)$$

$$A_i = r_{ei} / r_i$$

$$\sigma_{ei}^2 = (1 - r_i^2)^2 / (n_i - 1)$$

3) 妥当性係数の真値の推定値 ($\hat{\rho}$) の算出

$$\bar{r}_e = \sum n_i r_{ei} / N$$

4) 妥当性係数の真値の分散 (σ_{ρ}^2) の算出

$$\sigma_{\rho}^2 = \bar{\sigma}_{re}^2 - \bar{\sigma}_{ee}^2$$

$$\text{ただし, } \bar{\sigma}_{re}^2 = \sum n_i (r_{ei} - \hat{\rho})^2 / N$$

$$\bar{\sigma}_{ee}^2 = \sum n_i \sigma_{ei}^2 / N$$

メタ分析に必要なデータのうち、予測変数の信頼性係数および範囲制限については、今回用いられた予測変数がいずれも同一の標準化された検査であったことから、すべてを把握することができた（範囲制限の比率を表す $1/U$ の範囲は0.66～1.24）。一方、基準変数の信頼性についてはデータがまったく入手できなかったため、先行研究を参考にしてその値を定める必要があった。Schmidt & Hunter (1977)によれば、妥当性研究の基準となる変数の信頼性は、異なる評価者による、適切なインターバルをおいた再検査法によって求められるべきであり、その値は0.60を超えないとしている。また、その後の多くの妥当性一般化の研究 (Pearlman, Schmidt & Hunter, 1980; Schmidt, Gast-Rosenberg & Hunter, 1980) が、基準変数の信頼性係数としてこの値を用いていることを考慮して、本研究においてもこの0.60という値を用いて計算を行う。

III. 結果

表2は、性格検査および一般知的能力検査に関するメタ分析の結果である。この表には、性格検査、一般知的能力検査の各尺度ごとに、補正前の妥当性係数、補正後の妥当性係数（妥当性係数の真値の推定値 $\hat{\rho}$ ）、妥当性係数の真値 ρ の分散 (σ_{ρ}^2)、妥当性係数の観測値の分散 (σ_{re}^2)、誤差による分散 (σ_{ee}^2)、誤差による分散説明率、および90%信頼限界下方値が示されている。

最初に、推定された妥当性係数の真値 ($\hat{\rho}$) に注目する。性格検査に関しては、「調整－統率」と「維持－変革」の妥当性係数がそれぞれ0.311, 0.259と高く、これは一般知的能力検査の妥当性係数（総合で0.257）と同水準である。またこの値は、Barrickら (1991) の研究において最も高い妥当性係数を示したConscientiousness（勤勉性、誠実性）の0.22（管理者を対象）と比べても高いものである。さらに、「調整－統率」「維持－変革」以外の尺度についても、「心情－理性」が0.183、「内向－外向」が0.179、「繊細－強靭」が0.135、「慎重－大胆」が0.126と、それぞれある程度の水準の値を示しており、「自律－承認」と「思索－行動」に関しては妥当性係数がそれぞれ0.061, 0.031と低かったものの、性格尺度全体では8尺度中6尺度で0.10を超え、そのうちの4尺度では0.15を超える妥当性係数が確認された。

一方、一般知的能力検査に関しては、「総合」で0.257という妥当性係数が示された。これは米国における研究

表2 性格検査および一般知的能力検査に関するメタ分析結果（総標本数4,420、妥当性研究数24）

	補正前平均 妥当性係数	$\hat{\rho}$ （補正後の 妥当性係数）	σ_{ρ}^2	σ_{re}^2	σ_{ee}^2	誤差による 分散説明率	90%信頼 限界下方値
性格検査							
内向－外向	0.136	0.179	0.0064	0.0162	0.0097	60%	0.076
調整－統率	0.219	0.311	0.0035	0.0138	0.0103	74%	0.234
心情－理性	0.121	0.183	0.0010	0.0137	0.0127	93%	0.143
繊細－強靭	0.100	0.135	0.0048	0.0147	0.0098	67%	0.046
維持－変革	0.175	0.259	0.0133	0.0241	0.0108	45%	0.111
思索－行動	0.020	0.031	0.0026	0.0134	0.0108	81%	-0.034
慎重－大胆	0.088	0.126	0.0000	0.0119	0.0119	100%	0.126
自律－承認	0.039	0.061	*0.0000	0.0114	0.0128	*100%	0.061
一般知的能力検査							
概念的理解	0.137	0.200	0.0279	0.0402	0.0123	31%	-0.013
論理的思考	0.193	0.272	0.0263	0.0382	0.0119	31%	0.064
総合	0.183	0.257	0.0325	0.0436	0.0111	25%	0.026

• 表中「自律－承認」の σ_{ρ}^2 や誤差分散の説明率は、計算ではそれぞれ -0.0014, 112% であった。

• 誤差による分散説明率とは、観測された妥当性係数間の見かけの分散のうち、標本誤差、基準変数と予測変数の信頼性の低さによる誤差、予測変数の範囲制限による誤差によって説明される分散の割合を示す。

と比べるとその値は小さいものの、高橋ら（1994）の0.300とはほぼ同等の値であり、一定水準の妥当性係数が確認されたといえる。下位尺度別に見てみると、「概念的理解」が0.200、「論理的思考」が0.272であり、「概念的理解」に比べて「論理的思考」の妥当性係数のほうがやや高くなっている。

つぎに妥当性一般化の観点から結果を見てみる。まず性格検査に関しては、全体的に誤差による分散説明率が高いということがいえる。その値は、「自律－承認」、「慎重－大胆」、「心情－理性」、「思索－行動」において75%を超えており（それぞれ100%, 100%, 93%, 81%）、「調整－統率」、「繊細－強靭」、「内向－外向」においても75%にまでは達しないものかなり高い値を示している（それぞれ74%, 67%, 60%）。このことは、管理者層を対象としたときに、性格検査の妥当性係数の真値が異なる状況を通じて一つに定まる可能性があることを示している。一方、90%信頼限界下方値を見ると、「思索－行動」を除く7尺度でいずれも0よりも大きい値を示しており、「調整－統率」が0.234と最も高く、これに次いで「心情－理性」が0.143、「慎重－大胆」が0.126となっている。これらの7尺度については、その妥当性係数がさまざまな状況を超えて適用可能性（Transportability）を持つものであると解釈することができる。

性格検査の結果とは対照的に、一般知的能力検査に関しては誤差による分散説明率が「概念的理解」、「論理的思考」でともに31%、「総合」で25%と低い値となっている。この結果から判断すると、一般知的能力検査に関しては妥当性を一般化することはできず、妥当性係数の真値は個々の状況によって異なる可能性が高いといえる。しかし、90%信頼限界下方値を見ると「総合」では0.026とその値は0を超えており、一般知的能力検査の妥当性係数は真値が一つに定まるとはいえないものの、適用可能性を持つものであることが示唆される。

IV. 考察

本研究では、性格検査のいくつかの尺度に関して一般知的能力検査を上回る、あるいはそれと同等の高さの妥当性係数が確認された。以下、主にBig 5の枠組みで行われた米国におけるメタ分析結果のうち、特に管理者を対象とした先行研究との比較を行う。まず、「調整－統率」と「内向－外向」の2尺度について、その測定内容からExtraversion（外向性）に属するものと考えられる。Hough（1992）は、Extraversionを対人的な強さや積極性にあたるPotencyと社交性にあたるAffiliationの2つに分けて解釈しているが、「調整－統率」

はこのPotencyに、また「内向－外向」はAffiliationに近いものと考えられる。米国におけるメタ分析の結果が、Extraversionに関して0.18（Barrickら, 1991, 補正值）であるのに比較して、「調整－統率」では0.311と高く、「内向－外向」では0.179とほぼ同程度の水準である。また、PotencyとAffiliationに分けた場合に、Potencyに関しては0.18（Hough, 1992, 補正なし）が報告されているのに対して、「調整－統率」では補正前で見ても0.219を示した。

「維持－変革」は、Open to Experience（開放性：Hough, 1992の枠組みではIntellectance）に属すると考えられるが、Conscientiousness（勤勉性、誠実性）に含まれるAchievement（達成動機）のような意欲的侧面をも一部含んでいる。Barrickらの研究では、Open to Experienceが0.08、Conscientiousnessが0.22（ともに補正值）、またHoughの研究では、Intellectanceが0.09、Achievementが0.18（ともに補正なし）であるとの比較すると、「維持－変革」の0.259（補正前で0.175）はやや高い水準である。

また「繊細－強靭」は、Emotional Stability（情緒安定性）に近い尺度である。Barrickら（1991）の研究ではEmotional Stabilityについて0.08（補正值）が、Hough（1992）の研究では対応する尺度であるAdjustment（適合性）について0.11（補正なし）という値が報告されているが、それらと比べて「繊細－強靭」の0.135（補正前で0.100）は同等もしくはわずかながら高い値であるといえる。

一方、これ以外の尺度に関しては、Big 5の枠組みとの対応が明確ではないため直接の比較はできないが、「心情－理性」（0.183）や「慎重－大胆」（0.126）も一定水準の値を示している。このことは、Hough（1992）の指摘にもあるように、職務行動の予測という観点から見たときにBig 5の枠組みは必ずしも十分ではないということを示唆するものと考えられる。

上記の比較結果から見て、本研究で確認された性格検査の妥当性係数は、おしなべて米国におけるそれに比べ大きいということができる。このような結果が得られた理由として、以下の2点が考えられる。

1点目にあげられるのは、日本と米国における人事評価に関する風土の違いが影響している可能性である。日本企業においては従来よりまじめさや意欲・気配りといった仕事に対する態度や姿勢という側面が重視される傾向があるが、これらの要素はOCB（組織市民行動）に近いものと考えられる。性格検査の妥当性係数はOCBを基準変数にしたときに高まることが米国の研究で指摘さ

れている。例えば陸軍を対象にしたプロジェクトAでは、まじめさや職務に対する前向きな姿勢、達成動機の強さなどを反映していると思われる基準変数に対する性格検査の妥当性係数は0.30～0.36（補正なし）と高かったのに対して、専門的な知識やスキルを評価した基準変数に対する妥当性係数は0.15（補正なし）と相対的に低い水準にとどまっていることが報告されている（McHenry, Hough, Toquam, Hanson & Ashworth, 1990）。情意考課の存在に代表されるように、日本においては人事評価の中にOCB的な要素がより高い比重で含まれているとすれば、そのことが影響して今回の結果ではリーダーシップや変革意欲に関わる性格尺度の妥当性係数が高くなっている可能性が考えられる。

2点目にあげられるのは、今回使用した適性検査が管理者という特定対象の適性把握を目的として開発されたものであることが影響している可能性である。米国における先行研究の場合、一般性の高い性格検査が用いられていることが多い、そのため性格検査の妥当性が低くなっている可能性があることが指摘されている（Hough & Schneider, 1996）。すなわち、人事管理場面での使用を目的として性格特性と職務行動の結びつきを理論的に想定した上で開発された検査を用いれば、より高い妥当性が期待できるというのである。本研究で性格検査に関してある程度高い妥当性係数が確認されたのも、使用した性格検査のこの特性によるものである可能性も考えられる。

メタ分析ではなく特定のサンプルを用いて行われた妥当性研究であるため一般性のある結果と解釈することはできないが、米国においても管理者適性の測定を目的に開発された性格検査に関する研究がいくつかあり、CPI Managerial Potential Scale (Gough, 1984) については軍隊での総合評価を基準変数として妥当性係数が0.20（補正なし）、またHPI Managerial Potential Scale (Hogan & Hogan, 1986) については職務遂行能力評価を基準変数として妥当性係数が0.26（補正なし）と、いずれも比較的高い値が報告されている。

一方、本研究における一般知的能力検査の妥当性係数は「総合」で0.257であり、これは高橋ら（1994）が報告している0.300という結果とほぼ同等の水準である。

2つのメタ分析の結果をもって一般的な結論を導くことはできないものの、日本においては一般知的能力の妥当性係数が米国で示されている0.4～0.5といった高い水準ではなく、0.25～0.30程度に収まる可能性がある。前述したように、日本においてはまじめさや職務に対する前向きな姿勢、意欲の高さといった要素を評価する傾向が米国に比べて強いと考えれば、知的能力検査が良く予測

すると考えられる知的な労働やテクニカルな能力の発揮度を重視する程度が相対的に低くなり、その結果妥当性係数がそれほど高くならないという可能性もある。すなわち、米国と異なる日本の特殊な事情が、性格検査の妥当性を高める一方で知的能力検査の妥当性を相対的に低めていることが考えられる。

続いて、今回の結果を妥当性一般化の観点から検討してみる。「思索－行動」を除く性格検査のすべてと、能力検査の「総合」および「論理的思考」については90%信頼限界下方値が0を超えてのことから、少なくとも管理者層を対象とした場合には、産業特性や企業特性あるいは職務特性などの状況個別性を超えて、性格検査や一般知的能力検査が安定的に有効な妥当性係数を示すことが示された。一般知的能力検査の適用可能性については米国における妥当性一般化の研究を通じて広く確認されており、今回の結果は日本においても一般知的能力検査が適用可能性を有することを、高橋ら（1994）の研究に次いで確認するものとなった。一方、性格検査については、Barrickら（1991）の研究の中で管理者を対象として適用可能性が確認されたのはConscientiousness（勤勉性、誠実性）、Agreeableness（親和性）、Extraversion（外向性）の3尺度であり、これらに対応する下限値がそれぞれ0.09, 0.06, 0.01であったことと比較すると、本研究で得られた下限値は、「調整一統率」で0.234、「心情一理性」で0.143、「慎重一大胆」で0.126、「維持一変革」で0.111とより大きな値であった。このことは、少なくとも日本の管理者層においては、特定の性格特性がさまざまな状況を超えて安定的に比較的高い妥当性係数を示すことを示唆している。

次に、誤差分散の説明率に注目する。分散の比であるこの説明率という指標には不安定な側面があることが米国の研究の中で指摘されており（McDaniel, Hirsh, Schmidt, Raju & Hunter, 1986），本研究においてもこの問題については未解決なままである。しかしながら、今回の分析が性格検査、一般知的能力検査とも同一の対象者に対して行われたものであることから、いくつかの誤差要因についてはより統制されているものと考えられる。そこでここでは、あえて数値的な比較を行ってみると、性格検査については8尺度中4尺度で誤差による分散説明率が75%を超える、残りの尺度も1つを除きすべてが60%以上と全体的に高い値を示したのに対し、能力検査では25%～31%と低い説明率にとどまっている。このことは、性格検査については妥当性係数の真値が状況の個別性にかかわらず同程度の水準に収まる可能性が高く、反対に能力検査については妥当性係数の真値が状況

ごとに異なる可能性があることを意味している。

一般知的能力検査に関するこの結果の解釈にあたっては、高橋ら（1994）の研究が一つの示唆を与えてくれる。高橋らは、管理者層を対象にして職務内容別（一般管理、営業、企画研究）に分析を行った結果、知的能力検査の妥当性が一般化されたことを報告している。本研究では職務内容に関する情報が十分には得られなかつたため同様の分析を試みることはできなかつたが、米国における研究（Hunter & Hunter, 1984）において職務の複雑さの違いが一般知的能力検査の妥当性係数に影響を与えることが示されているのを考え合わせると、職務内容の相違がモデレータとして作用している可能性も十分考えられる。

一方、性格検査については日本におけるメタ分析の先例がほとんどなく、あくまで米国の研究結果に基づいて推測するしかないが、まじめさや意欲・気配り等による行動が重視されるという日本の特殊な環境がここでも影響しているのではないかと考えられる。つまり日本においては、企業風土や職務内容の違いを超えて一般的に求められる管理者としての姿勢や態度が存在し、それが評価に強く反映されたために今回の結果が得られたという可能性も考えられるのである。

本研究の結果は、さまざまな知見をもたらすものであった。しかしながら、結果の解釈にあたって提示してきたいくつかの仮説は現時点ではあくまで推測にすぎず、それを検証していくためには今後さらなる研究を積み重ねることが必要である。今後の課題として以下の4点をあげておきたい。

第1に、本研究で対象とした個別妥当性研究の数は24であり、この種のメタ分析としては必ずしも十分とはいえない。引き続きデータを蓄積し、より頑健性の高いメタ分析を行っていくことが大切である。第2に、本研究ではデータ上の制約のため実施できなかつたが、職務内容や職場環境あるいは経営・人事管理に関する風土の違いなど、妥当性係数および妥当性一般化に影響を与える可能性があるモデレータ変数について研究していくことが求められる。第3に、予測変数としての性格検査の性質の違い、すなわち特定対象の適性把握を目的とした検査を用いた場合とより一般性の高い検査を用いた場合とで、妥当性係数の値が異なるかどうかについて確認することが重要である。第4に、本研究においてともに有効な妥当性が確認された性格検査、一般知的能力検査に関して、両者を総合したときの増分妥当性（Incremental Validity）が一般にどの程度の値を示すのかについての研究が必要だと考えられる。性格検査と一般知的能力

検査との間の相関は一般に低いことから、妥当性係数はある程度高まることが期待されるが、この問題についても今後の研究が求められる。

[注]

- 1) この尺度はNMATでは「承認-自律」という形で得点が算出されるが、妥当性が「承認」の方向に作用することが多いことが過去の分析を通じて確認されているため、ここでは解釈を容易にするために尺度を反転させて「自律-承認」としている。

引 用 文 献

- Barrick, M. R. & Mount, M. K. 1991 The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R. & Mount, M. K. 1993 Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111-118.
- Ghiselli, E. E. 1966 *The validity of occupational aptitude tests*. New York: Wiley.
- Gough, H. G. 1984 A managerial potential scale for the California Psychological Inventory. *Journal of Applied Psychology*, 69, 233-240.
- Guion, R. M. 1965 *Personnel testing*. New York: McGraw-Hill.
- Hogan, J. & Hogan, R. 1986 *Hogan Personnel Selection Series Manual*. Minneapolis, MN: National Computer Systems.
- Hough, L. M. 1992 The "big five" personality variables — Construct confusion: Description versus prediction. *Human Performance*, 5 (1 & 2), 139-155.
- Hough, L. M. & Schneider, R. J. 1996 Personality traits, taxonomies, and applications in organizations. In K. R. Murphy (Ed.), *Individual Differences and Behavior in Organizations*. San Francisco: Jossey-Bass.
- Hunter, J. E. & Hunter, R. F. 1984 Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J. E. 1986 Cognitive ability, cognitive

- aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J. E., & Schmidt, F. L. 1990 *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- James, L. R., Demaree R. G., Mulaik, S. A., & Mumford, M. D. 1988 Validity generalization: Rejoinder to Schmidt, Hunter, and Raju (1988). *Journal of Applied Psychology*, 73, 673-678.
- Kemery, E. R., Mossholder, K. W., & Roth, L. 1987 The power of the Schmidt and Hunter additive model of validity generalization. *Journal of Applied Psychology*, 72, 30-37.
- McDaniel, M. A., Hirsh, H. R., Schmidt, F. L., Raju, N. S., & Hunter, J. E. 1986 Interpreting the results of meta-analytic research: A comment on Schmitt, Gooding, Noe, and Kirsch (1984). *Personnel Psychology*, 39, 141-148.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. 1990 Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335-354.
- Mischel, W. 1968 *Personality and assessment*. New York: Wiley.
- Motowidlo, S. J., & Van Scotter, J. R. 1994 Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.
- 二村英幸, 1998 人事アセスメントの科学, 東京: 産能大学出版部。
- Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. 1983 Statistical power of tests of the situational specificity hypothesis in validity generalization studies: A cautionary note. *Journal of Applied Psychology*, 68, 115-122.
- Pease, P. W., & Switzer III, F. S. 1988 Validity generalization and hypothetical reliability distributions: A test of the Schmidt-Hunter procedure. *Journal of Applied Psychology*, 73, 267-274.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. 1980 Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373-406.
- Raju, N. S., Anselmi, T. V., Goodman, J. S., & Thomas, A. 1998 The effect of correlated artifacts and true validity on the accuracy of parameter estimation in validity generalization. *Personnel Psychology*, 51, 453-465.
- Sackett, P. R., Harris, M. M., & Orr, J. M., 1986 On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error. *Journal of Applied Psychology*, 71, 302-310.
- Schmidt, F. L., & Hunter, J. E. 1977 Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540.
- Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. 1980 Validity generalization results for computer programmers. *Journal of Applied Psychology*, 65, 643-661.
- Schmidt, F. L., Hunter, J. E., & Caplan, J. R. 1981 Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, 66, 261-273.
- Schneider, R. J. & Hough, L. M. 1995 Personality and industrial/organizational psychology. In C. L. Cooper & I. T. Robertson(Eds), *International Review of Industrial and Organizational Psychology*, Vol. 10, 75-129. Chichester, UK: Wiley.
- 高橋 潔・西田直史, 1994 知的能力検査に関する妥当性一般化. 産業・組織心理学研究 第8巻, 第1号, 3-12.
- (平成12年2月10日受稿, 平成12年3月10日受理)