

一般企業人を対象とした 性格検査の妥当性のメタ分析と一般化

株式会社リクルートマネジメント 都澤真智子・二村英幸
ソリューションズ組織行動研究所 今城志保・内藤淳

Meta-analyses and validity generalization studies of a personality test
with salaried workers

Machiko TOZAWA, Hideyuki NIMURA, Shiho IMASHIRO, and Jun NAITO
(RECRUIT MANAGEMENT SOLUTIONS CO., LTD.)

Institute for Organizational Behavior Research)

This study reports the results of two meta-analyses of validation studies of a personality test using performance appraisals as criteria. The purpose of this study is twofold. Firstly we seek out the general levels of personality validity coefficients with Japanese samples and their generalizability across situations such as different jobs or organizations. Secondly we address issues of aggregating studies with different criteria, study purposes, and study periods. Five out of seventeen scales showed corrected validity coefficients higher than .10 in absolute value, with the highest coefficient of .21 on "Vitality". The results were similar with those of the previous meta-analyses on U.S. and European samples. Our second meta-analysis included only the studies that met the all three conditions of being conducted within a certain periods of time, using the same criterion, and being confined to research purposes. As a result, the validity coefficients became higher and more generalizable compared to the initial analysis. It is argued that the differences among studies that were often not dealt with in preceding meta-analyses have a significant effect.

Keywords: meta-analysis; personality test; validity generalization; validity.

I. 問題の背景

企業経営における人材の性格的側面に対する関心は、多様な局面で多様に展開されている。人材配置の場面における職務や職場との適合性の検討、採用選考や昇格昇進選考の場面における性格特性の職務遂行能力との関連性に対する関心、精神的健康管理の場面における精神的病理の診断などである。そこでは性格特性の個人差をとらえるために性格検査が適用され、人格の理解や診断を経て人事判断が下されている。

こうした性格検査の基本的な技術は1900年代の前半に確立されたものであるが、当初よりその適用の有効性を検討するためには妥当性(Validity)の概念が用いられてきている。妥当性の概念が提唱されたのは1921年のCourtisに溯り、さまざまな議論が繰りひろげられてきたが(Cronbach, 1988; 池田, 2001; Linn, 1989; 柴山, 2000), 最近では『そのスコアを利用するときの解釈がそれまで

積み重ねられてきた証拠や理論によって、どれだけ支持されているかを示す度合』(American Educational Research Association, American psychological Association, and the National Council on Measurement in Education, 1999)と定義されるに至っている。焦点が当たられるのは、テストの内容、テストの内部構造、テストの実施・採点・解釈の過程、他の変数との関連性、他の変数との並行性や特異性、基準変数との関連性などで、これらに関する証拠や理論をもとにしながら、検査尺度の是非が議論される(Validity argument)。

これらの証拠の中では基準変数との関連性が最も有力で関心が持たれやすいものである。人事評価や業績評価などの基準変数と性格検査尺度(予測変数)との関係に注目し、その相関係数の水準によって選抜の有効性が議論される。この基準変数と予測変数との相関係数は特に妥当性係数と呼ばれ、定量的な評価が可能であることなどから客観的な証拠としての意義は大きい。それは単に

尺度得点と職務行動との関連性に関する議論にとどまらず、人格理解を深める契機となりうるものと言えよう。

妥当性研究は個別の企業や職務における検証にとどまらず、蓄積されてきた個別研究を検索、収集しメタ分析がなされるに及んでいる。妥当性係数の全体的な傾向に関するメタ分析にはつぎの2点のテーマが存在する。すなわち個々の研究によって報告された複数の妥当性係数を集約して真の水準を推定することと、その真値が状況を超えた一般性を備えているかどうかを検証することの2つである。これらを分析するための統計的な手続きがSchmidt & Hunter (1977) によって提唱され、その後それを適用した妥当性研究が多数報告されるに至っている。このメタ分析の証拠などをもって、同種の検査尺度の妥当性係数を新たな検証をせずに他の状況に適用することができるかを検討することは、「妥当性の一般化(Validity generalization)」の議論と呼ばれている。

妥当性に対するアプローチは以上のように整理することができるが、日本においてはその重要性が指摘されながらも先行研究は米国に比べて多くない。そこには日米各々における社会的な背景が存在する。

まず、米国では1964年に施行された公民権法を契機に雇用における人権擁護の施策がとられた結果、性格検査などによる選抜の正当性に関する証拠の提出が求められるところとなり、その有力な証拠の一つとして妥当性係数が利用された。この間の事情は高橋 (1994)、大沢 (2000) に詳しく述べられている。

一方、日本の経営人事においては、臨床用の性格検査が一部転用されつつ固有の質問紙法によるツールが開発され、1970年代以降一定の普及をみている (大沢, 1986; 横原, 2000)。しかしながら、わが国では性格検査が能力検査のように選抜基準として利用されることはあるが、主に人格理解を支援するツールとして利用されることが多い (二村・国本, 2002)。そこでは人格理論や尺度構成、測定結果の解釈法などの有効性に関する視点が重視され、妥当性係数による実証に関心が向けられにくい事情がある。このため、性格検査の妥当性は研究対象にされなかったり、研究されたとしても広く開示されなかったりするなどの状況で、検査開発事業者の頒布活動と実務的な効用感の蓄積によって普及が進んできているように思われる。

このように日米における社会的な背景や人間観の違いが妥当性に対するアプローチを特徴づけているものと理解できるが、このことは、妥当性に関する実証的な研究が日本の社会状況において不必要であることを意味するものではない。実証的な情報の蓄積に基づいてこそ検査得点の意味が職務行動との関連で明確にされるのであ

り、妥当性係数は法的な必要の有無に関わらず妥当性の議論における有効な証拠である。開示されている実証研究の少なさこそ懸念されるべきものであろう。

II. 研究の目的

性格検査の妥当性に関するメタ分析は、1990年代に米国を中心として数多く行われた。その結果、状況の個別性を超えて得られる妥当性係数の平均値や、妥当性に影響を与える要因に関する研究が進み、分析手法や観点などはひととおり確立された感がある。従来、米国では、Guion & Gottier (1965) やMischel (1968) に代表されるように、性格検査と職務遂行能力 (Performance) の関連性は非常に弱く性格検査を予測変数とするのは有効ではないとされていた。それが1990年代前半にBarrick & Mount (1991) が性格特性主要5因子モデル (Big5) を用いてメタ分析を行った結果一定の方向性が見出され、それ以後再び盛んに研究され始めたという経緯がある。

妥当性に関するメタ分析には、前述したように2つのテーマが存在する。第1は個々の研究によって報告された複数の妥当性係数を集約して真の水準を推定すること、第2はその真値が状況を超えた一般性を備えているかどうかを検証することである。欧米における性格検査の妥当性に関するメタ分析の結果を総括すれば、第1の視点である妥当性係数の水準については一定以上の値を示すことが広く認められたのに対し、第2の視点については妥当性は一般化されるものではなく、性格検査の妥当性は職務特性や企業特性などの状況によって異なるという方が大方の見解となっていると考えられる。

欧米における代表的な先行研究としては、Barrick & Mount (1991) がBig5を用いて性格検査に関する妥当性研究のメタ分析を行い、Conscientiousness (勤勉性、誠実性) がさまざまな職種平均で.22 (補正值) の妥当性係数を示すことを報告している。欧州においては、Salgado (1997) がConscientiousnessについて.15 (補正值)、Emotional Stability (情緒安定性) について.13 (補正值) を報告している。また、Judge & Bono (2001) は、Self-esteem, Self-efficacy, Locus of Control, Emotional Stabilityの4つの枠組みにおいて、職務遂行能力を基準とした際の妥当性係数が.19~.26 (補正值) であることを報告している。

一方、妥当性係数の水準が一般化されるかどうかという観点については、Conscientiousnessがさまざまな職務の違いを超えて一定水準の妥当性係数を安定的に示すというBarrick & Mount (1991) の報告もあるが、妥当性係数の真値が一つに定まるという厳密な意味での一般

化の証拠を示すものはSalgado (1997) における Emotional Stability以外には先行研究に見あたらない。

日本では欧米に比べて性格検査に関する妥当性研究は数が少なく（木島・大内・渡辺,2002; 二村・今城・内藤,2000; 二村,2000; 大庭・森,2000; 関本,1978; 若林,1991），メタ分析に関してはさらに限られる（二村ら,2000; 大沢ら,2000）。二村ら（2000）は、管理者層を対象にした性格検査のメタ分析を行い、「調整一統率」が.31,「維持一変革」が.25,「心情一理性」が.18,「内向一外向」が.18,「慎重一大胆」が.13（以上、いずれも補正値）という結果を報告している。また、一般企業人を対象にした性格検査の妥当性研究に関しては、大沢ら（2000）が追跡的な妥当性研究のレビューを行い、「活動意欲」における.16（妥当性係数の単純平均値）が17尺度中最も高い値であったことを報告している。ただし、この研究は欧米のメタ分析とは手法を異にするものであり、同列に比較することは難しい。このような状況を踏まえたとき、日本において一般企業人を対象にした場合に性格検査の妥当性係数がどの程度の水準を示すのか、その妥当性係数は一般化することが可能かどうかをメタ分析によって確認することの意味は大きい。

ところで、性格検査の妥当性に関するメタ分析については、手続き上の問題点がいくつか指摘されている。第1に予測変数の扱い方の問題がある。先行研究では、予測変数をBig5などの枠組みに集約するために予測変数の構成概念からのずれが生じて予測力が低減したり、妥当性の一般化が阻害されている可能性が考えられる。先行研究においては、予測変数は必ずしもBig5そのものを測定する性格検査のみが用いられているわけではなく、Big5の概念に近いさまざまな尺度を統合することで、何らかの関係が導き出せるかどうかに着目したものである。しかし、Hurtz & Donovan (2000) は、Big5と職務遂行能力の関係性を分析するのであれば、Big5を測定概念とする検査のみでメタ分析を行うべきであり、さまざまな概念を測定している検査が混在しているために、職務遂行能力との関連を見極めることができ難くなっていることを指摘している。この個別研究間における予測変数の構成概念からのずれの問題は、メタ分析のごく初期の段階すでにSchmidt & Hunter (1977) によって、妥当性係数に混在する人為的な誤差（Artifacts）の一つとして示されているものであるが、できるだけ多数の個別研究を統合したいというメタ分析の実務上の要請からこれまであまり焦点が当てられてこなかった事情がある。もし、Big5よりも詳細な概念を測定する同一の予測変数を使用した個別妥当性研究に絞ってメタ分析を行うことができれば、先行研究とは異なる結果が得られる可能性

がある。

第2の問題点として、基準変数に関しても予測変数と同種のことが指摘できる。Barrickら（1991）やHough (1992) は、先行研究において妥当性が一般化されていない理由の一つとして基準変数の性質の問題をあげている。すなわち、職務遂行能力評価、研修評価、人事考課など基準変数の種類ごとに区分して分析を行えば、結果が異なる可能性があるという指摘である。基準変数として研究目的で集められた評価を用いる場合には、現実場面での評価を用いる場合と比べて妥当性係数が大きくなるという議論もあり（McDaniel, Whetzel, Schmidt & Maurer, 1994），基準変数の内容を統制することによって先行研究とは異なる傾向が確認される可能性が指摘できる。

第3に個別妥当性研究が実施された時期や、研究デザインの違いの問題があげられる。時期の問題については、Schmitt & Chan (1998) が指摘するように、研究間で5年、10年の開きがある場合には、時代背景の影響を受ける可能性があるため、基準変数をとらえる際に職務遂行能力の変化を考慮しなければならない。また、研究デザインの問題については、予測変数と基準変数のデータ収集時期が異なる予測的妥当性の研究と両者の時期が同時に近接している併存的妥当性の研究、あるいは、同じ予測的妥当性の研究でも入社5年後の従業員を対象とする研究と入社10年後の従業員を対象とする研究では、結果として得られた妥当性係数の水準にずれが生じている可能性がある。これらの点についても、もし対象とする個別研究を条件を満たすものに絞ってメタ分析を行うことができれば、先行研究とは異なる結果が得られることが予想される。

以上のような問題意識に立ち、本研究では一般企業人を対象とした場合に、性格検査の職務遂行能力に対する妥当性係数がどの程度の水準および一般化の程度を示すかを確認することを第1の目的としている。加えて、分析対象とする個別妥当性研究を、基準変数、研究目的、研究時期の点で同じ条件のものだけに限定することにより、先行研究に関して指摘されている手続き上の問題点を統制した場合に、そのことが妥当性係数の水準および一般化の程度にどのような影響を与えるかについて検証を行うことを第2の目的とする。

III. 方法

分析 1

＜分析対象とした個別妥当性研究＞

本研究においてメタ分析の対象とした33の個別妥当性研究は、1991年3月から2000年12月までの期間にいずれ

も企業における採用選考や従業員の昇進昇格選考実務場面で実施された、あるいは研究目的で実施されたものである。のべ標本人数は5844名で、1社あたりの人数は52名から1090名、平均177名である。なお、33の個別妥当性研究のうち21は大沢ら(2000)が分析に用いたものと同一である。

<予測変数>

予測変数に用いられているのは、いずれも総合検査SPI(Synthetic Personality Inventory: 株式会社リクルートマネジメントソリューションズ)の性格検査部であり、17尺度で構成されている。それぞれの信頼性係数(Cronbachの α 係数)を表1に示す。

表1 性格検査の信頼性係数 ($n=85,500$)

尺度名	α 係数
社会的内向性	.91
内省性	.86
身体活動性	.89
持続性	.87
慎重性	.87
達成意欲	.84
活動意欲	.87
敏感性	.89
自責性	.89
気分性	.85
独自性	.80
自信性	.84
高揚性	.80
内向-外向	.88
直観-感覚	.88
感情-思考	.83
知覚-判断	.82

<基準変数>

基準変数に用いられているのは、現実場面における人事考課と研究目的で収集した調査票による評価の2種であり、どちらが用いられているかは個別妥当性研究によって異なる。このうち、調査票による評価は、直属上長が対象者の職務遂行能力を1項目6段階で評定したものであり、すべて共通の評定項目を用いている。

<分析の手法>

メタ分析の手続きについては、広く用いられているHunter & Schmidt(1990)に従う。この手続きは、先行研究であるBarrick & Mount(1991)や二村ら(2000)の用いた手続きと同様である。妥当性一般化の考え方はSchmidt & Hunter(1977)が唱えたものであるが、それによれば、妥当性係数は個々の研究でかなり異なる値をとることが多く、そのばらつきには人為的な誤差が含まれていると考えられている。これらを完全に除去することができれば、妥当性係数の真値は、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1

つに定まる可能性がある。人為的な誤差として、Schmidt, Gast-Rosenberg & Hunter(1980)はつきの7種類をあげている。1) 標本誤差、2) 基準変数の測定誤差、3) 予測変数の測定誤差、4) 予測変数の範囲制限、5) 基準変数の構成概念からのずれや研究間での違い、6) 研究者の計算、タイピング、転記の間違い、7) 予測変数の構成概念からのずれや研究間での違い、である。そのうち、5) から7)までの誤差については影響を取り除くことはできないが、1)から4)までは統計的な手法により影響を取り除くことができるとしている。妥当性係数の値の分散の大半がそれら誤差により説明できる場合には、妥当性が一般化されたものとみなされる。その判断基準として、Schmidtら(1980)は「75%ルール」を提示している。妥当性係数の分散のうち誤差に起因する部分の割合が75%以上ならば妥当性は一般化できるとするものである。

またこの基準に加え、80%確信区間(80% Credibility Interval; 80% CV)を用いて、推定された妥当性係数の真値がどの範囲の区間に収まるかどうかを検証する。具体的には、推定された妥当性係数の真値の分布のそれぞれ上方・下方から10パーセンタイルにあたる値が揃って0を含まず正もしくは負で一定水準の値となっていれば、状況の違いを超えて意味のある値をとるものと考える(Schmidt, Hunter & Caplan, 1981)。

<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、最後に企業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(σ_{ρ}^2)」を求める。

なお、以下の式における記号はそれぞれつきの意味を表す(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標準偏差] / [範囲制限を受けた予測変数の標準偏差]
- r_i : 補正前の妥当性係数
- 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
- σ_{ei}^2 : 補正前の妥当性係数の標本誤差分散
- σ_{eci}^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
- σ_{ρ}^2 : 妥当性係数の真値の分散

一般企業人を対象とした性格検査の妥当性のメタ分析と一般化

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

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

2) 補正後の標本誤差分散 (σ_{eci}^2) の算出

$$\sigma_{eci}^2 = \alpha_i A_i \sigma_{ei}^2$$

ただし、

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

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

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

3) 真の妥当性係数の推定値 ($\hat{\rho}$) の算出

$$\bar{r}_c = \sum n_i r_{ci} / N$$

4) 真の妥当性係数の分散 (σ_{ρ}^2) の算出

$$\sigma_{\rho}^2 = \sigma_r^2 - \sigma_e^2 = \bar{\sigma}_{rc}^2 - \bar{\sigma}_{ec}^2$$

ただし、

$$\sigma_r^2 = \bar{\sigma}_{rc}^2 = \sum n_i (r_{ci} - \hat{\rho})^2 / N$$

$$\sigma_e^2 = \bar{\sigma}_{ec}^2 = \sum n_i \sigma_{eci}^2 / N$$

用いられた予測変数は標準化された検査であるため、信頼性係数および範囲制限について把握することができた。範囲制限の比率を示す1/Uの範囲は0.70~1.35で、いずれのデータも範囲制限の影響はそれほど大きく受け

ていないといえる。

一方、基準変数については、調査票を用いたものに関しては職務遂行能力に関する1項目による評定であったため信頼性係数が算出できず、また、基準変数に人事考課を用いたものについても信頼性係数に関する情報は得られなかったため、先行研究を参考にして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.60を超えないとしている。また、Viswesvaran, Ones & Schmidt (1996)が行ったメタ分析では、上長による職務遂行能力評価の信頼性係数は.52であると報告されている。本研究では過度の補正を避けるというスタンスをとり、.60を用いて計算を行う。

分析 2

分析1で用いた33の個別妥当性研究のうち、2000年7月から12月までの期間に実施された12の個別妥当性研究のみを対象とする。手続きは分析1と同じであるが、基準変数、研究目的、研究時期を限定している。すなわちこれらの個別妥当性研究は、いずれも研究目的で実施されたものであり、基準変数にはすべて同一の調査票、同一の評定項目を用いている。評価は、直属上長が対象者の職務遂行能力を1項目6段階尺度で行っている。のべ標本人数は1607名で、1社あたりの人数は72名から209名で平均134名であった。なお、予測変数の範囲制限の比率を示す1/Uは0.83~1.32であった。

表2 分析1結果

	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_r^2	σ_e^2	誤差による 分散説明率	80% 確信区間
社会的内向性	-.08	-.12	.0151	.0231	.0080	35%	-.28~ .04
内省性	-.03	-.04	.0176	.0253	.0077	30%	-.21~ .13
身体活動性	.09	.13	.0134	.0211	.0077	36%	-.02~ .27
持続性	-.02	-.02	.0137	.0212	.0075	35%	-.17~ .13
慎重性	-.06	-.08	.0111	.0185	.0074	40%	-.22~ .05
達成意欲	.09	.12	.0282	.0359	.0077	21%	-.09~ .34
活動意欲	.15	.21	.0100	.0172	.0071	42%	.08~ .34
敏感性	-.03	-.04	.0112	.0191	.0079	41%	-.18~ .09
自責性	-.06	-.08	.0172	.0248	.0076	31%	-.25~ .09
気分性	-.02	-.03	.0229	.0311	.0082	26%	-.22~ .17
独自性	-.03	-.05	.0074	.0157	.0083	53%	-.16~ .06
自信性	.05	.07	.0136	.0213	.0077	36%	-.08~ .22
高揚性	.05	.07	.0065	.0144	.0079	55%	-.03~ .17
内向→外向	.09	.11	.0233	.0309	.0076	25%	-.08~ .31
直観→感覚	-.00	-.00	.0073	.0154	.0081	52%	-.11~ .11
感情→思考	.07	.01	.0112	.0191	.0078	41%	-.04~ .23
知覚→判断	-.00	.03	.0306	.0383	.0077	20%	-.23~ .22

注) すべての尺度において、総標本数は5844、妥当性研究数33であった。

表3 分析2結果

	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_{ϵ}^2	σ_e^2	誤差による 分散説明率	80% 確信区間
社会的内向性	-.11	-.16	.0023	.0126	.0104	82%	-.23～-.10
内省性	.06	.09	.0193	.0294	.0101	34%	-.09～.26
身体活動性	.08	.11	.0011	.0117	.0106	91%	.06～.15
持続性	-.01	-.02	.0168	.0266	.0097	37%	-.18～.15
慎重性	-.05	-.08	.0085	.0187	.0102	55%	-.20～.04
達成意欲	.14	.18	.0170	.0258	.0088	34%	.01～.35
活動意欲	.20	.27	.0000*	.0066	.0093	100%*	.27～.27
敏感性	-.09	-.12	.0132	.0236	.0104	44%	-.27～.03
自責性	-.13	-.17	.0090	.0186	.0096	52%	-.29～-.05
気分性	-.06	-.08	.0316	.0415	.0099	24%	-.31～.15
独自性	-.02	-.03	.0114	.0216	.0103	47%	-.17～.10
自信性	.14	.19	.0008	.0105	.0097	92%	.15～.22
高揚性	.03	.03	.0091	.0190	.0100	51%	-.09～.16
内向一外向	.09	.13	.0066	.0170	.0104	61%	.02～.23
直観一感覚	-.01	-.02	.0073	.0169	.0096	57%	-.13～.09
感情一思考	.12	.16	.0154	.0247	.0093	38%	.00～.32
知覚一判断	.04	.06	.0226	.0330	.0104	32%	-.14～.25

注)すべての尺度において、総標本数は1606、妥当性研究数12であった。

*「活動意欲」の σ_{ρ}^2 および誤差による分散説明率は、計算では-.0027, 141%であった。

V. 結果

表2は33の個別妥当性研究にもとづく分析1の結果である。この表には、性格検査の各尺度ごとに、補正前の平均妥当性係数、補正後の平均妥当性係数(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_{ϵ}^2)、誤差による分散(σ_e^2)、誤差による分散説明率、80%確信区間が示されている。

まず、推定された妥当性係数の真値($\hat{\rho}$)に注目する。もっとも値の高い尺度は、「活動意欲」の.21であった。また、「身体活動性」が.13、「社会的内向性」が-.12、「達成意欲」が.12、「内向一外向」が.11であり、妥当性係数の大きさの絶対値が.10以上だったのは17尺度中5尺度であった。

次に誤差による分散説明率に注目すると、75%以上の尺度はなかった。これは妥当性係数の分散のうち、誤差では説明できないものも一定以上の割合が含まれていることを意味しており、真の妥当性係数の水準は状況の違いにより変動する可能性を示している。ただし、「活動意欲」の80%確信区間は.08～.34と0を含まない値となっており、「活動意欲」の妥当性係数は状況の違いを超えて安定して正の値をとることが認められた。

続いて表3は12の個別研究を対象にした分析2の結果である。ここで用いた12の個別研究は、研究目的で検査実施と評価情報の収集を半年の間に行ったもので、基準変数にはすべて同じ調査項目からなる調査票が用いられて

いる。

先ほどの分析結果と同様に、推定された妥当性係数の真値($\hat{\rho}$)に注目すると、もっとも値が高かったのは、「活動意欲」の.27であり、続いて「自信性」が.19、「達成意欲」が.18、「自責性」が-.17、「社会的内向性」が-.16、「感情一思考」が.16となっている。妥当性係数の大きさの絶対値が.10以上だったのは17尺度中9尺度であり、そのうち6尺度は.15以上であった。

次に誤差による分散説明率に注目すると、75%以上の尺度は「活動意欲」が100%、「自信性」が92%、「身体活動性」が91%、「社会的内向性」が82%であった。これらの性格特性については、妥当性係数の分散の多くが誤差によるものであり、真の妥当性係数は状況の違いに関わらず1つに定まる可能性を示している。

さらに80%確信区間に注目すると、17尺度中7尺度が0を含まない値となっている。「活動意欲」(.27～.27), 「自信性」(.15～.22), 「身体活動性」(.06～.15), 「社会的内向性」(-.23～-.10), 「内向一外向」(.02～.23), 「自責性」(-.29～-.05), 「達成意欲」(.01～.35)の7尺度であり、これらの性格特性については状況の個別性を超えて妥当性係数が安定して正の値を示す、あるいは負の値を示す尺度であることが認められた。

V. 考察

まず分析1の結果から、日本の一般企業人を対象とした場合の性格検査の妥当性の水準と妥当性係数の状況の違いを超えた安定性の程度について、欧米の先行研究と

の比較を行いながら考察する。

妥当性係数の水準について表2の補正後の妥当性係数をみると、もっとも高い値を示した尺度は「活動意欲」の.21である。「活動意欲」は行動が機敏で意欲的である程度を測定している尺度で、Big5ではExtraversion（外向性）の一部と考えられる。Extraversionと職務遂行能力との妥当性係数は、Barrickら（1991）が.13（補正值）、Salgado（1997）は.08（補正值）を報告している。Hough（1992）は、ExtraversionをAffiliation（社交性）とPotency（支配性）に分け、それぞれ.00（補正なし）、.10（補正なし）を報告している。「活動意欲」は補正值で.21（補正なし.15）を示し、上記の先行研究で得られた結果よりも高い水準である。次に高い値を示した「身体活動性」は、「活動意欲」と関連が深く、体を動かすことを好む傾向を測定している。補正值.13（補正なし.09）は、先行研究とほぼ同等の水準といえる。3番目に高い値を示したのは「達成意欲」と「社会的内向性」であった。「達成意欲」は目標を高く掲げやりがいを求める傾向を測定している。先行研究ではHough（1992）のAchievement（達成欲求）に対応し、またBig5のConscientiousness（勤勉性、誠実性）と関連が深い。Hough（1992）のAchievementは.15（補正なし）、Barrickら（1991）とSalgado（1997）のConscientiousnessはそれぞれ.22（補正值）、.15（補正值）であるのに対し、「達成意欲」は補正值.18（補正なし.14）であり、ほぼ同等の水準となっている。一方、「社会的内向性」は社交意欲の程度を測定する尺度で、補正值-.12（補正なし-.08）が得られた。また、「社会的内向性」との関連が深い尺度である「内向一外向」についても補正值.11（補正なし.09）が確認された。この2つの尺度は、概念的にHough（1992）のAffiliationに近く、Big5ではExtraversionの一部であると考えられる。Hough（1992）はAffiliationの妥当性係数.00（補正なし）を報告しており、Extraversionについては、Barrickら（1991）が.13（補正值）、Salgado（1997）が.08（補正值）を報告している。今回の結果はこれらとほぼ同等の水準であるといえる。なお、「社会的内向性」は対人面における消極性を測定する尺度であるため、先行研究とは正負の符号が反転している。

妥当性係数の状況の違いを超えた安定性に関しては、75%ルールに照らし合わせると誤差による分散説明率が75%を超えた特性はなかった。80%確信区間でみると、「活動意欲」で.08～.34と0を含まない結果となった。このことから、職種や企業特性などの状況の違いによって妥当性係数の値は変動するものの、「活動意欲」については0を超える妥当性が安定して得られることが示唆さ

れた。「活動意欲」はBig5のExtraversionと関連が深いが、このExtraversionにおいてBarrickら（1991）は90%下方限界信頼値-.01を、Salgado（1997）は-.08を報告している。90%下方限界信頼値は、今回算出している80%確信区間の下方値（係数の値が負である尺度は上方値）と同義である。これらの先行研究の下方値も大きく0を下回るものではないが、本研究では.08と0を超える値となった。

以上のように比較を行うと、概して本研究での妥当性係数や確信区間の下方値は、先行研究と比べて同等かやや高い値となる傾向が見られた。このことの考えられる理由の一つとして、前述した予測変数が特定の検査に統一されたことがあげられる。すなわち、同一の検査を使用することで予測変数間での構成概念のずれがなくなり、結果的に妥当性の水準は高く、また変動は小さくなつたと考えられる。

一方、水準にはやや違いが見られたものの、本研究で比較的高い妥当性係数が得られた尺度は、先行研究であるメタ分析と同様の特性を測定する尺度であった。このことから、職務遂行能力と関係のある性格特性は日本と欧米で大きな相違はないと推測される。職務を中心に成り立っている欧米の人事慣行と人を中心とする日本の人事慣行の違いを指摘する意見もあるが（大沢、2000），本研究の結果では少なくとも性格特性と職務遂行能力の関係に関しては日米で類似した傾向が確認された。

次に分析2の結果より、メタ分析に使用する研究を基準変数、研究目的、研究の時期が同じものに限定することで、分析結果に差がみられるかについての考察を行う。

表3（分析2）と表2（分析1）を比較すると、もっとも高い値を示した補正後の妥当性係数はいずれも「活動意欲」で、研究を限定した場合（分析2）では.27、しない場合（分析1）では.21と前者のほうが値が大きい。また、限定しない場合（分析1）は妥当性係数の絶対値が.10以上の尺度は5尺度であったが、研究を限定した場合（分析2）では9尺度である。さらに、研究を限定した場合（分析2）では、妥当性係数の絶対値が.15以上の尺度が6尺度あり、変数の大半で妥当性係数の値が限定しない場合（分析1）に比べ大きくなっている。

状況の違いを超えた妥当性係数の安定性の観点については、メタ分析に用いる研究を限定しない場合（分析1）には誤差による分散説明率が75%を超えた特性はなく、80%確信区間に0を含んでいないのは「活動意欲」だけであった。これに対して、研究を限定した場合（分析2）には4尺度において誤差による分散説明率は75%を超え、これらの尺度では80%確信区間も0を含まない結果とな

った。すなわちこれら4尺度では産業特性や企業特性、職務特性などの状況個別性を超えて、安定的に一定水準以上の妥当性係数を示すことが確認された。

以上のように、妥当性係数の水準と状況を超えた安定性のいざれに關しても、基準変数、研究目的、研究時期の条件を統制した個別研究を用いた場合（分析2）に結果が顕著になった。これは、分析対象とした個別妥当性研究が基準変数に同一の調査票を用いていたために、Schmidtら（1980）が統計的な補正が難しい誤差としてあげた基準変数の構成概念からのずれや研究間での違いを抑えられたことが一つの原因と考えられる。また、個別研究が実施された時期がすべて半年以内であること、研究デザインもすべて予測変数と基準変数を同タイミングで収集する併存的妥当性の研究で同一であることから、通常のメタ分析ではコントロールすることが困難である多くの誤差が抑えられているということがいえる。この意味で分析2は、妥当性係数に混在する誤差をできるだけ少なくした場合の性格検査と職務遂行能力の関連性の水準を求めることになり、性格特性の妥当性係数としては上限に近い値といえるかもしれない。現実場面でのテストの利用状況を考えるとかなり特殊な条件下での結果であるが、性格特性が職務遂行能力にどの程度関連があるかを知るうえでは本研究の結果は重要な意味をもつ。同様のことは多くの先行研究にもあてはまると思われ、性格特性と職務遂行能力との関係性は、状況を統制した場合には現在考えられているよりも高い水準を示す可能性がある。

他方、分析2において標本誤差、予測変数と基準変数の信頼性の希薄化に加えて、予測変数の構成概念の研究間の違い、基準変数の構成概念からのずれや研究間での違いを抑えた後も誤差による分散が75%に満たなかったり、80%確信区間が0をまたぐ性格特性が存在することは、なお統制されない分散が残っていることを意味している。この残った分散の原因としては、職務特性、業種や産業特性、企業規模、地域特性などが考えられ、これらを統制すれば妥当性係数は状況の個別性を超えてより安定した値を示す可能性も否定できない。この点は、今後さらなる研究が求められる。

VI. 今後の課題

以上の性格特性と職務遂行能力評価の相関関係は、単に性格検査尺度の妥当性の検討にとどまらず、経営人事の実践に対する示唆も提供している。本研究では、職務遂行能力と関連が深い性格特性として、自己を高い目標に動機づけたり関係者に機敏に働きかけたりするなどの行動傾向が示された。それは職務や職場との適合性をあ

らわすだけでなく、企業人一般に求められる実践的能力・スキルの一側面をあらわす要素のように思われる。近年、人材要件として、身体的能力や知的能力以外に情動的知性（Emotional Intelligence）があげられることがあるが（Gardner, 1993; Goleman, 1995など）、その自己統制や対人関係のスキルと符合する一面が感じられる。このように捉えるとき、明らかになった性格特性は単に適合性の観点からではなく、採用選考、配置、能力開発など多様な経営人事場面における評価・育成の基本的な要件と位置づける意義が感じられる。

ただし、現実の採用選考などにおいて性格検査尺度スコアを選抜基準として適用するか否かは、性格が扱いに慎重さが求められる内面の特性であることや受検者に納得感があるかどうかなど他の側面も視野に入れた広範な議論が必要となることは言うまでもない。

本研究の結果の解釈にあたっては、予測変数に用いた総合検査SPIの性格検査17尺度をBig5と関連づけてみたときにAgreeablenessとOpenness to Experienceの2特性に対応する尺度が相対的に少ないと留意が必要である。他の検査尺度を予測変数として用いた同様の妥当性研究が広く蓄積されることが望まれる。また、本研究が扱っているのが1990年代を中心とするデータであり、背景となっている経営環境や社会文化的な様相あるいは経営人事の機軸はこの間大きく変質していることは、さまざまな実態調査で知られている通りである（日本能率協会マネジメントセンター, 2000など）。こうした背景の変化の影響については今後新たな検証を待たねばならない。

本研究では、一般的な傾向を特定するアプローチによって一定の知見が得られたが、このことは同種のデータの蓄積を重ねる必要がなくなったということを意味しない。本研究で用いた標本が欧米の同種の研究に比べて少ないということもちろんあるが、それのみならず妥当性の概念の原義からもその必要性を指摘できる。この点についてはGuion(1992)が、『妥当性に関する証拠は、どこまでも暫定的なもので、たとえ得られた値が感動的なものであったとしても、それは検討の途上であると認識すべきである』(p.375)と述べている。

また得られたデータの解釈に際しては、妥当性の他の側面に関する検討も重要である。つまり妥当性の概念は妥当性係数を重要な証拠としつつも構成概念の確かさ、実施や結果の適用が社会や個人に与える影響、適用の意義や価値などと関連づけた議論も必要であるとされる（Cronbach, 1988）。とりわけ性格検査の日本における適用が、尺度得点による予見よりも人事決定のための人格

理解を確実にする目的が優先される事情を鑑みて、安直に結論づけることなく慎重に議論していく姿勢が求められる。

引用文献

- American Educational Research Association, American psychological Association, and the National Council on Measurement in Education, 1999, *The Standards for Educational and Psychological testing*, American Educational Research Association
- Barrick, M.R. & Mount, M.K. 1991 The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Cronbach, L.J. 1988, Historical and Epistemological Bases of Validity, In Wainer, H. & Braun, H.I. (Eds) *Test Validity*. New Jersey : Lawrence Erlbaum Associates, Inc.
- Cronbach, L.J. 1988, Five Perspectives on Validity Argument, In Wainer, H. & Braun, H.I. (Eds.) *Test Validity*. New Jersey : Lawrence Erlbaum Associates, Inc.
- Gardner, H. 1993, *Frames of mind : The theory of multiple intelligences*. New York : BasicBooks.
- Goldberg, L.R. 1993a The structure of personality traits: Vertical and horizontal aspects. In Funder, D.C., Parke, R.D., Tomlinson-Keasey, C. & Widaman, K. (Eds.), *Studying lives through time: Personality and development*. Washington, DC : American Psychological Association.
- Goleman, D. 1995, *Emotional Intelligence*. New York : Bantam Books. (土屋京子訳 EQこころの知能指數 講談社 1996)
- Guion, R.M. 1992 Personnel Assessment, Selection, and Placement, In Dunnett, M.D. (Ed.) *Industrial and Organizational Psychology Handbook* (2nd. ed. vol.2). Palo Alto, CA : Consulting Psychologist Press.
- Guion, R.M. & Gottier, R.F. 1965 Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.
- Hough, L.M. 1992 The "big five" personality variables—Construct confusion: Description versus prediction. *Human Performance*, 5(1&2), 139-155.
- Hurtz, G.M. & Donovan, J.J. 2000 Personality and Job Performance: The big five revisited. *Journal of Applied Psychology*, 85, 869-879.
- Hunter, J.E. & Schmidt, F.L. 1990 *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA : Sage Publications, Inc.
- 池田央 2001 アセスメント技術から見たテスト法の過去と未来 日本教育工学会誌 24(1), 3-13.
- Judge, T.A. & Bono, J.E. 2001 Relationship of core self-evaluations traits — Self-esteem, generalized self-efficacy, locus of control, and emotional stability—With job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, 86, 80-92.
- 木島伸彦・大内健・渡辺直登 2002 パーソナリティ尺度と医薬情報担当者の営業成績との関連性：気質・性格尺度Temperament and Character Inventory (T C I) を用いて。経営行動科学, 16(2), 151-161.
- Linn, R.L. ed. 1989 *Educational Measurement* 3rd, : Collier Macmillan Publishing Company.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. & Maurer, S.D. 1994 The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599-616.
- Mershon, B. & Gorsuch, R.L. 1997 Number of factors In the personality sphere: Does increase in factors increase predictability of real life criteria ? *Journal of Personality and Social Psychology*, 55, 675-680.
- Mischel, W. 1968 *Personality and assessment*. : John Wiley & Sons, Inc.
- 二村英幸 2000 採用選考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 第4章, 金子書房
- 二村英幸・今城志保・内藤淳 2000 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 経営行動科学, 13(3), 159-167.
- 二村英幸・国本浩市 2002 採用選考ハンドブック 人事測定研究所 (現、株式会社リクルートマネジメントソリューションズ)
- 大庭さよ・森崇王 2000 大卒新入社員の適応に対するビッグ・ファイヴの役割. 産業・組織心理学研究第13巻第1号, 39-53.
- 大沢武志 1986 わが国における企業の採用試験と心理テスト 行動計量学第14巻第1号, 53-59.
- 大沢武志 2000 人事アセスメントの社会性 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 第3章, 金子書房
- Paunonen, S.V. & Adelheid, A.A.M.N. 2001 The

- personality hierarchy and the prediction of work behaviors. In Roberts, B.W. & Hogan,R.(Eds.), *Personality psychology in the workplace*, American Psychological Association.
- Paunonen, S.V. & Ashton, M.C. 2001 Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, 81, 524-539.
- 榎原國城 2000 性格の測定 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 第10章, 金子書房
- Salgado, J.F. 1997 The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- 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., 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.
- Schmidt, F.L., Gast-Rosenberg, I. & Hunter, J.E. 1980 Validity generalization results for computer programmers. *Journal of Applied Psychology*, 65, 643-661.
- Schmitt, N. & Chan, D. 1998 *Personnel selection: A theoretical approach*. Newbury Park, CA : Sage Publications, Inc.
- 柴山直 2000 古典的テスト理論 大沢武志・芝祐順・二村英幸 2000 人事アセスメントハンドブック 第17章, 金子書房
- 高橋潔 1994 米国における採用テスト妥当性の専門的基準と法的規制 日本労働研究雑誌4月号, 51-64.
- Viswesvaran,C., Ones, D.S. & Schmidt, F.L. 1996 Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574.

謝辞

妥当性の定義に関する翻訳は、池田央先生（教育測定研究所）にご指導いただいた。記して感謝を申しあげたい。

（平成16年1月10日受稿、平成16年10月29日受理）