

日本テスト学会 第7回研究会(公開シンポジウム)

『テストの妥当性の概念および検証方法の新たな展開』

人事アセスメントにおける妥当性検証と そのメタ分析

RECRUIT
MANAGEMENT SOLUTIONS

株式会社リクルートマネジメントソリューションズ
組織行動研究所 舩田博之

会社概要

- 社名 株式会社リクルートマネジメントソリューションズ
- 創業 1963年
- 設立 1989年
- 資本金 1億10百万円
- 売上高 164億83百万円(2006年3月期)
- 従業員数 448名
- 代表者 代表取締役社長 釘崎 広光
- 本社所在地 〒108-6108
東京都港区港南2-15-2 品川インターシティB棟 8F

- 2004年10月
(株)HRR(旧人事測定研究所)とリクルートのトレーニング
事業部門を統合、
新会社 (株)リクルートマネジメントソリューションズ 設立

マネジメントソリューションを提供する6つの領域

- 人材採用・戦略化支援
 - 人材開発支援
 - 経営・組織変革支援
 - 戦略実践支援
 - 制度構築・運用支援
 - 営業力強化支援
-
- アセスメント、コンサルティング、トレーニングによるソリューションサービスを展開

人事アセスメントとは

- 人事アセスメントとは
 - 人材や組織の特性を測定し、その特徴を把握すること
 - 人事評価の科学性、実証性を高めるために使用される
- 代表的な人事アセスメントツール
 - 適性テスト
 - 多面観察評価ツール(360°フィードバック)
 - アセスメントセンター方式
 - 面接 …
- 適性テストの利用場面
 - 採用選考
 - 昇格昇進選考

人事アセスメントにおける妥当性

- 米国で先行する妥当性研究
 - 採用選考における差別問題は、人事上の差別をめぐる代表的な問題
 - 米国における雇用差別をめぐる法的規制
 - EEO(Equal Employment Opportunity)の法的規制
 - 基本法 公民権法第7編(Title VII of the Civil Rights Act) 1964年
 - 雇用機会均等委員会(EEOC)による規制や裁定
 - 選考手続きに関する統一ガイドライン(1978年)
 - EEOの法的規制の考えを反映
 - 適性テストを実施する場合、利用する合理的根拠を示さなければならない
 - 合理的根拠=妥当性についての実証的研究成果
 - 積極的にとりあげられる妥当性の一般化
 - 教育・心理テストのスタンダード(APA他;1966年,1974年,1985年,1999年)
 - 科学性に裏付けられた専門的基準としてEEOCが支持
 - 係争事件に発展した場合、法廷の場の判断基準となるスタンダード

人事アセスメントにおける妥当性

- 妥当性の一般化

- 基準関連妥当性への要請

- 追跡的研究に基づく予測的妥当性が優先されることが望ましい
 - 実証的研究に耐えるテストの確保の難しさ

- 妥当性一般化への期待

- ある状況下での基準関連妥当性が新しい状況にも一般化可能
 - 計量的なメタ分析手法の適用により客観的な判断基準が利用可能

【方法論】妥当性一般化の手続き

- メタ分析の手法
 - Hunter & Schmidt(1990)の手続き
 - 妥当性一般化(Validity Generalization)の考え方
 - 観測される妥当性係数のばらつきは人為的な誤差(Artifacts)の混入による
 - これらの阻害要因を除去することができれば、真の妥当性係数は一つに定まる

妥当性係数に混入する誤差

- 混入する人為的な誤差

- 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(Range restriction)

- 5) 基準変数の構成概念からのずれ
(Contamination & Deficiency)

- 6) 研究者の計算、タイピング、転記ミス

- 7) 予測変数の構成概念からのずれ

- 1)~4)については統計的手法によって除去することが可能

【参考】妥当性係数の補正

- 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

$$r_{ci} = r_i' / \sqrt{\rho_{XX}} \sqrt{\rho_{YY}}$$

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
2006/9/9 希薄化の補正後の妥当性係数

【参考】誤差分散の算出

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

、 $A_i = r_{ci} / r_i$

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

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

【参考】妥当性係数の分散の算出

- 真の妥当性係数の推定値 ($\hat{\rho}$) の算出

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

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

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

妥当性一般化の検証

- 職務や企業の違いを超えて妥当性係数が安定した値を示すといえるか？
- 状況個別性 (Situational Specificity)
VS.
妥当性一般化 (Validity Generalization)
- Hunter & Schmidt (1990) の手法に従う

妥当性一般化の検証

- 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

- 75%ルール(Schmidt, Gast-Rosenberg & Hunter 1980)

$$\sigma_e^2 / \sigma_r^2 \geq 75\%$$

– 観測された妥当性係数の分散の大部分が1)~4)によって説明できる場合に一般化されたと判断する

→ 状況個別性の棄却(絶対的な根拠に基づくというよりも、一つの判断基準)

- 90%信頼限界下方値(80%信頼区間)

– この値が一定の水準を示せばその妥当性は異なる状況への適用可能性を持つ(Transportability)

→ 意味のある妥当性係数の水準(一般化の適用可能性)

米国における妥当性研究の動向

【性格テストの妥当性】

- 米国における従来の見解
 - 職務遂行能力との関連は薄く、予測変数としては現実的ではない (Guion, 1965; Mischel, 1968)
 - 近年の動向 (1980～1990年代)
 - BIG5
 - メタ分析
- 性格テストの復権
- 代表的な研究 (Barrickら, 1991; Hough, 1992)

米国における妥当性研究の動向

【性格テストの妥当性】

- 米国における近年の見解
- 性格テストは職務遂行能力の予測変数として有力である
 - Conscientiousness (Achievement)
 - Extraversion (Potency)
- ただし、その妥当性係数の値は、職務特性、企業特性などによって異なる
 - ⇒状況個別性が高く、妥当性の一般化はできない

米国における妥当性研究の動向

【知的能力テストの妥当性】

- 「g因子」に関して、高い値の妥当性係数が、状況の違いを超えて安定的に得られる
- 性格検査に比べ、その妥当性が一般化されることが多い(ただし、通常は職種別)
- 職務の複雑さが妥当性係数に影響する
(Hunterら1984)

海外の先行研究

❖ 性格検査・能力検査と職務遂行能力評価との関連性の研究は
欧米諸国で数多く行われてきた

【性格テスト】－ BIG5のConscientiousness（誠実性） >

Barrick & Mount(1991)	.22(補正值)
Hough(1992)	.15(補正なし)
Salgado(1997)	.15(補正值)

【一般知的能力テスト】

Hunter & Hunter(1984)	.23~.56(複数職務Gr.のレンジ)
Schmitt et al.(1984)	.28(複数職務Gr.の平均)
Salgado et al.(2003)	.24~.63(複数職務Gr.のレンジ)

弊社の妥当性研究

❖ 日本での研究の蓄積は少ない。

	予測変数		研究デザイン		階層		その他	
	性格	能力	予測的	併存的	管理職	一般	職種別	業種別
A		○		○	○		○	
B	○	○		○	○			
C	○	○		○	○		○	○
D	○	○	○			○		
E	○		△	△		○		
F	○	○	○			○	○	○

A: 高橋・西田(1994)

B: 二村・今城・内藤(2000)

C: 今城・内藤・二村(2000)

D: 持主・今城・内藤・二村(2004)

E: 都澤・二村・今城・内藤(2005)

F: 飯塚・持主・内藤・二村(2005)

弊社の妥当性研究の紹介

- 管理者層対象の妥当性研究
 - 二村・今城・内藤(2000)
 - 今城・内藤・二村(2000) ー職務内容別のメタ分析ー
能力テスト・性格テスト、職種別(同一データを使用)
 - 一般層対象の妥当性研究
 - 飯塚・持主・内藤・二村(2005) 適性検査の予測的妥当性
能力テスト・性格テスト、職種別
- ☆研究数がさほど多くないので、誤差分散の説明率による一般化の判断は？
→一応、75%ルールでも見ておく
- ☆状況個別性の範囲にも興味
→「妥当性係数の水準」&「80%信頼区間」に着目

管理者層を対象とした妥当性一般化研究

<分析の対象>

- メタ分析の対象
→ 22社の企業における24の妥当性係数(r)
- 予測変数 → 管理者適性検査NMAT
(管理者選考場面、研修場面で実施)
- 基準変数 → 上司による人事評価、昇進スピード、
研修場面での評価

<職種ごとの個別妥当性研究数>

営業	一般管理	技術・研究
14件(N=884)	12件(N=504)	11件(N=943)

管理者層を対象とした妥当性一般化研究

＜妥当性係数の補正＞

1) 標本誤差 → $\bar{n}=184$

2) 基準変数の測定誤差 → $\rho_{yy}=0.6$

3) 予測変数の測定誤差 → $\rho_{xx}=0.81\sim 0.93$

4) 予測変数の範囲制限による誤差

→ $1/U=0.74\sim 1.71$

(制限されたSD / 制限されていないSD)

妥当性係数の水準

米国の先行研究と比較して、全体的に高水準

	尺度	全体	職種別		
			営業 N=884	一般管理 N=504	技術・研究 N=943
性格	内向—外向	0.18 (0.14)	0.20 (0.15)	0.18 (0.13)	0.10 (0.06)
	調整—統率	0.31 (0.22)	0.32 (0.22)	0.33 (0.22)	0.25 (0.17)
	心情—理性	0.18 (0.12)	0.11 (0.07)	0.22 (0.16)	0.13 (0.09)
	繊細—強靱	0.14 (0.10)	0.05 (0.04)	0.30 (0.21)	0.14 (0.10)
	維持—変革	0.26 (0.18)	0.31 (0.20)	0.26 (0.19)	0.20 (0.14)
	思索—行動	0.03 (0.02)	0.06 (0.05)	0.04 (0.02)	0.01 (0.00)
	慎重—大胆	0.13 (0.09)	0.17 (0.12)	0.14 (0.09)	0.08 (0.06)
	自律—承認	0.06 (0.04)	0.21 (0.12)	-0.05 (-0.04)	-0.03 (-0.01)
一般能力的	基礎能力総合	0.26 (0.18)	0.29 (0.18)	0.25 (0.18)	0.17 (0.12)
	概念的理解	0.20 (0.14)	0.22 (0.14)	0.25 (0.17)	0.11 (0.07)
	論理的思考	0.27 (0.19)	0.28 (0.18)	0.23 (0.16)	0.19 (0.13)

米国の先行研究と比較して、若干低め

誤差による分散の説明率

	尺度	全体	職種別
性格	内向－外向	60	24～53
	調整－統率	74	56～100
	心情－理性	93	47～81
	繊細－強靱	67	49～72
	維持－変革	45	38～100
	思索－行動	81	43～69
	慎重－大胆	100	48～100
	自律－承認	100	42～85
一般能力的	基礎能力総合	31	31～50
	概念的理解	31	35～68
	論理的思考	25	29～67

＜妥当性一般化の確認＞

【全体】

- ・ 75%ルールを超えたのは、性格4尺度
- ・ 75%を超えないまでも比較的説明率の高い性格尺度に対して、説明率の小さい能力検査

【職種別】

- ・ 性格検査は一貫した傾向なし
- ・ 能力検査はいずれの職種も『全体』より高い

* 数値はパーセンテージ

90%信頼限界下方値

	尺度	全体	職種別
性格	内向－外向	0.08	-0.20～0.18
	調整－統率	0.23	0.13～0.25
	心情－理性	0.14	-0.03～0.03
	繊細－強靱	0.05	-0.11～0.08
	維持－変革	0.11	-0.01～0.26
	思索－行動	-0.03	-0.15～-0.10
	慎重－大胆	0.13	-0.03～0.08
	自律－承認	0.06	-0.28～0.17
一般能力	基礎能力総合	-0.01	-0.15～0.10
	概念的理解	0.06	-0.04～0.08
	論理的思考	0.03	-0.10～0.10

<適用可能性の確認>

【全体】

・性格2尺度(「思索－行動」「自律－承認」と能力の1尺度(概念的理解)を除いて0を超える
→ほとんどの尺度でtransportabilityが確認された

【職種別】

・すべての職種で0を超えたのは「調整－統率」のみ

一般層を対象とした予測的妥当性研究

<分析の対象>

– メタ分析の対象

→ 32社の企業における32の妥当性係数(r)

❖ 予測変数 総合検査SPIおよびSPI2

収集期間 1982年～2004年

(採用選考・内定者・新入社員の時期に実施)

❖ 基準変数 個別研究ごとの職務遂行能力評価

収集期間 1998～2004年

(人事考課・昇進スピードなど)

❖ 検査実施から評価までの期間 平均6.7年

一般層を対象とした予測的妥当性研究

✦ 標本全体	全体
研究数(k)	32
人数(n)	7198

1研究あたり標本数 30～846

標本数平均 224

✦ 職種別	営業	スタッフ	開発・技術
研究数(k)	25	11	14
人数(n)	3945	1097	2156

一般層を対象とした予測的妥当性研究

<妥当性係数の補正>

1) 標本誤差 → $\bar{n}=224$

2) 基準変数の測定誤差 → $\rho_{yy}=0.6$

3) 予測変数の測定誤差 → $\rho_{xx}=0.80\sim 0.89$

4) 予測変数の範囲制限による誤差

→ $1/U=0.51\sim 1.32$

(制限されたSD / 制限されていないSD)

【全体】メタ分析結果

	性格32社 (N=7155), 能力30社 (N=6722)			
	補正前平均妥当性係数	補正後の妥当性係数 $\hat{\rho}$	誤差による分散説明率	80%確信区間
社会的内向性	-0.10	-0.16	56%	-0.25 ~ -0.06
内省性	-0.03	-0.04	81%	-0.09 ~ 0.01
身体活動性	0.11	0.16	44%	0.04 ~ 0.27
持続性	-0.02	-0.03	81%	-0.07 ~ 0.02
慎重性	-0.04	-0.05	101%	~
達成意欲	0.06	0.10	51%	-0.01 ~ 0.20
活動意欲	0.13	0.18	77%	0.12 ~ 0.23
敏感性	-0.03	-0.04	49%	-0.15 ~ 0.07
自責性	-0.06	-0.09	85%	-0.13 ~ -0.04
気分性	-0.03	-0.04	136%	~
独自性	-0.03	-0.05	52%	-0.15 ~ 0.05
自信性	0.05	0.08	96%	0.06 ~ 0.10
高揚性	0.06	0.10	65%	0.02 ~ 0.17
言語能力	0.00	0.00	44%	-0.12 ~ 0.12
非言語能力	0.06	0.10	21%	-0.11 ~ 0.30
一般知的能力総合	0.04	0.06	24%	-0.13 ~ 0.26

90%信頼
限界下方
値と同じ

【職種別】妥当性係数の水準

	営業(25)		スタッフ(11)		開発・技術(14)	
	補正前平均妥当性係数	補正後の妥当性係数 $\hat{\rho}$	補正前平均妥当性係数	補正後の妥当性係数 $\hat{\rho}$	補正前平均妥当性係数	補正後の妥当性係数 $\hat{\rho}$
社会的内向性	-0.09	-0.15	-0.09	-0.14	-0.12	-0.17
内省性	-0.04	-0.05	0.01	0.01	-0.05	-0.07
身体活動性	0.11	0.16	0.10	0.13	0.13	0.18
持続性	-0.02	-0.03	-0.03	-0.03	0.00	-0.01
慎重性	-0.05	-0.06	-0.04	-0.05	-0.02	-0.03
達成意欲	0.04	0.07	0.11	0.18	0.06	0.09
活動意欲	0.12	0.17	0.12	0.17	0.13	0.17
敏感性	-0.02	-0.02	0.02	0.03	-0.07	-0.10
自責性	-0.04	-0.07	-0.06	-0.09	-0.08	-0.12
気分性	-0.02	-0.03	0.01	0.02	-0.07	-0.10
独自性	-0.03	-0.05	-0.04	-0.05	-0.03	-0.04
自信性	0.02	0.04	0.04	0.06	0.11	0.16
高揚性	0.08	0.12	0.06	0.10	0.03	0.04
言語能力	0.00	-0.01	0.02	0.03	-0.03	-0.04
非言語能力	0.05	0.08	0.09	0.13	0.02	0.04
一般知的能力総合	0.04	0.06	0.06	0.10	-0.01	0.00

【職種別】誤差による分散説明率と80%確信区間

	営業(25)		スタッフ(11)		開発・技術(14)	
	誤差による分散説明率	80%確信区間	誤差による分散説明率	80%確信区間	誤差による分散説明率	80%確信区間
社会的内向性	38%	-0.32 ~ 0.01	98%	-0.16 ~ -0.11	88%	-0.21 ~ -0.13
内省性	69%	-0.13 ~ 0.03	141%	~	58%	-0.16 ~ 0.03
身体活動性	30%	-0.02 ~ 0.34	119%	~	45%	0.06 ~ 0.30
持続性	81%	-0.09 ~ 0.03	63%	-0.15 ~ 0.08	37%	-0.15 ~ 0.13
慎重性	78%	-0.12 ~ 0.00	219%	~	77%	-0.09 ~ 0.03
達成意欲	54%	-0.05 ~ 0.18	57%	0.04 ~ 0.32	92%	0.05 ~ 0.12
活動意欲	45%	0.04 ~ 0.30	92%	0.13 ~ 0.22	55%	0.08 ~ 0.27
敏感性	43%	-0.17 ~ 0.12	70%	-0.08 ~ 0.13	36%	-0.25 ~ 0.04
自責性	50%	-0.19 ~ 0.06	271%	~	47%	-0.24 ~ 0.00
気分性	58%	-0.14 ~ 0.08	146%	~	69%	-0.18 ~ -0.03
独自性	40%	-0.20 ~ 0.10	55%	-0.19 ~ 0.09	32%	-0.21 ~ 0.12
自信性	56%	-0.07 ~ 0.15	117%	~	66%	0.08 ~ 0.24
高揚性	35%	-0.05 ~ 0.29	95%	0.06 ~ 0.13	126%	~
言語能力	45%	-0.15 ~ 0.14	112%	~	21%	-0.27 ~ 0.19
非言語能力	26%	-0.13 ~ 0.29	59%	0.00 ~ 0.27	23%	-0.18 ~ 0.26
一般知的能力総合	29%	-0.14 ~ 0.26	84%	0.03 ~ 0.17	19%	-0.25 ~ 0.25

人事アセスメントにおける妥当性研究の課題

- 妥当性一般化は個別研究があって成り立つ
- 一般的傾向との比較(結果の利用)
 - 解釈の際の基準、よりどころとなる
- 利用者としての関わりの影響
- 人事テストの役割

改めて両研究をふりかえると…

- 対象の違い
 - 管理者層に共通して求められる特性
 - NMAT開発の背景(MATからの変遷)
- テストの位置づけ
 - 昇格昇進選考の支援機能
 - 採用選考時の人物理解機能
- テストの果たす役割と妥当性
 - 選考は『判断』
 - 有用な情報として利用すること=妥当性

2006/9/9