検査主体の求める人物像に合わせた尺度の個別設定を行える測定方法の試み

佐々木 研一 EI リサーチ株式会社 豊田 秀樹 早稲田大学

New assessment scale based on the type of person desired by an employer

Kenichi Sasaki (EI Research, Inc.) and Hideki Toyoda (Waseda University)

In many cases, aptitude tests used in the hiring process fail to connect the measurement scale with the emotional type of the person desired by an employer. This experimental study introduced a new measuring method, in which the measurement scale could be adjusted according to the type of person an employer is seeking. Then the effectiveness of this method was verified by comparing the results of an aptitude test utilizing the method and the results of the typical hiring process carried out by the new method in hiring.

Key words: aptitude test, emotional intelligence, analytic hierarchy process, paired comparison.

The Japanese Journal of Psychology
2011, Vol. 82, No. 4, pp. 379-384

日本企業の社員採用時に用いられる適性検査には様々な種類があるが、一つの類型として、最大値測定と典型値測定(Thorndike, 1997)の区別を挙げることができる。最大値測定は、知能・学力検査(以下、基礎能力検査とする)に代表される測定であり、受検者の特性の最も大きな値の測定を目的とする。一方、典型値測定は、性格・興味検査(以下、適合度検査とする)に代表される測定であり、受検者の特性の典型的な値を測定することを目的とする。ただし、典型値測定を実施する際、二つの問題がある。一つは、社会的望ましさへの配慮から反応歪曲が生じる傾向があること(岩脇, 1973)、もう一つは、企業の社風や職種によって重視される特性の内容が異なることである。このため、採用時には、最大値測定が効果的であり、基礎能力検査が主体となる傾向があった。

しかし、有能な受検者を入社させるという目的に照らす時、基礎能力検査では測れない性格・興味等の受検者特性を選抜の際に考慮し、企業の社風や職種に適合しているかを測定する適合度検査の実施も重要となる。そのため、採用時に典型値測定を行う場合に生じる反応歪曲に抗する研究(例えば、豊田・川端・松下(2005)により、社会的望ましさによる反応歪曲に耐性を持った検査の開発)が行われている。また、特定

の職種において重視される特性の研究(飯塚・持主・内藤・二村, 2005)が試みられている。これらの先行研究により、受検者の特性の一部を高い精度で予測可能な適合度検査の開発が行われており、現在では多くの企業で、採用時に最大値測定及び典型値測定が併用されている。しかし、実際に適合度検査を活用する場面において、そもそも検査を実施する主体(以下、検査主体とする)が求める人物像と適合度検査の測定内容にずれが生じることも多く、検査結果の内容妥当性や基準関連妥当性の低下が懸念される。理由としては、検査主体において評価の基準を決める人(以下、基準設定者とする)が、検査の専門家ではなく、測定内容を完全には理解できないこと、また、一方で適合度検査の提供者が、仮に心理検査の専門家であったとしても、検査主体とは別の人間であり、検査主体の求める人物像を完全には理解できていないことが考えられる。そこで、心理学の特定の理論に基づく複数の尺度に対し、当該検査主体が求める人物像に合わせて、尺度に重みづけ(例えば、複数の尺度の重みの合計が1になる配点に)してテスト得点を構成することができれば、検査主体毎に適切な適合度検査を構成することが可能になる。そのような方法の一つとして、AHP (Analytic Hierarchy Process) が考えられる。AHP は、Saaty (1980) によって提唱された意思決定法である。手順としては、手順1、問題の要素を、意思決定したい目的、観点、代替案の関係でとらえて、階層構造を作る。手順2は、(a)目的からみた観点の重

要度と、(b)各観点からみた代替案の重要度を求める。手順 3 は、総合評価を(a)と(b)の加重和により求める。例えば、Benaïm, Perennou, Pelissier, & Daures (2010) は、AHP を用いて検査主体が項目単位に重要度を設定し、テスト得点を構成した際の有効性を示している。ただし、採用時の適合度検査のように多数の項目を持つ場合には、尺度単位に重要度を設定する必要があり、この方法は適用できない。

本研究の目的は、まず採用時に活用されている適合度検査において、検査主体が評価基準を個別設定するために、尺度単位に重要度を設定し、テスト得点を構成できる新しい測定法（以下、評価基準個別設定法とする）を提案し、次に評価基準個別設定法の有効性を検証することである。

AHP の考え方に従えば、目的は検査主体が求める受検者を採用することであり、観点は、当該企業が受検者を採用する際の評価基準となる。また、当該企業が、就職を希望する受検者のうち、どの受検者を採用するかを選ぶことから、代替案を受検者として行うことができる。総合評価は、観点の重要度と受検者の適合度検査の結果との加重和（以下、適合度得点とする）である。AHP の手順 2 (b)各観点からみた代替案の重要度は、受検者が適合度検査を受検することで得られるため、評価基準個別設定法では、AHP の手順 2 (a)目的からみた観点の重要度を定めるプロセスの部分のみ利用し、代替案である受検者の適合度得点を求める。

評価基準個別設定法は、どのようなテストバッテリーにも適用可能であるが、本研究では反応歪曲に耐性を持った典型値測定による EI (Emotional Intelligence) を測定する検査（豊田他, 2005）を用いて研究を行った。したがって、目的からみた観点の要素は、EI 検査の 12 尺度を用いることになる。企業における優秀な人材とは、知的な基礎能力が高いばかりでなく、自制心、根気・やる気、対人スキル、共感性・協調性等を有する EI が高い人物でもある（Goleman, 1998 梅津訳 2000）ため、企業が採用時に受検者の EI の特性を適合度検査で測定することには意味があると考えられる。Goleman (1998), Salovey & Mayer (1990) は、EI 検査で測定する基本的な領域として、(a)自分自身の情動を知る、(b)自分の感情を制御する、(c)他者の情動を認識する、(d)他者との人間関係をうまく処理するという四つを挙げている。豊田らによる EI 検査は、四つの領域内で、それぞれ三つの下位尺度、合わせて 12 尺度から構成され、各尺度はそれぞれ 6 項目で構成されている（EI 検査全体で 72 項目）。豊田他 (2005) は、反応歪曲への耐性を比較するため、同様の項目によるリッカート法における EI 検査と一対比較法による EI 検査を、採用時及び非採用時で比較した。その結果、一対比較法は、リッカート法に比べ、次の 3 つの点で反応歪曲に対する耐性

があることが示されている。一つ目は、採用時においても、無意味に平均値が高くない。二つ目は、採用時においても、社会的望ましい回答に集中しないため、分散が小さくならない。三つ目は、非採用時の得点と採用時の得点における相関が高い。一方で、リッカート法は、収束的妥当性が低く、妥当性の観点から一対比較法はリッカート法に対し、優位性を有していることが明らかにされている。

方 法

AHP を活用した新しい測定法

階層構造の作成 まず、AHP の手順 1 に従い、階層図の作成を試みた。作成した階層図を Figure 1 に示す。

目的からみた観点の重要度の決定（基準設定検査の実施）多くの企業は、採用を効果的に実施するために、採用の早い段階で検査主体と適合しない受検者を選考から省き、採用の過程が進んでいく中で、最も検査主体と適合する受検者を選抜するという方法をとっている。適合度検査は、採用の早い段階で検査主体と適合しない受検者を省くために活用されることが多い。したがって、基準設定者は、適合しない受検者を省くという目的に照らして、観点の重要度を決定するために、一対比較法検査（以下、基準設定検査とする）を実施する。基準設定検査の測定方法は、階層図の直下の項目である 12 の観点のうち二つの観点を取り上げて対とし、直上の目的からみて、その対のどちらをどの程度重視しているのかについて、当てはまり具合を 4 段階の評定尺度上に 1 箇所だけチェックを付け、比較・評価する（基準設定検査全体は、66 項目で構成されている）。ただし、基準設定者が基準設定検査を受検する際、観点の測定内容を正確には把握していないため、観点の名称（モニタリング等）のみをみて比較・評価することは困難であり、測定内容を誤認識した場合、正しい結果が得られない恐れがある。そのため、基準設定者が観点を容易に理解できるように、各観点の得点が低い場合に受検者がとる可能性の

Figure 1. AHP における階層図

Table 1
基準設定検査の設問例

尺度	項目
モニタリング	自分がやりたいことを考えていないか、もしくは、気づいていないため、仕事に意欲をもって取り組めない
適切な自己評価	自分を過信して、ミスを繰り返したり、周囲の助言を受け入れようとしたりしない、逆に、自分を過小評価して、チャレンジしようとししない
肯定的自己像	自分の能力・知識・経験に自信をもてないため、何事にも不安に感じ、自分を卑下するような行動を取る

高い行動をこれまでに検査を実施した経験や採用後の受検者の行動を担当者にヒアリングした結果から設問文を作成した（項目の具体例を Table 1 に示す）。基準設定検査から観点の重要度を決定する計算方法として固有値法と幾何平均法とがあるが、本研究においては、あまり結果が変わらなかったため、より簡便な幾何平均法を用いた。評価基準個別設定法では、検査主体の観点をより明確にするために、基準設定検査の12の観点のうち、基準設定者が重視する上位六つの観点をを用いることとした。

適合度得点の算出 適合度得点は、取り出した上位六つの観点の重要度と受検者の EI 検査の同じ六つの観点の得点を掛けた積の総和（加重和）により求める。適合度得点が低い場合、当該企業と受検者との適合度が低いと解釈する。

適合度得点の標準化 検査主体が基準設定検査を実施する度に、適合度得点を標準化する手続きを行った。標準化は、基準設定者が重視する上位六つの観点と、あらかじめ準備してある EI 検査の結果（2003 年 7 月から 11 月までに得られた大学生、会社員、教員、主婦等を含む 673 名）との加重和を算出した結果をもとに行った。

検定方法

標本数の決定（事前の分析） 評価基準個別設定法の有効性を検討するにあたり、各標本において、合格者と不合格者の適合度得点の平均値から導かれる検出したい実質的な差を十分な検定力で検定するために、必要な標本数を実験前に確認した。実験前のため、2 群の標本数を同じとして n と表現すると式（1）で表現できる（豊田，2009，p. 36）。 α は有意水準を表す。

$$n = f(\text{母効果量}, \text{検定力}, \alpha) \quad (1)$$

本研究では、母効果量 = .8，検定力 = .8， α = .05 の場合の標本数 n を求める。結果、必要な標本数は、合格者、不合格者がそれぞれ 26 名（ $n = 25.5$ ）となる。

平均値差の検定力分析（事後の分析） 実験実施後の標本検定力は式（2）を用いて確認した（豊田，

2009，p. 37）。

$$\text{標本検定力} = f(n_{\text{実験}}, n_{\text{対象}}, \text{標本効果量}, \alpha) \quad (2)$$

受検者に対する EI 検査の実施方法 受検者は EI 検査を Web にて受検した。

結 果

以下、評価基準個別設定法を適用した三つの事例結果を示す。全事例とも、実際の採用時に評価基準個別設定法を実験的に導入したものであり、測定の結果は実際の可否にはまったく影響しておらず、採用の可否は面接の結果において決定された。また、全事例とも、基準設定者は、募集人員の採用における責任者であり、直接面接を実施した。したがって、実際の合格者と不合格者の平均値差の評価は評価基準個別設定法の妥当性を示すものとなる。平均値差の評価は、標本効果量を用いる。標本効果量は、Cohen（1962）によって 0.2 が小さな効果、0.5 が中くらいの効果、0.8 が大きな効果の目安として提唱されている。評価基準個別設定法が面接の可否評価の一部を再現することに成功しているならば、コストのかかる面接の一部に代替することが可能であり、実践的・実務的に有効であるといえる。

事例 1

実験協力者 A 社の 2010 年度新卒採用に応募後、1 次面接に合格した大学生 78 名。

調査期間 2009 年 1 月から 4 月までの 4 ヶ月間。

基準設定者と重視する観点 人事部長が基準設定検査を受検し、結果は Figure 2 となった。重視する上位六つの観点は、重要度の高い順に、克己抑制、サービス精神、チームワーク、適切な自己評価、共感性、達成動機。

Figure 2. A 社の基準設定者が決定した基準値

Table 2
A 社の標本検定力

α	5%	1%	0.1%
$1-\beta$	.992	.959	.832

Table 3
A 社の“明日への分析” (必要標本数)

$\alpha \setminus 1-\beta$	.7	.8	.9	.95
5 %	14	17	22	27
1 %	21	25	31	37
.1 %	32	37	44	51

までは個人の能力を重視してきたが、企業として成熟してきたこともあり、今後はチームで業務を行うことを重視した採用を行いたいとする意向があった。今回の基準設定の結果は A 社のイメージに合ったプロトコルと考えられよう。

結果 合格者は 45 名で適合度得点の平均 48.429, 標準偏差 8.244, 不合格者は 33 名で適合度得点の平均 39.015, 標準偏差 10.028 となった (Figure 3)。 t 検定の結果は, $t=4.410$, $p<.01$ であり, 標本効果量は 1.011, 標本効果量の 95% 信頼区間の下限は 0.534, 上限は 1.488 である。このことから, 合格者と不合格者の適合度得点の平均値差には有意差があることが確認でき, その差は, 標準効果量から大きいと解釈される。また, 実験実施後の標本検定力の結果から (Table 2), 1% 水準でも有意差を得る確率が 95% を超えているため, 適切な検定力で検定したといえる。実験結果を踏まえて引き続き実験する際に必要な標本数の目安をつける分析を行う。この分析を豊田 (2009, pp. 36-37) は, “明日への分析” と称してい

Figure 3. A 社の合格者と不合格者における新しい測定法による適合度得点の平均と±標準偏差の図

る。結果を Table 3 に示す。検定力が 80% で α が 5% なら, 17 名以上で有意差が出ると予測される。

事例 2

実験協力者 B 社の 2011 年度新卒採用に応募後, 会社説明会に参加した大学生 575 名。

調査期間 2010 年 3 月の 1 ヶ月間。

基準設定者と重視する観点 人事部長が基準設定検査を受検し, 結果は Figure 4 となった。重視する上位六つの観点は, 重要度の高い順に, サービス精神, アサーション, チームワーク, 克己抑制, リーダーシップ, 適応性である。B 社では, 新人であってもアルバイトの管理をするため, リーダーシップが発揮できることを重視した採用を行いたいとする意向があった。今回の結果は B 社のイメージに合ったプロトコルと考えられよう。

Figure 4. B 社の基準設定者が決定した基準値

Figure 5. B 社の合格者と不合格者における新しい測定法による適合度得点の平均と±標準偏差の図

Table 4
B社の標本検定力

α	5%	1%	0.1%
$1-\beta$	.940	.826	.586

 Table 5
B社の“明日への分析”（必要標本数）

$\alpha \setminus 1-\beta$	.7	.8	.9	.95
5%	42	53	70	86
1%	65	79	100	119
.1%	98	115	140	163

結果 合格者は44名で適合度得点の平均48.145、標準偏差8.598、不合格者は531名で適合度得点の平均43.403、標準偏差8.308となった（Figure 5）。 t 検定の結果は、 $t=3.525$, $p<.01$ であり、標本効果量は0.553、標本効果量の95%信頼区間の下限は0.244、上限は0.862である。このことから、合格者と不合格者の適合度得点の平均値差には、有意差があることが確認でき、その差は、標本効果量から中程度と解釈される。実験実施後の標本検定力の結果から（Table 4）、1%水準でも有意差を得る確率は80%を超えているため、適切な検定力で検定したといえる。“明日への分析”の結果をTable 5に示す。検定力が80%で α が5%なら、54名以上で有意差が出ると予測される。

事例3

実験協力者 C社の中途採用の1次面接に参加した社会人22名（事例3は事前の分析による目標の標本に足りていない）。

調査期間 2009年2月から5月までの4ヵ月間。

基準設定者と重視する観点 営業部長が基準設定検

Figure 6. C社の基準設定者が決定した基準値

 Table 6
C社の標本検定力

α	5%	1%	0.1%
$1-\beta$	.928	.764	.432

 Table 7
C社の“明日への分析”（必要標本数）

$\alpha \setminus 1-\beta$	.7	.8	.9	.95
5%	4	5	6	7
1%	6	7	8	9
.1%	9	10	12	13

査を受検し、結果はFigure 6となった。重視する上位六つの観点は、重要度の高い順に、サービス精神、センシブル、チームワーク、共感性、適切な自己評価、克己抑制である。C社では、社内業務が主である庶務職を募集しており、他者の利益のために自発的かつ、意図的に行動できることを重視した採用を行いたいとする意向があった。今回の結果はC社のイメージに合ったプロトコルと考えられよう。

結果 合格者は3名で適合度得点の平均55.233、標準偏差4.626、不合格者は19名で適合度得点の平均43.395、標準偏差8.378となった（Figure 7）。 t 検定の結果は、 $t=3.598$, $p<.05$ であり、標本効果量は2.235、標本効果量の95%信頼区間の下限は0.834、上限は3.636である。このことから、合格者と不合格者の適合度得点の平均値差には、有意差があることが確認できた。標本数はA社・B社と比較して少なく、必要な標本数である26名にも足りていないが、標本効果量及び、標本効果量の信頼区間の下限値からも、この平均値の差は大きいと解釈できる。実験実施後の

Figure 7. C社の合格者と不合格者における新しい測定法による適合度得点の平均と±標準偏差の図

Table 8
3 基準×3 社の効果量

	受検者データ		
	A 社	B 社	C 社
A 社基準	1.011	-.009	.692
B 社基準	.684	.553	-1.385
C 社基準	.712	.267	2.235

標本検定力の結果から (Table 6), 5%水準であれば有意差を得る確率は 90%を超えているため, 適切な検定力で検定したといえる。“明日への分析”の結果を Table 7 に示す。検定力が 80%で α が 5%なら 5 名以上で有意差が見込める。

考 察

本研究で提案した新しい測定法を企業の採用時に試み, 有効性の検証を行った。三社の企業において, 実際の採用時に実験的に導入, 検証した結果, 合格者と不合格者の適合度得点の平均値差において, 実質的に意味のある効果量が観察され, 統計的にも有意な差があることが示された。このことから, 評価基準個別設定法の基準連関妥当性が示された。また, 基準設定者との面談において, 選択された観点 (基準設定検査の結果) の重要度を提示し, 解説したところ, A 社, B 社, C 社が重視しているイメージに合ったプロトコルであることが示されたことから評価基準個別設定法の内容的妥当性が示された。

各社で設定した基準を, 自社ばかりでなく, 各社間に適用し, 合格者と不合格者の適合度得点の平均値差の効果量を示したものが Table 8 である。各社にとってどの基準が最良であるかを確認すると, 主対角の効果量が大きいこと, 非主対角の効果量が小さいことが観察される。例えば A 社の観点の重要度 (A 社基準) で B 社の受検者を評価すると効果量は無くなるし, B 社の観点の重要度 (B 社基準) で C 社の受検者を評価すると効果量はマイナスにすらなる。効果量の定義式は, t の絶対値を取ることもあるが, ここでは不合格者の平均値が高くなったことを示すために, 絶対値ではなく, t 値そのものを用いて効果量を算出した。自社の観点の重要度では効果量が最大になるが, 他社の観点の重要度を用いると効果量が小さくなってしまふという事実は, 評価基準個別設定法の“各社の欲しい人物像”に関する構成概念妥当性を示しているといえよう。このことから, 評価基準個別設定法が面接の合否評価の一部を再現することに成功していることが

示され, 実践的・実務的に有効な測定法であると考えられる。

引用文献

- Benaïm, C., Perennou, D. A., Pelissier, J. Y., & Daures, J. P. (2010). Using an analytical hierarchy process (AHP) for weighting items of a measurement scale: A pilot study. *Revue d'Épidémiologie et de Santé Publique*, **58**, 59-63.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology*, **65**, 145-153.
- Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam Books.
- (ゴールマン, D. 梅津 祐良 (訳) (2000). ビジネス EQ ——感情コンピテンスを仕事に生かす——東洋経済新報社)
- 飯塚 彩・持主 弓子・内藤 淳・二村 英幸 (2005). 適性検査の予測妥当性——職種別および製造・非製造業別の分析——産業・組織心理学会第 21 回大会発表論文集, 159-162.
- (Iizuka, A., Mochinushi, Y., Naito, J., & Nimura, H. (2005). Predictive validity of personality and general mental ability tests: Analysis by occupational groups and manufacturing-nonmanufacturing business sectors. *Proceedings of the 21th Annual Conference of The Japanese Association of Industrial / Organizational Psychology*, 159-162.)
- 岩脇 三良 (1973). 心理検査における反応の心理 日本文化科学社
- (Iwawaki, S.)
- Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.
- Salovey, P., & Mayer, J. D. (1990) Emotional intelligence. *Imagination, Cognition, and Personality*, **9**, 185-211.
- Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education*. 6th ed. Englewood Cliffs, NJ: Prentice-Hall.
- 豊田 秀樹 (2009). 検定力分析入門 東京図書 (Toyoda, H.)
- 豊田 秀樹・川端 一光・松下 信武 (2005). 採用場面における EQ 検査の改善 教育心理学研究, **53**, 456-466.
- (Toyoda, H., Kawahashi, I., & Matsushita, N. (2005). Improvement of Emotional Intelligence Quotient (EQ) Test used for employment. *Japanese Journal of Educational Psychology*, **53**, 456-466.)

—— 2010. 7. 8 受稿, 2011. 3. 5 受理 ——