

採用場面における EQ 検査の改善

豊田 秀樹* 川端 一光** 松下 信武***

構造方程式モデリングを用いた一対比較型 EQ 検査の反応歪曲に対する耐性の確認, ならびに標準化が試みられる。実験 1 では一対比較型 EQ 検査の反応歪曲への耐性を検討するために, 140 名の被験者に対して一対比較型 EQ 検査と, 同一項目で構成されたリッカート測定法による EQ 検査が, 選抜的評価場面, 非選抜的評価場面の 2 場面を想定して実施された。その結果, 一対比較型 EQ 検査は測定場面にかかわらず, 標準偏差, 効果量, 収束的妥当性の観点からリッカート測定法よりも反応歪曲に耐性を持っていることが確認された。実験 2 では一対比較型 EQ 検査のスコアリングを安定させる為, 新たに用意された 383 名の被験者のデータを実験 1 の被験者のデータに加えて計算を行った。また代替検査の開発の為にテストを 20 パタンに折半し, 10 パタンそれぞれで尺度間相関を求めた。その結果併存的妥当性の観点から支持される代替テストが得られた。

キーワード: 一対比較型 EQ 検査, 構造方程式モデリング

問題・目的

心理測定には様々な種類があり, 1 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, 学力検査, 技能検査に代表される測定であり, 被験者特性の最も大きな値を測定することを目的とする。それに対して典型値測定は, 性格検査・興味検査に代表される測定であり, 被験者の特性の典型的な値を測定することを目的とする。

心理測定の結果の評価は, 選抜的評価と非選抜的評価に分けて考えることができる。選抜的評価は, 入学試験や入社試験に代表される評価であり, 受験者が受験結果によって何らかの(多くの場合に合格か不合格かの)処遇を受ける。非選抜的評価は, 教育場面における単元末ドリルや職業興味検査に代表される評価であり, 形成的判断や自己診断や進路指導の資料的情報として利用されることの多い評価である。正確に特性値が測定されたほうが受験者にとって望ましいので, 非選抜的評価場面において反応歪曲は深刻ではない。

選抜的評価場面では, 社会的望ましき (social desirability) への配慮から反応歪曲が生じる傾向がある (岩脇三良, 1973)。社会的望ましきが反応に及ぼす影響の除去

は, EPPS (Edwards, 1959) によって試みられている。また肥田野・岩原・岩脇・杉村・福原 (1970) は EPPS の日本語版を作成している。EPPS では社会的望ましきがほぼ同じであるような項目の対へ, 強制選択法と呼ばれる手続きで反応させることによって, 反応歪曲の少ない回答を得ることに成功している。

この反応歪曲は本来の自分の姿に社会的望ましき加わることにより生じると考えられる。反応歪曲は典型値測定の検査に関して顕著であり, たとえば入社試験に

1. 朝起きて理由もなく学校に行きたくない。
2. 隣のひとは電波を出し私を見張っている。

という性格検査の項目が出題され, “はい”, “どちらでもない”, “いいえ” というリッカート法で回答する場合には, 少なからず実態よりも “いいえ” の方向への反応歪曲が生じることが予想される。被験者は, 自分がどのような特性を有しているかというよりも, むしろどのような特性が望ましいかを想像して回答する為である。学力入学試験に対する過度な受験テクニックは, 意識的な反応歪曲の一種と考えることもでき, 最大値測定にも反応歪曲は存在するが, 選抜的評価場面では, 明らかに典型値測定のほうが意識的な反応歪曲が容易である (続・村上, 1975)。

この為選抜的評価場面では, 最大値測定が効果的であり, これまで入社試験では能力検査や学力検査が主体となる傾向があった。しかし, 有能な受験者を入社させるという目的に照らすとき, 能力テストや学力テストでは測れない特性を選抜の際に考慮することは重

* 早稲田大学

** 早稲田大学大学院文学研究科

〒162-8644 東京都新宿区戸山 1-24-1 早稲田大学文学部

*** 株式会社 EQ ジャパン

〒107-0052 東京都港区赤坂 2-22-15 赤坂加藤ビル 1 F

要である。企業における優秀な人材とは、IQ ばかりでなく、自制心、根気・やる気、対人スキル、共感性・協調性等が高い水準にある EQ (emotional intelligence quotient) も高い人物である (ゴールマン, 2000)。

典型値測定は、リッカート測定法で構成された質問紙によって行われることが多い。たとえば YG 性格検査, MPI モーズレイ性格検査 (Maudsley Personality Inventory) 等の代表的な性格測定はリッカート法によってスコアリングされる。テスト得点が項目得点の和で表現されているために、古典的テスト理論から導かれる (α 係数等の) 公式を利用しやすいこと、非選抜的評価場面では反応歪曲の悪影響が深刻でないことから、リッカート法による典型値検査は多数提案されている。尚、EQ は性格特性ではないが、最大値測定と典型値測定の観点からは、典型値測定に頼らざるを得なく、本研究ではその点において EQ と性格特性とを同様のものとして扱っている。

入社試験において性格・EQ に関する特性は、通常、面接試験で確認する。しかし面接試験は手間がかかるので、中間選抜されたあとの少人数の受験者にしか実施できないことも多い。妥当に選抜に反映できるならば、1 次試験から性格・EQ 検査を実施することが望ましい。しかし上述のように、リッカート法による性格・EQ の測定は、反応歪曲にさらされる危険が高い。

反応歪曲を少しでも回避する為に、入社試験では性格・EQ に関する特性をイプサティブスコアを用いて測定することが多い。イプサティブスコアとは、被験者ごとに尺度を考え、変数を尺度上に位置付けるような場合に用いられる得点であり、そのようなデータを実測値として持つ変数をイプサティブ変数という (豊田, 2000)。イプサティブ変数はその和が一定に調整され、当該被験者にとっての変数の重要度を表現する。

たとえば 5 件法の一対比較法ならば“どちらでもない”への反応は両対象に 3 点を、“A より B がやや当てはまる”への反応に、A に 4 点、B に 2 点を与えるならば総和は一定でありイプサティブ変数となる。

イプサティブスコアを算出するための質問項目形式としては、2 者択一形式の強制選択法、7 件 (5 件) の評定値による一対比較法、複数制限選択法などが挙げられる。イプサティブデータを用いた測定は、リッカート法による測定と比べて反応歪曲に耐性があるという利点を持つ反面、数理的扱いが難しいという問題点を有する。

リッカート法で測定されたデータは、集団に対して同一の尺度を考え、その尺度上に被験者を位置付ける

ことを目的として収集される。この場合被験者を通じて変数の相関係数を計算し、変数間の関連を考察することができる。しかしながらイプサティブデータを用いた典型値測定では、被験者ごとに尺度を考え、変数を尺度上に位置付けているため、単純に被験者を通じて変数間の相関関係を考察することは困難である。また標本分散共分散行列・相関行列の行列式が 0 になり、尤度を用いた統計モデルによる分析が困難になるという性質も挙げられている (豊田, 2000)。

本論文では典型値測定尺度として、新たに EQ 検査を作成する。さらに新尺度に関してリッカート法による測定と、構造方程式モデリングによって表現された一対比較法による測定とを比較することにより、一対比較型 EQ 検査の有効性を確認する。また一対比較型 EQ 検査は構造方程式モデリングを用いることによって、イプサティブスコアではないテスト得点を算出するため、上述したイプサティブデータに関連する数理的難点を克服する。

本論文は、上述した採用場面における典型値測定に関する難点の改善を意図したものである。新尺度の考案と、スコアリングのための数理モデルの提示を同時に行うことによって、この問題に関して建設的な議論を提起する。

方 法

反応歪曲が想定されるような選抜・採用・考課などの診断評価場面で一対比較法によって典型値測定を行う一般的方法を提案し、一対比較型 EQ 診断検査 (diagnostic EQ test using paired comparison) を作成する。

測定する領域・尺度 Salovey & Mayer (1990), Mayer & Salovey (1993), ゴールマン (2000) は、EQ の基本的な領域として (1) 自分自身の情動を知る, (2) 自分の感情を制御する, (3) 自分を動機付ける, (4) 他人の感情を認識する, (5) 他者との人間関係をうまく処理する, という 5 つを挙げている。(1), (2), (3) は自分に関する領域, (4), (5) は他者に関する領域, (1), (4) は知ることに関する領域, (2), (3), (5) は働きかけに関する領域である。そこで本研究では、自己か他者か、知ることか、働きかけに注目し、(2) と (3) を 1 つにまとめて、

- a. “自己認識”：自己感情の識別, 上述 (1)
- b. “自己制御”：自己感情の利用, 上述 (2), (3)
- c. “対人理解”：対人感情の認知, 上述 (4)
- d. “対人技能”：対人感情の調整, 上述 (5)

の 4 つとする。

次にこの 4 つの領域内で、それぞれ 3 つの下位尺度

x , y , z を構成する。

“自己認識”は、自分の気持ちや感情を読み取り、自覚する能力とし、下位尺度 ax “モニタリング”, ay “適切な自己評価”, az “肯定的自己像”を設定する。

“自己制御”は、自分の気持ちや感情を適切に統制し、活用する能力とし、下位尺度 bx “克己抑制”, by “達成動機・統制感”, bz “安定感・柔軟性”を設定する。

“対人認知”は、相手の気持ちや感情を読み取り、共感する能力とし、下位尺度 cx “共感性”, cy “センシブル”, cz “サービス精神”を設定する。

“対人技能”は、相手の気持ちや感情を適切に活用する能力とし、下位尺度 dx “リーダーシップ”, dy “アサーション”, dz “チームワーク”を設定する(付録に下位尺度の定義を掲げた)。

12の尺度で、それぞれ6項目(テスト全体で72項目)を作成する。それぞれの尺度に関して、項目例を1つずつTABLE 1に示した。

一対比較測定 これらの項目をたとえばTABLE 2のように4領域間で12回比較する。これを1セットと呼ぶ。1セットの中では、たとえばTABLE 1の項目が、それぞれ2回ずつ出現する。テスト全体では6セット作成され、セット間にまたがって登場する項目はない。これは実験2において、項目の重複のない代替検査を作成するためである。また同一尺度での比較も行わない。(1)式では、観測変数の期待値を得るには、異なる次元の α の差得点の期待値が計算されなければならない。同次元を測定する信頼性の高いテスト間ではその差得点の信頼性が悪くなることが知られているから、(1)式における同次元の項目の比較はテスト全体の信頼性という観点からは好ましくないことになる。

具体的な評定方法は、2つの項目、上 $[mik]$ と下 $[njl]$ を比較して、どちらが被験者に当てはまっている

TABLE 2 項目の組み合わせ

尺度	a	b	c	d
a 自己認識	*	$az-by$	$ay-cy$	$ax-dy$
b 自己制御	$bz-ay$	*	$by-cx$	$bx-dx$
c 対人認知	$cz-ax$	$cy-bx$	*	$cx-dz$
d 対人技能	$dz-az$	$dy-bz$	$dx-cz$	*

るかを考えさせ、当てはまりぐあいの差を5段階評定尺度上に1個所だけ○印を付けることで表現させる。

m と n は領域を示す添え字であり、 a , b , c , d の値をとる($m \neq n$)。 i と j は下位尺度を示す添え字であり、 x , y , z の値をとる。下位尺度は領域の入れ子であるから、 i と j は、それぞれ m と n にネストし、正確に書くなら $i(m)$ と $j(n)$ である。

k と l は、下位尺度を構成する項目番号であり、1から当該下位尺度の項目数まで動く。項目は、下位尺度と領域の入れ子であるから、それぞれ $i(m)$ と $j(n)$ にネストし、正確に書くなら $k(im)$ と $l(jn)$ である。ただし、入れ子の関係は変化しないので、本論文では、以後省略して表記する。

たとえば、 az “肯定的自己像” dz “チームワーク”の1番目の項目を比較する場合には、

上 $[az1]$ 自分には人並み以上の知力がある

下 $[dz1]$ 仲間とはいっしょに楽しい時間を過ごすとなり、“上が下よりとても当てはまる”, “上が下よりやや当てはまる”, “どちらともいえない”, “下が上よりやや当てはまる”, “下が上よりとても当てはまる”に、それぞれ5点から1点を与える。以上のことから、観測変数は x_{miknjl} と表現される。

測定方程式 観測変数は以下の測定方程式によって

$$x_{miknjl} = (\alpha_{mik} - \alpha_{njl}) + (f_{mi} - f_{nj}) + d_{minj} + e_{miknjl} \quad (1)$$

のように分解する。右辺は、 α_{mik} と α_{njl} が母数¹(定数項)

TABLE 1 項目の具体例

尺度	項目
ax	あせり・恐怖・喜び・悲しみなど、いつも自分の今の感情に名前をつけられる
ay	自分の得意分野で勝負するように心がけている
az	自分は成功するだけの能力がある
bx	あせりや恐怖に打ち勝って冷静に行動できる
by	自らのスキルを磨き、常に自分を向上させる
bz	失敗の恐れより、成功の喜びを期待して信じるほうだ
cx	人から悩みを相談されることが多い
cy	すこし話をすると相手の好みや性格が把握できる
cz	ハイキングや飲み会などの集まりを企画して仲間を喜ばせることが好き
dx	人を引っ張っていく力がある
dy	顔をつぶさずに反対者を説き伏せる自信がある
dz	仲間を出し抜いて自分の手柄にしようとは思わない

であり、残りが確率変数（外生変数）である。右辺の確率変数の期待値は全て0であり、分散は自由母数として全て推定する。共分散は f_{mi} と f_{nj} の間に仮定される以外は全て0に固定する。分析プログラムには、SASの平均共分散構造分析のプロシジャCALISを利用した。

右辺の各項の解釈的意味を説明する。まず α_{mik} は（ α_{njl} は）、個人の特性とは独立の項目 mik の重要度（項目 njl の重要度）であり、ここに社会的望ましさが含まれる。

$$E[x_{miknjl}] = \alpha_{mik} - \alpha_{njl} \quad (2)$$

であるから、 α_{njl} が一定で、 α_{mik} が大きくなれば、当該観測変数の期待値も大きくなる。またここでは社会的望ましさを、EPPSにおける定義と同様に、集団における値として定義している²。よって(1)式における α は、各被験者ではなく被験者集団の代表値として算出される。

f_{mi} は（ f_{nj} は）、尺度 m_i における（尺度 n_j における）個人の尺度値である。平均共分散構造モデルでは、因子パターンが1か-1に固定された構成概念として表現される。テスト得点は、構成概念のスコアを求める重み行列を、観測変数にかけることによって計算する。また因子パターンが1に固定されていることから、パス係数の確認の意味がないので本研究ではパス図を割愛した。

構成概念スコアを求めるための重み行列は、選抜場面と非選抜場面のデータを併合して求める。この手続きにより選抜場面と非選抜場面を合わせた被験者の構成概念スコアの平均が0になり、反応歪曲があればリッカート測定法と同様に選抜場面の平均値が高くな

ることが予想される。

d_{minj} は尺度 m_i と尺度 n_j の交互作用を表現する項である。複数の尺度の尺度値間の比較は、厳密には1次元元的ではないので、その誤差成分を表現する。 e_{miknjl} は測定誤差を表現している。

実験 1

目的

選抜的評価場面における一対比較法によるEQ検査の反応歪曲への耐性の有無を、リッカート法との比較によって明らかにする。

方法

要因 評価場面（非選抜的評価・選抜的評価）と測定方法（リッカート法・新測定法）の2要因とした。まず非選抜的評価場面でリッカート法と新測定法を実施し、1週間の間において選抜的評価場面でリッカート法と新測定法を実施した。また被験者がテストに対して予断を持つことを防ぐ為、常に非選抜場面での検査を最初に実施した。

教示 非選抜的評価場面では「できるだけ素直に、正直に、思ったままお答え下さい。」と、選抜的評価場面では「あなたは今就職活動をしており、どうしても就職したい第1希望の会社の入社試験会場にいます。この性格検査は入社試験の選抜資料の一部として利用されることを想像してお答え下さい。」と、文章と口頭の両方で教示した後、測定を実施した。また項目への社会的望ましさの影響が考慮できるように調査票への記名が求められた。

得点化 リッカート法では、各尺度に対する反応の和を当該尺度得点とした。新測定法では被験者ごと12尺度の因子得点を算出し、当該尺度得点とした。

被験者 関東圏の大学生。非選抜的評価実験では147名、選抜的評価実験では157名。両実験に共通した被験者は140名である。

結果

測定の信頼性 新測定法の検証に入る前に、各下位尺度を測定する6項目の測定に関して、Cronbachの α 係数により信頼性を確認した。本論文の主張内容は新測定法の提案であるため、古典的テスト理論の枠組みでの議論は信頼性係数に関してのみとした。TABLE 3には12尺度における信頼性が記載されている。 ay において信頼性が低いことが伺えるが、これはテスト理論の観点からは項目 x_{17} が他の項目と相関が低い為である。しかしながら実験2において x_{17} を削除しなくても、新測定法が悪いという結果は示されなかった

¹ EPPSでは社会的望ましさが近似的に等しい2つの項目の一対比較が求められる。このとき2つの項目は事前に詳細な項目分析を行うことにより、社会的望ましさ α が等しく、 $E[\alpha_{mik} - \alpha_{njl}]$ が近似的に0となるように設定されている。しかしながらこの方法には、同じ α の値を持つ項目を探すのが非常に難しいことや、 $E[\alpha_{mik} - \alpha_{njl}] = 0$ という仮定が粗い近似でしかないといった問題点がある。本研究で示すモデルでは、事前に α を推定して同程度の社会的望ましさをもった項目を探索する必要はなく、異なる社会的望ましさをもつ項目の対による一対比較が可能である。

² m_i と n_j は異なる次元の下位尺度であり、一見比較は難しいように思われるが、異なる次元であっても自分がどちらの行動をとるかという観点からは比較可能である。たとえば職業選択の際に、事務系か対人関係系かという判断をするが、これは異なる尺度間での比較であり日常的によく行う判断である。仮に m_i と n_j が同一尺度上の点であるとする、差得点の信頼性は著しく低いものとなりむしろ望ましくない。

TABLE 3 12 尺度の信頼性係数

尺度	α
<i>ax</i>	0.754
<i>ay</i>	0.443
<i>az</i>	0.781
<i>bx</i>	0.761
<i>by</i>	0.717
<i>bz</i>	0.934
<i>cx</i>	0.707
<i>cy</i>	0.785
<i>cz</i>	0.685
<i>dx</i>	0.830
<i>dy</i>	0.701
<i>dz</i>	0.702

ので、当該尺度を含めた。それ以外の尺度に関しては、特に *bx*, *dx* が 0.8 以上の高い値を示している。他の尺度に関しては、0.685～0.785 の間の値をとっていることから、本テストの信頼性は許容範囲以内であると判断した。

新測定法の有効性と妥当性

1. 平均値と標準偏差の比較 非選抜的評価場面、選抜的評価場面において、両測定法における被験者の各尺度得点を合計し、新測定法の有効性、妥当性を検証する為にその平均値と標準偏差を求め比較した。その結果を TABLE 4, TABLE 5 に記載する。まずリッカート法に注目すると、評価場面の違いによる平均値は、全ての尺度において選抜的評価場面における平均

TABLE 4 平均値の比較

新測定法					リッカート法				
尺度	非選抜的 評価場面		選抜的 評価場面	効果量	非選抜的 評価場面		選抜的 評価場面	効果量	
<i>ax</i>	-0.198	<	0.173	0.765	15.286	<	17.123	0.411	
<i>ay</i>	-0.065	<	0.059	0.461	12.626	<	18.716	1.427	
<i>az</i>	0.070	>	-0.056	0.311	15.136	<	18.013	0.738	
<i>bx</i>	0.104	>	-0.090	0.623	11.701	<	16.503	1.091	
<i>by</i>	0.093	>	-0.081	0.469	16.041	<	19.348	1.140	
<i>bz</i>	0.078	>	-0.068	0.481	15.272	<	20.316	1.415	
<i>cx</i>	-0.061	<	0.060	0.433	14.735	<	17.471	0.678	
<i>cy</i>	-0.100	<	0.103	0.457	12.088	<	17.090	1.285	
<i>cz</i>	-0.028	<	0.027	0.144	12.476	<	17.368	1.268	
<i>dx</i>	0.044	>	-0.036	0.120	13.905	<	19.368	1.432	
<i>dy</i>	0.030	>	-0.026	0.170	14.823	<	19.084	1.110	
<i>dz</i>	-0.012	<	0.009	0.079	17.116	<	20.045	0.867	
6 対 6					0 対 12				

TABLE 5 標準偏差の比較

新測定法					リッカート法				
尺度	非選抜的 評価場面		選抜的 評価場面		非選抜的 評価場面		選抜的 評価場面		
<i>ax</i>	0.469	<	0.500		4.461	<	4.474		
<i>ay</i>	0.275	>	0.262		4.653	>	3.888		
<i>az</i>	0.398	<	0.411		4.258	>	3.540		
<i>bx</i>	0.290	<	0.329		4.456	>	4.351		
<i>by</i>	0.411	>	0.331		3.025	>	2.786		
<i>bz</i>	0.314	>	0.295		4.169	>	2.918		
<i>cx</i>	0.274	<	0.276		4.279	>	3.801		
<i>cy</i>	0.434	<	0.450		4.103	>	3.693		
<i>cz</i>	0.396	>	0.366		4.085	>	3.640		
<i>dx</i>	0.382	<	0.414		4.217	>	3.411		
<i>dy</i>	0.325	<	0.334		4.131	>	3.555		
<i>dz</i>	0.276	>	0.267		3.618	>	3.147		
5 対 7					1 対 11				

値が高い。新測定法ではリッカート法のように系統的な差はみられず、平均値に差がなかった。TABLE 4には、評価場面の違いに関する標準化された平均値差(効果量:南風原, 2002)も示した。新測定法のほうがリッカート法よりも標準化された平均値差が大きいのは、*ax* “モニタリング”のみであり、あとの11尺度に関しては、新測定法のほうがリッカート法より小さい。

TABLE 5の標準偏差に注目すると、リッカート法では、“モニタリング”という1尺度を除いて、全ての尺度で非選抜的評価場面のものが、選抜的評価場面のものよりも大きい。

2. 相関係数の比較 次に各測定法における、収束的妥当性を検証する為に同一測定法の評価場面間で、尺度間の相関係数を求めた。TABLE 6にその結果を記載する。両測定法の相関係数を比較すると“モニタリング”、“肯定的自己像”、“克己抑制”の3尺度を除く9尺度で新測定法のもとでの相関係数が、リッカート測定法におけるそれよりも大きかった。

考察

まず、平均値に関しては、リッカート法では選抜的評価場面において一貫して平均値が高くなるという傾向がみられたが、新測定法ではそのような系統的な差はみられなかった。本検査では逆転項目は挿入されていないから尺度得点が高いほどポジティブな回答をしていると解釈できる。

標準偏差に関しては、リッカート法においては非選抜的評価場面のほうが標準偏差が一貫して小さい。これは選抜的評価場面において、被験者は一様にポジティブな方向へのアピールを試み、満点が天井効果を生じさせ回答が歪んだものと解釈できる。また選抜的評価場面で標準偏差が大きい“モニタリング”は自分

の感情に対する客観性を表現する尺度であるが、“自分は本当はどうしたいのだろう、どんな気持ちなのだろう”といった、個人によっては、社会的望ましさに大きな違いが生じる項目が含まれていた為に、選抜的評価場面において評価の分散が大きくなったものと考えることができる。

以上、平均と標準偏差の比較から得られた知見を総合すると、選抜的評価場面では、リッカート法において新測定法よりも、被験者は一様に自分を好ましいイメージにアピールしていたと考えることができる。

両測定法の相関係数を比較すると、*ax*, *az*, *bx* における場面間の相関係数がリッカート測定法において高かった。リッカート尺度では個人差が顕著に現れるので、採用場面においてこれらの尺度の重要性の認識には大きな違いがあったと考えられる。

両測定法の相関係数の比較すると、新測定法において尺度間の相関が高いものが多く、収束的妥当性の観点からは、新測定法のほうが選抜・非選抜場面によらず値が安定していることが示された。

EQ検査の信頼性に関しては、Cronbachの α 係数により測定信頼性が確認された。また本検査は国内でベストセラーとなったゴールマン(2000)を基に、自己感情の識別、自己感情の利用、対人感情の認知、対人感情の調整の4領域の区別を踏襲したオリジナルの項目によって構成した。ゴールマン(2000)はEQ理論に関して最も重要視されているSalovey & Mayer(1990), Mayer & Salovey(1993)の情動知能理論を基に記述されており(子安, 2005)、またEQという構成概念は、ゴールマンの著書によって我が国の一般社会に広く認知され、受け入れられている。以上の2点から、本検査の内容的妥当性が保証されると判断した。

一対比較型EQ検査の妥当性という観点から結果を考察すると、以下の3点は本検査の妥当性を支持していると考えられる。(a)選抜的評価場面においても、無意味に平均値が高くない。(b)選抜的評価場面においても、ステレオタイプの分散が小さくない。(c)非選抜的評価場面と選抜的評価場面における相関が高い。特に(c)は新測定法の収束的妥当性の存在を支持していると解釈することができる。一方リッカート測定法においては、尺度得点の収束的妥当性は低く、妥当性の観点から新測定法はリッカート測定法に対して優位性を保持していることが明らかになった。

TABLE 6 選抜場面と非選抜場面の相関係数の比較

尺度	新測定法		リッカート法
<i>ax</i> モニタリング	0.536	<	0.682
<i>ay</i> 適切な自己評価	0.598	>	0.294
<i>az</i> 肯定的自己像	0.446	<	0.590
<i>bx</i> 克己抑制	0.444	<	0.488
<i>by</i> 達成動機	0.534	>	0.360
<i>bz</i> 安定感	0.478	>	0.236
<i>cx</i> 共感性	0.495	>	0.479
<i>cy</i> センシブル	0.638	>	0.258
<i>cz</i> サービス精神	0.658	>	0.373
<i>dx</i> リーダーシップ	0.561	>	0.293
<i>dy</i> アサーション	0.565	>	0.418
<i>dz</i> チームワーク	0.649	>	0.394

9対3

実験 2

目的

実験1で一対比較型EQ検査の有効性が確認されたので、実験2では標本数を増やして本検査の標準化を試みる。同時に代替検査の作成、項目に及ぼす社会的望ましさの影響の考察を行う。

方法

質問紙 実験1で一対比較型EQ検査の有効性が確認されたので、本実験では非選抜場面での一対比較型EQ検査のみを用いた。各被験者ごとの、12尺度における因子得点を求める為の重み行列を安定させる為に、実験1で非選抜場面において新測定法で得点化されたデータも利用して標本数を増やした。

被験者 大学生、会社員、教員、主婦等を含む383名から回答が得られた。性別の内訳は男性174名、女性203名、性別未記入が6名であった。さらに実験1において非選抜の評価実験で回答した147名の被験者も被験者に含まれる為、分析に用いられる被験者数は530名となった。

調査時期 2003年7月から11月までの期間に実施した。

教示 一対比較型EQ検査の実施にあたっては「できるだけ素直に、正直に思ったままお答え下さい。」と文章と口頭の両方で教示した。本検査の実践的な使用を考慮し、実験1のように回答時間に制限を設けなかった。調査票は無記名で実施した。

代替検査 本検査は、何らかの処遇の前後にプリテスト、ポストテストとして実施されることが想定される。このような使用方法に應える為に、質問項目をTABLE 2に表される行列を1セットとするように分類し、全6セットの組み合わせ行列を作成し、これをもとに10組の代替検査を作成した。10組の代替検査の内、併存的妥当性の観点から最も精度の高い折半の組み合わせを探索する為に、折半した検査間で対応する尺度の相関係数を求めた。もちろん各セット間で項目に重複はないので、折半した検査間でも項目に重複はない。

結果

代替検査の妥当性 TABLE 7には、全10通りの折

TABLE 7 代替検査の組み合わせ

test1	S_{136}	S_{123}
test2	S_{245}	S_{456}
<i>ax</i>	0.639	0.553
<i>ay</i>	0.521	0.356
<i>az</i>	0.700	0.510
<i>bx</i>	0.115	0.350
<i>by</i>	0.607	0.430
<i>bz</i>	0.558	0.453
<i>cx</i>	0.682	0.557
<i>cy</i>	0.610	-0.020
<i>cz</i>	0.699	0.277
<i>dx</i>	0.738	0.578
<i>dy</i>	0.163	0.501
<i>dz</i>	0.588	0.558

半の組み合わせにおける尺度間の相関係数のうち、2組分のみ³が記載されている。表において S_{136} は、セット1、セット3、セット6を一組とした代替検査を表現している。代替検査 $S_{136} \cdot S_{245}$ における尺度間の相関係数は全体として0.521~0.738の間の値をとっている。この折半の組み合わせは他の組み合わせの中で最良のものであり、他の組み合わせに比較して並存的妥当性における優位性を保有していた。

尺度*bx*と*dy*は他の尺度と比較すると若干相関が低い。この2尺度で相関が低いのは、代替検査 $S_{123} \cdot S_{456}$ を除く他の代替検査において概してみられる傾向であった。

代替検査 $S_{123} \cdot S_{456}$ は代替検査 $S_{136} \cdot S_{245}$ に比較して、全体的に尺度間の相関係数は低い。しかしながら当組み合わせにおいて、他の代替検査では低く算出されていた*bx*と*dy*における相関係数は比較的高い値をとっている。

項目が影響を受ける社会的望ましさの指標である α_{mik} を、12尺度6項目、計72項目に関して求め、その絶対値を高順位から降順(昇順)に並べ上位5位、下位5位を表示したものがTABLE 8である。社会的望ましさの影響を特に受ける上位5位以内の項目には、総じて対人技能因子*d*を測定する項目が多く含まれていた。具体的には、“集団の中ではリーダーシップを発揮する”、“仲間とは一緒に楽しい時間を過ごす”、“自分は仲間の中でまとめ役である”、“不当な扱いには笑顔で堂々と反論できる”という項目である。それに対して下位5項目では自己制御因子*b*を測定する項目が多く含まれていた。具体的には“将来、大きな仕事をしようと心に決めている”、“怒りや不快を顔にださず速やかに静められる”、“あせりや恐怖に打ち勝って冷静に行動できる”、“失敗の恐れより、成功の喜びを期待

³ 折半信頼性とは項目を無作為に半分に分けて計算するものではなく、内容、領域を考慮して、できるだけ信頼性が高くなるようバランスよく配分するものである。仮にどのように折半しても相関係数が高かったらそれは似たような項目しかないということであり、内容的妥当性の低いテストといえる。

TABLE 8 項目が受ける社会的望ましさの影響

上位 5 位の項目 (降順)		$ \alpha_{mik} $
cy1	微妙な表情や会話の間から相手の気持ちの変化を感じる	0.429
dx3	集団の中ではリーダーシップを発揮する	0.322
dz1	仲間とはいっしょに楽しい時間を過ごす	0.304
dx5	自分は仲間の中でまとめ役である	0.302
dy4	不当な扱いには笑顔で堂々と反論できる	0.292
下位 5 位の項目 (昇順)		$ \alpha_{mik} $
dy5	相手と自分の双方の利益になる結論を見つけるようにしている	0.001
by1	将来、大きな仕事をしようと心に決めている	0.004
bx1	怒りや不快を顔にださず速やかに静められる	0.006
bx2	あせりや恐怖に打ち勝って冷静に行動できる	0.013
bz2	失敗の恐れより、成功の喜びを期待して信じるほうだ	0.015

TABLE 9 モデルの適合度

自由度	χ^2	GFI	AGFI	CFI	RMSEA
2406	6310.861	0.696	0.659	0.6990	0.057

して信じるほうだ”という項目である。因子 d は対人関係における自己評価であり、これらの項目の α の値が高いことは、当該因子が社会的望ましさの影響を強く受けることを示していると考えられる。

モデルの適合度 TABLE 9 には算出された適合度が記載されている。GFI, AGFI, CFI とともに低い値をとっているが、これは本モデルの自由度が 2406 と非常に大きいことを反映しており、後に詳述するように必ずしもモデルの適合が悪いことを意味していない。また RMSEA は 0.057 であり、他の指標に比較してモデルの適合を示唆していた。

⁴ 本テストの評価メカニズムは、以下のようなたとえ話で示される。母集団における鰻丼と牛丼の好みの確率変数 (人が実現値) を、それぞれ $y_{\text{鰻}}$, $y_{\text{牛}}$ と表記し、それをさらに $y_{\text{鰻}} = \alpha_{\text{鰻}} + f_{\text{鰻}}$ と $y_{\text{牛}} = \alpha_{\text{牛}} + f_{\text{牛}}$ と分解しておく。ここで $\alpha_{\text{鰻}}$ と $\alpha_{\text{牛}}$ は母数であり、それぞれ母集団における平均的な鰻丼と牛丼の望ましさである。 $f_{\text{鰻}}$ と $f_{\text{牛}}$ は平均が 0 の確率変数であり、それぞれ鰻丼と牛丼に関する平均からの好みの個人の偏差である。このとき鰻丼と牛丼のどちらをより好むかという質問の回答 $x_{\text{鰻牛}}$ は

$$x_{\text{鰻牛}} = y_{\text{鰻}} - y_{\text{牛}} = (\alpha_{\text{鰻}} - \alpha_{\text{牛}}) + (f_{\text{鰻}} - f_{\text{牛}})$$

と導かれ、これが誤差変数 d と e を省略した場合の (1) 式のたとえの式である。 $\alpha_{\text{鰻}} > \alpha_{\text{牛}}$ と推定されている (鰻丼のほうが高価だから社会的望ましさが高い) 場合には、たとえば「どちらも同じくらい好き」と回答した当該個人は、かなり牛丼が好きであると判定される。逆に、高価な鰻丼を直ちに選ばないことは、集団内の相対評価として、当該個人はそれほど鰻丼が好きでないと判定される。本モデルでは、こうした判定を積み重ね、社会的望ましさの影響を取り除いて、多次元的個人特性を評価する。

考察

実験 2 では一対比較型 EQ 検査のスコアリングを、実験 1 のデータを含めて標本数を増やし実施した。標本数を増加させることにより、各被験者ごとの 12 尺度における因子得点を求めるための重み行列はより安定したものとなっている可能性が高い。

代替検査に関しては、項目セットのとり得る全ての組み合わせで求められた尺度間の相関係数が概して中程度の値をとっていた。性格検査のような典型性検査では並存的妥当性の指標としての相関係数は 0.3 から 0.6 程度で妥当性の根拠として報告されることを考慮すると、本検査で項目の重複の無いような代替検査を作成することの可能性が示されたといえる。

bx , dy の 2 つの尺度は代替検査 $S_{123} \cdot S_{456}$ を除く、全ての代替検査で相関係数が小さく、他尺度に比較して尺度の妥当性が低いものであると解釈できるが、代替検査 $S_{123} \cdot S_{456}$ と代替検査 $S_{136} \cdot S_{245}$ の両検査において、 bx , by の 2 尺度に関してセット 2 とセット 6 の項目を交換すれば、代替検査 $S_{136} \cdot S_{245}$ は並存的妥当性の観点からは最も精度の高い検査となる。

項目への社会的望ましさの影響については、他者技能因子を測定する多数の項目と、他者理解因子を測定する一部の項目が大きな影響を受けることが示唆される結果となった。自己認識因子、自己制御因子を測定する項目では α_{mik} は上位 5 位に含まれる項目に比較して極めて小さい。これは他者との交流という社会的場面では、社会的望ましさを考慮して反応歪曲が生じたためであると考えられる。反応歪曲に耐性をもった一対比較法を用い、さらに非選抜場面を想定して実施した検査であるにもかかわらず、社会的望ましさの影響は混入するようである。ただし本検査では数理的に α_{mik} と分離された 12 尺度の因子得点をスコアリングの対象とするので、具体的な結果の解釈にお

いて項目に対する社会的望ましきによる反応歪曲を考慮しなくて済む。

適合度指標は、慣習的な基準に照らし合わせるとき、本モデルを支持しなかった。しかし自由度が5369(豊田・真柳, 2001), 1225(豊田, 2001)という非常に大きい値をとるモデルでは、低いGFI, AGFIの値は必ずしもモデルの不適合を示さない。この為自由度が非常に大きい場合、自由度による影響が少ないRMSEAを報告すべきである(豊田・真柳, 2001)との研究もある。自由度が大きいモデルにおいて、RMSEAが唯一の指標であるかということに関しては議論の余地があるが、先行研究に従うならば本研究ではRMSEAは0.057であり他の指標に比較して悪くなく、モデルの適合が十分に示唆されていると考えられる。CFIの低さに関しては、GFIと同様に自由度が大きいために、 χ^2 値が大きくなり、値が低くなったと考えられる。

総合考察

本研究では選抜的評価場面における典型値測定において、反応歪曲に対して耐性をもった一対比較型EQ検査の有効性の確認、標準化を目的としていた。

有効性に関しては効果量、収束的妥当性といった指標が、選抜的評価場面における一対比較型EQ検査の反応歪曲への耐性を明らかにした。

標準化に関しては標本数を増やして重み行列を安定させたので、素データさえあれば標準得点と、重み行列の積和で被験者ごとに因子得点が算出できる。従って構造方程式モデリングを実行できるようなソフトウェアが手元になくとも標準化された本検査を使用することが可能である。

代替検査に関しては一般的な典型値検査における併存的妥当性の基準は満たしていることが示された。代替検査として一対比較型EQ検査を折半したもので処遇の効果を検討することは、 bx , dy の2尺度に配慮するならば、実務、研究レベルでの使用に十分耐えうるものであることが示された。

実験1ではEQ検査に関して測定の信頼性を示したが、妥当性研究に関しては明示していない。これは(a)著者の一人はEQ教育に長年かかわっており、その経験に照らしても作成されたテストは内容的に妥当であると判断されたこと、(b)妥当性研究は終わることのない不断の確認作業と称されるように、多数の追研究によって検証されるものだからである。EQ検査に関する妥当性研究に関して、今後複数の研究がなされることが期待される。

さらに一対比較型EQ検査の実践的使用に関する研究の展望として、社会的望ましき α_{mik} を困難度母数とみなし、ラッシュモデルのような項目反応モデルを適用することの可能性の検証が挙げられる。

一対比較型EQ検査は反応歪曲への耐性を保有しているだけでなく、併存的妥当性が高い代替検査を提供することも可能であり、処遇の効果に関して精度の高いアセスメントを可能にするような非常に稀有な典型値検査であるといえよう。

引用文献

- Edwards, A. L. 1959 *Manual, Edwards Personal Preference Schedule*. New York: The Psychological Corporation.
- エドワーズ, A. L. 肥田野 直・岩原信九郎・岩脇三良・杉村 健・福原真知子(訳編) 1970 EPPS 性格検査手引 日本文化科学社 (Edwards, A. L. 1970 *Edwards Personal Preference Schedule*. Tokyo: Nihon Bunka Kagakusha.)
- ゴールマン, D. 土屋京子(訳) 1996 EQ—こころの知能指数— 講談社 (Goleman, D. 1996 *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 (Goleman, D. 2000 *Working with emotional intelligence*. New York: Bantam Books.)
- 南風原朝和 2002 心理統計学の基礎 有斐閣
- 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
- 子安増生 2005 社会的知能 中島義明・重柁算男・箱田裕司(編) 新・心理学の基礎知識 有斐閣 p.304.
- Mayer, J. D., & Salovey, P. 1993 The intelligence of emotional intelligence. *Intelligence*, 17, 433-442.
- 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 2000 共分散構造分析 [応用編]—構造方程式モデリング— 朝倉書店 (Toyoda, H.

2000 *Application of covariance structure analysis : Structural equation modeling*. Asakura Shoten.)

豊田秀樹 2001 確認的ポジショニング分析—印象変化の発見と認知マップの評価の為の多変量解析法— 心理学研究, 72, 399. (Toyoda, H. 2001 Positioning method, change of image, cognitive map, three-mode multivariate data, semantic differential (SD) method. *Japanese Journal of psychology*, 72, 399.)

豊田秀樹・真柳麻誉美 2001 繰り返し測定を伴う実験のための因子分析モデル—アイスクリームに関する官能評価— 行動計量学, 28, 1-7. (Toyoda, H., & Mayanagi, M. 2001 A factor analysis model for an experiment having repeated measures: A sensory evaluation of ice cream. *Japanese Journal of Behaviormetrics*, 28, 1-7.)

続 有恒・村上英治(編) 1975 心理学研究法 9 質問紙調査 東京大学出版会
(2004.8.30 受稿, '05.4.14 受理)

付 録

尺度の内容と項目例

ax：モニタリング 自分の感情をモニターする能力。自分自身の本心に気がつかず、本意と異なる行動をとり続けることは、神経症の主原因といわれている。ビジネスパーソンとして生き生きと活動するために、「自分は本当はどうしたいのだろうか」と考え、心の中の本当の気持ちと言動が食い違っていないかを自己監視する能力。

ay：適切な自己評価 自らの得手不得手を客観的に評価する能力。自分の特性の中で、自信のある領域を強調し、欠点を自覚してカバーする技量。たとえば孫子の兵法「敵を知り、己を知れば百戦危うからず」を認識している度合い。

az：肯定的自己像 自分自身を価値あるものとする感情。自尊感情は、言動や意識や態度を基本的に方向付け、自分自身の存在や仕事を基本的に価値あるものとして評価し、信頼する。肯定的自己像は、積極的・意欲的にビジネスに向かわせ、精神的健康を維持する。

bx：克己抑制 人は怒りや不快に任せて、我を忘れ

た行動をすることによって、ビジネスパーソンとして失敗する。不快な事態・不利な状況にも我を忘れず、冷静さを保ち、単調な作業も粘り強くがんばり続ける能力。

by：達成動機・統制感 ある優れた目標に対して、それを高い遂行レベルで完遂する決意を伴った感情。また自らが努力して環境に働きかければ、自分は有効に環境を変え得るという自信。クラーク博士が札幌農学校を去るときに言った「Boys, be ambitious! (少年よ、大志を抱け!)」に相当する意志。

bz：安定感・柔軟性 新しい環境に、または望ましくない状況変化に対して、柔軟に楽観的に適応する能力。安定的な自我を有する同僚は、職場の仲間からの評価が高く、ビジネス・パートナーとして部下として頼りにされる。

cx：共感性 他者の立場におかれた自分を想像することにより、他者の感情を察する能力。共感性は他者の理解を深め、円滑な対人関係の形成を促す。

cy：センシブル 政治的観察力に優れ、重要な社会的ネットワークを認識し、決め手となる力関係を見抜く能力。職場の価値観や暗黙のルールを読み取る。人間関係のダイナミクスを予測し、効果的にアプローチする能力。

cz：サービス精神 他者の利益のために意図的になされる利他的行動を自発する能力。向社会的行動を通じて、仲間を喜ばせ、接客に長け、周囲と良好な関係を形成することに貢献する能力。

dx：リーダーシップ 集団の目標達成、および集団の維持・強化のために効果的な行動を自発する能力。強力な牽引力を発揮する場面と、民主的にメンバー間の関係を調節する場面を適切に行き分けられることにより、効果的な指導力が発揮される。

dy：アサーション 自分の要求や主張を、他者の利益にも配慮しつつ、効果的かつ論理的に他者に提示できる能力。アサーション能力は、本人の精神的衛生に寄与し、ストレスを低減させる。同時に他者からは、交渉事態に有能であるとの評価を受ける。

dz：チームワーク 基本的興味が人・仲間に向けられ、仲間との良好な関係を保つこと自身に喜びを感じる。自分が高い評価を受ける事と同等に、あるいはそれ以上に仲間全体として高い評価を受けることを重視し、向上しようとする能力。

Improvement of Emotional Intelligence Quotient (EQ) Test Used for Employment

HIDEKI TOYODA (WASEDA UNIVERSITY), IKKO KAWAHASHI (GRADUATE SCHOOL OF LETTERS, ARTS AND SCIENCES, WASEDA UNIVERSITY)
NOBUTAKE MATSUSHITA (EQ JAPAN) JAPANESE JOURNAL OF EDUCATIONAL PSYCHOLOGY, 2005, 53, 456—466

In the present study, the robustness of the Emotional Intelligence Quotient (EQ) Test was examined against response bias, using paired comparison methods expressed by structural equation modeling. The EQ test was then standardized. In Experiment 1, paired-comparison and Likert-type EQ tests were administered to 140 participants in recruitment and non-recruitment situations, in order to evaluate the robustness of the test. Standard deviations, effect size, and convergent validity showed that in both situations, the paired comparison test was more robust against response bias than the Likert-type test. In Experiment 2, in order to stabilize the scoring, data obtained from another set of 383 participants were added to the data from Experiment 1. Items on the test were then divided into 10 subsets, yielding parallel tests, and the correlation of scales among those tests was calculated. The parallel tests were showed to have concurrent validity.

Key Words : paired comparison, Emotional Intelligence Quotient (EQ) Test, structural equation modeling