

SEMによる反応歪曲に抗する一対比較型テストモデル

豊田秀樹* 川端一光* 渡辺徹** 日高美英子**

豊田・川端・松下(2005)の拡張モデルとして、社会的望ましさに関する個人差である f_0 を導入し、かつ尺度を構成する項目のタウ等価測定の制約を外したモデルを考案した。予備調査で作成された8つのコンピテンシー尺度を用いて一対比較検査を構成し、これをA・Bの2検査に折半し標準検査を作成した。400名の被験者に対して両検査を実施した結果、検査の信頼性・妥当性が確認された。妥当性研究では標準化検査の反応歪曲への耐性を検討するために、新測定法とリッカート法による検査が、選抜的評価場面、非選抜的評価場面の2場面を想定して、199名の被験者に対して実施された。その結果、新測定法は評価場面にかかわらずリッカート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。

キーワード：一対比較コンピテンシー検査、反応歪曲、社会的望ましさ、構造方程式モデリング

問題と目的

進路指導や形成的評価のための単元テストなど、受験者自身への情報提供を目的に利用される非選抜場面では、心理テストの測定に対してあからさまな反応歪曲を心配する必要性は少ない。それに対して入社試験や昇進試験など、受験者自身の処遇に影響する選抜的評価場面においては、知能・学力検査のような最大値測定(Thorndike, 1997)はともかく、特に性格検査に代表される典型値測定では反応歪曲の悪影響(岩脇, 1973)が懸念される。

豊田・川端・松下(2005)では、典型値測定における反応歪曲に抗するために、SEM(Structural Equation Modeling)(狩野・三浦, 2002)の測定方程式を利用した一対比較テストを提案している。

提案されたテストはSEMにおける

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

という測定方程式で表現されている。これは7件法による観測変数 x_{miknjl} に関する測定方程式である。右辺の添え字 m と n は下位尺度を包括する少数の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。また k と l は項目番号を示す添え字である。

この右辺は、 α_{mik} と α_{njl} が母数(定数項)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。 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。ただし α_{mik} には社会的望ましさばかりでなく、客観的に当てはまっている人の割合も含まれる。また f_{mi} は(f_{nj} は)、尺度 mi における(尺度 nj における)個人の尺度値である。 d_{minj} は、尺度 mi と尺度 nj の交互作用であり、 e_{miknjl} は、測定誤差である。

上述の測定方程式は領域と尺度と項目という高次の構造を表現しているが、心理検査は通常、尺度と項目という構成で作成されるから、(1)式の本質は

$$x_{ikjl} = (\alpha_{ik} - \alpha_{jl}) + (f_i - f_j) + d_{ij} + e_{ikjl} \quad (2)$$

と簡略化される。

豊田ら(2005)は(1)式を用いた一対比較型検査が社会的望ましさによる反応歪曲に対して耐性を保持していることを、同一項目により作成されたリッカート型検査との比較により明らかにし、SEMによる一対比較型検査の有効性を報告している。

ただ(2)式には更なる2つの改良の余地がある。1つは、項目の社会的望ましさが全ての受験者に同様に影響していると仮定している点である。社会的規範の一部ともいえる項目の社会的望ましさの体系を比較的強固に受け入れている個人と、全く気にしない個人の違いを表現することは、個人特性(あるいは個人の状態)の記述という観点からは有効かもしれない。

もう1つは、(2)式は因子負荷量が1か-1かに固定された平均構造のある確認的因子分析の形式で表現されている点である。一対比較モデルは、特性の差を表現する際に1か-1の係数を利用することが多い。しかし因子負荷量の等しい因子分析モデルは、全ての項目

* 早稲田大学 〒162-8644 東京都新宿区戸山1-24-1
早稲田大学文学部心理学教室

** 株式会社EIリサーチ
〒162-8616 東京都新宿区神楽坂3-2 神楽坂Kビル7F

の信頼性が等しいというタウ等価測定のモデル化である。現実には、必ずしもタウ等価測定が成立立つとは限らないので因子負荷量を項目の信頼性にあわせて推定することは有効かもしれない。

また(2)式における f_i は、モデルとしてはイプサティップスコア¹として導入されていないが、タウ等価の仮定が入っているために、通常のSEMのソフトで構成概念スコアを推定すると推定量がイプサティップスコアになってしまふという弊害がある(これを回避するようにテスト得点を定義することは可能であるが、通常のソフトの範囲でできないのは不便である)。因子負荷量を自由母数にすることにより、モデルばかりでなく、構成概念スコアの推定値もイプサティップスコアでなくなる。

そこで本論文では、豊田・川端・松下(2005)で提案された(2)式の一対比較型テストモデルを

$$x_{ikjl} = (\alpha_{ik} - \alpha_{jl})f_0 + (\alpha_{ik}f_i - \alpha_{jl}f_j) + d_{ij} + e_{ikjl} \quad (3)$$
と拡張し、それに従ってオリジナルのテストを作成しその効果を調べる。

一対比較テストモデル： x_{ikjl} は項目 (i, k) と項目 (j, l) の比較を表現する観測変数である。 f_0 は、平均 1 で正の値をとることが期待される潜在変数であり、値が大きいほど α_{ik} による社会的望ましさの体系を尊重し、0 に近いほど社会的望ましさの体系を気にしないという個人傾向である。また「自分に当てはまっていると素直に言える」という別の解釈も可能であるが、この表現は社会的望ましさの体系を内面化した被験者の状態を表現していると考えれば、 f_0 は社会的望ましさの体系を表現していると解釈することができる。 f_0 の導入によって(2)式から増加する母数は、その分散だけであり、適合度の観点からは大きな影響はない予想される。

次に $\alpha_{ik}(\alpha_{jl})$ は、項目 ik (項目 jl) の因子負荷量であり、その値が高いほど当該項目の信頼性は高いと解釈する。また前述の通り f_i (f_j) は因子を表現している。この改良は導入した時点でイプサティップでないテスト得点を構成できるという長所を有するので比較成績が豊田ら(2005)と同程度か、それ以上であれば導入の効果があると判断できる。また d_{ij} は交互作用を、 e_{ikjl} は測定誤差をそれぞれ示している。

¹ イプサティップスコアとは被験者毎に尺度を考え、変数を尺度上に位置づける得点である(豊田, 2000)。イプサティップ変数はその和が一定に調節されており、例えば 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 6 で一定となっている。

本研究は豊田ら(2005)における測定方程式に個人差 f_0 を導入し、かつ $\alpha_{ik} = \alpha_{jl}$ というタウ等価測定の仮定を緩めた SEM による一対比較測定法の有効性・妥当性検証の試みであり、従って 8 つのコンピテンシーに関する尺度作成を主目的とするものではない。

予備調査

目的

後に標準化を試みる 8 つのコンピテンシー尺度に関して、項目作成・項目分析・項目選択を行い、次に選定された尺度に関してその信頼性・妥当性を確認する。8 つのコンピテンシー並びにそれに基づくテスト項目は、企業人として必要な資質として、コーチングに長年携わってきた複数の専門家によって、実際の企業人からの聞き取り調査の内容から、KJ 法によりいくつかの類似した概念群から抽出・選定された。8 つのコンピテンシーに関する既知の心理学的理論背景、先行研究は現時点ではない。

コーチングによる効果の確認・検証という観点から、実践場面ではプリテスト、ポストテストに利用できる代替性の高い 2 つの検査が求められている。本研究では実践場面における代替検査の重要性と、一対比較測定法による検査の信頼性の検証を考慮して 2 つの検査を作成する。

各項目の内的整合性等、項目の信頼性に関する議論は一対比較法による尺度では考察できないため、予備調査では後に一対比較型調査法で用いられるものと同じ項目を、リッカート法によって表現される項目として被験者に提示する。

方法

被験者：予備調査における被験者は都内の私立大学生 390 名であった。このうち 1 名のデータに不備がみられたのでこれを削除した。

コンピテンシーの選定：専門家による評定によって抽出された 8 つのコンピテンシーは以下の通りである。第 1 尺度：“現状認識力”，第 2 尺度：“ビジョン創出力”，第 3 尺度：“達成行動力”，第 4 尺度：“自己管理力”，第 5 尺度：“メンタルタフネス”，第 6 尺度：“人間関係構築力”，第 7 尺度：“コミュニケーション力”，第 8 尺度：“組織適応力”という 8 つのコンピテンシーを測定対象とした。各コンピテンシーの内容は TABLE 1 に記載されている。

項目分析：各尺度に 5 つの予備項目を付け加え 19 項目とした(本調査では 2 × 7 項目となる)。従って予備調査では全 152 項目が作成・実施された。項目の選択に際

しては先ず平均値と標準偏差が極端な値でないことを確認した後、残った項目に対して各尺度で1因子の因子分析を実行し、因子負荷量の大きい上位14個の項目を選択するという方法を採用した。また母数の推定法には最尤法を採用した。

項目の折半：項目分析の結果得られた各尺度14個の項目を用いて、代替検査信頼性の観点から7項目ずつの最適な折半の組み合わせを探査した。すなわち尺度内の一貫性と、尺度間の相関が最大となる項目の折半を探査した。

結果と考察

TABLE 2には尺度毎にCronbachの α 係数と代替検査間の相関係数が記載されている。 α 係数を参照すると、代替検査を構成する全ての下位尺度においてその値は検査Aで0.83～0.86、検査Bで0.82～0.85の間をとっており、慣習的な基準から尺度の信頼性が確認される結果となった。また尺度の代替検査信頼性の指標である折半検査間における、下位尺度毎のPearsonの積率相関係数も0.82～0.89の間の値を取っており、尺度の信頼性が確認される結果となった。

また抽出されたコンピテンシー並びに、作成された項目群は専門家によって作成されており、内容的妥当性が保証されているものと判断した。

一対比較テストの標準化（本調査）

目的

予備調査において8つのコンピテンシー尺度の信頼性・妥当性が確認されたので、次に(3)式に基づくSEMによる一対比較型テストを作成し、これを標準化する。

一対比較測定：一対比較型検査では、8つのコンピテンシーから2つのコンピテンシーを選んで、順列で比較する。従って比較回数は56回である。比較回数は観測変数の数と一致するので、社会的望ましさが含まれる a_{ik} も56個推定する。従って8のコンピテンシーで、それぞれ7項目、テスト全体で56項目を作成する。

特定のコンピテンシーは、自分以外の全てのコンピテンシーと2回ずつ比較される。合計56 (=8×7) 個の項目で、56回の一対比較測定を行うわけであるから、同じ組み合わせが登場しないように、1つの項目が、テスト中で必ず2回ずつ出現する。TABLE 3には本調査で用いられる一対比較計画が記載されている。例えば、TABLE 3の2行1列目の2312というセルは、コンピテンシー2である“ビジョン総出力”の項目番号3と、コンピテンシー1である“現状識別力”の項目番号2の一対比較を表現している。

TABLE 1 選定された8つのコンピテンシー

尺度名	尺度の内容
1 現状認識力	的確に状況を把握し、適切な判断ができる能力
2 ビジョン創出力	自律的に行動し、目的や目標を明確化できる能力
3 達成行動力	状況に合わせて適切に感情をコントロールできる能力
4 自己管理力	目標に向かって自発的に取り組み続けることができる能力
5 メンタルタフネス	プレッシャーやストレスに対して強い精神力を維持できる能力
6 人間関係構築力	思いやりを持ち、積極的に他人と協働ができる能力
7 コミュニケーション力	他人の話に耳を傾け、自己表現が適切にできる能力
8 組織適応力	組織風土や慣習を感じ取り、相応に行動変容できる能力

TABLE 2 尺度の信頼性・代替検査信頼性

尺度番号	尺度名	信頼性A	信頼性B	代替検査信頼性
1	現状認識力	0.83	0.82	0.82
2	ビジョン創出力	0.83	0.82	0.85
3	達成行動力	0.85	0.85	0.82
4	自己管理力	0.85	0.85	0.83
5	メンタルタフネス	0.84	0.82	0.86
6	人間関係構築力	0.85	0.83	0.89
7	コミュニケーション力	0.85	0.85	0.85
8	組織適応力	0.86	0.85	0.85

TABLE 3 本調査の一対比較計画

尺度名	1	2	3	4	5	6	7	8
1 現状認識力	—	1821	1731	1641	1551	1461	1371	1281
2 ビジョン創出力	2312	—	2132	2842	2752	2662	2572	2482
3 達成行動力	3513	3423	—	3243	3153	3863	3773	3683
4 自己管理力	4714	4624	4534	—	4354	4264	4174	4884
5 メンタルタフネス	5115	5825	5735	5645	—	5465	5375	5285
6 人間関係構築力	6316	6226	6136	6846	6756	—	6576	6486
7 コミュニケーション力	7517	7427	7337	7247	7157	7867	—	7687
8 組織適応力	8718	8628	8538	8448	8358	8268	8178	—

方法

評定は2つの項目、 x_{ik} と x_{jl} を比較して、どちらが被験者に当てはまっているかを考えさせ、その評価を対応する7件法の尺度上で回答させた。FIGURE 1は刺激文と、対応する尺度を図示したものである。また“ a が b より非常に当てはまる”，“ a が b より当てはまる”，“ a が b よりやや当てはまる”，“どちらともいえない”，“ b が a よりやや当てはまる”，“ b が a より当てはまる”，“ b が a より非常に当てはまる”に、それぞれ7点から1点を与えた。

i と j はコンピテンシーを示す添え字($i \neq j$)であり、1から8までの値をとる。 k と l は、それぞれ i と j の下位項目を構成する項目番号であり、1から8まで動く。実際の項目数は7であるが、($k=i$), ($l=j$)の比較は行わず、例えば($k=i$)以降では $k'=k+1$ と表記することによって項目を表現する為、 k と l は1から8までの値をとりうる。これはTABLE 3の表記に明らかである。またコンピテンシーの入れ子であるから、それぞれ i と j にネストし、正確に書くなら $k(i)$ と $l(j)$ である。ただし、入れ子の関係は変化しないので、以後省略して表記する。従って観測変数は x_{ikhj} と表現される。

一対比較検査の教示文：「 a と b を比較してどちら

- a : イライラしたときでも平静さを保てる
b : 本当に言いたいことが伝わるよう工夫して話をする

非常に当てはまる	aがbより当てはまる	やがbより当てはまる	どちらともいえない	やが当てはまる	bがaより当てはまる	非常に当てはまる
----------	------------	------------	-----------	---------	------------	----------

1. | | | | | | | |

FIGURE 1 項目例

があなたに当てはまるのかを考えて下さい。次に、当てはまる程度の差として最も適当と思われるものを選択肢より1つだけ選んで回答用紙にマークしてください。もし a と b が同じぐらい当てはまる場合、あるいはどちらも当てはまらない場合には「どちらともいえない」にマークしてください。

測定方程式

(3)式の測定方程式を用いる。ただし $a_i = a_{ik}$ ($a_j = a_{jl}$)というタウ等価測定の強い仮定が解除され、かつ因子スコアがイプサティブでなくなるという利点を考慮して、(3)式中の a_{ik} (a_{jl})を自由推定した。

f_0 は、期待値が1であり負の値をとらない潜在変数として導入した。

一対比較型検査：予備実験で信頼性・妥当性が確認されたリッカート型検査A・Bそれぞれに関して、TABLE 3の一対比較計画に基づき一対比較型検査を作成した。

被験者：都内の私立大学生に対して実施された。1人の被験者は検査A、検査Bの両方を受検する。検査A、検査Bの順番で受検した被験者が190名(平均年齢、20.57歳)，検査B、検査Aの順番で受検した被験者が210名(平均年齢、20.96歳)，合計400名(平均年齢、20.76歳)であった。また性別の内訳は男性が201名(平均年齢、20.95歳)女性が199名(平均年齢、20.57歳)であった。

分析プログラム：SAS for Windows V8 (SAS Institute Inc)におけるCALISプロシジャーを用いた。

結果と考察

モデルの適合度：一対比較型検査を実施した結果、両検査におけるモデルの適合はTABLE 4のようになつた。

代替検査Aにおいて $GFI=0.788$, $AGFI=0.766$, $RMSEA=0.039$, 代替検査Bにおいて $GFI=0.795$, $AGFI=0.772$, $RMSEA=0.038$ であった。適合度指標は、慣習的な基準に照らし合わせると、本モデルを支持しなかった。

TABLE 4 一対比較型検査A・Bの適合度

適合度指標	検査A	検査B
GFI	0.788	0.795
AGFI	0.766	0.772
RMR	0.774	0.607
RMSEA	0.039	0.038
χ^2 値	2413.332	2331.204
自由度	1497	1491

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)。本検査の自由度は検査Aで1497, 検査Bで1491と非常に大きく、先行研究の見解に従うならばRMSEAを参照すべきである²。RMSEAを参照すると両検査で慣習的基準を満たしており、採用されたモデルの適合が必ずしも悪いものではないことが伺える。

社会的望ましさの影響:(3)式の測定方程式で定義されているように、 α_{ik} を比較対を構成する項目ごとに算出した。検査が選抜に利用されないような形成的な自己評価を行う際にも、人は自分を良く見せようとする傾向があり、例えばEPPS(エドワース, 1970)が対処を試みた、形成的自己評価において影響するであろう社会的望ましさの影響が α_{ik} には混入している。

この値に基づいて社会的望ましさの影響を強く受けた尺度を、 α_{ik} の平均の観点から両検査を総合した上で順序付けた。TABLE 5は α_{ik} の平均値に基づいて算出された、両検査をプールした場合の尺度の順位関係を

TABLE 5 項目が受ける社会的望ましさの影響³

尺度名(降順)	平均	標準偏差
5 メンタルタフネス	0.399	0.169
4 自己管理力	0.385	0.267
8 組織適応力	0.309	0.239
2 ビジョン創出力	0.305	0.148
7 コミュニケーション力	0.290	0.145
3 達成行動力	0.288	0.213
1 現状認識力	0.261	0.132
6 人間関係構築力	0.257	0.170

² また Heywood case を避ける為の制約から検査Aよりも検査Bの自由度が小さくなっている。

³ ここでは各尺度を構成する項目にそれぞれ定義される α の平均と標準偏差を求めている。各尺度毎に7つの項目が存在し、かつ2つの検査があるので、 7×2 個のデータに基づいて統計値を求めている。

表している。

α_{ik} の平均値の高い上位2尺度は“5：メンタルタフネス”と“4：自己管理力”であり、それぞれ平均値が0.399, 0.385、標準偏差が0.169, 0.267であった。一方下位2尺度は“1：現状認識力”, “6：人間関係構築力”であり、それぞれ平均値が0.261, 0.257、標準偏差が0.132, 0.170であった。

メンタルタフネス尺度と自己管理力尺度を構成する項目群は、組織に属する企業人として特に必要とされる、ストレスマネジメント能力、状況に応じた感情のコントロール能力の有無を判断させていた。一方下位2尺度である現状認識能力、人間関係構築力を構成する項目群は、自分を取り巻く状況の的確な認知と、状況への対応、日常場面における対人行動に関する判断を求めていた。

上位2尺度と、下位2尺度の差異は、前者が企業人として優秀な人材像を表現する一方で、後者は日常場面における個人の自然な人物像を表現する点にある。後者の観点も確かに企業人として必要とされる特性ではあるが、自己への認知は高いが企業において積極的な行動を起こせない人よりも、挫折や困難に強く、かつ状況に応じて感情をコントロールできる個人の方が、企業にとっては優秀な人材であると考えられる。

従って上位2尺度は下位2尺度と比較するとき、社会的望ましさの影響を強く受ける尺度であると考えることができる。全尺度における標準偏差は0.132～0.267という小さい区間の間の値をとっており、被験者の反応が一様であることが示されている。標準偏差が一様に小さいということは、 α_{ik} が集団における価値判断を適切に表現していることの証拠であり、社会的望ましさを表現する α_{ik} の導入の妥当性を示唆する結果であると考えることができる。

f_0 の分布: f_0 は社会的望ましさの体系を尊重する個人傾向として導入した。 f_0 の分布が存在しているか(個人差があるか)否かを検証するために、推定値の分散の有意性検定を5%水準の z 検定で実行した。

TABLE 6には検査A、検査Bそれぞれにおける分散の推定値とその z 値及び p 値が記載されている。検査Aにおいて標準偏差の推定値は0.261であり5%水準

TABLE 6 検査A・Bにおける f_0 の標準偏差

検査	標準偏差	z値	p値
A	0.261	4.40 *	0.015
B	0.371	5.22 *	0.026

* は5%水準で有意であることを示す。

で有意であった($z=4.41, p<.05$)。また検査Bにおいて標準偏差の推定値は0.371であり、5%水準で有意であった($z=5.22, p<.05$)。社会的望ましさの体系に対して、それを受け入れる個人と受け入れない個人が平均1を中心に有意に散らばっていることが示される結果となった。従って本モデルで導入された f_0 は社会的望ましさの体系に関する個人傾向として統計的に意味のある指標であることが確認された。

代替検査信頼性：一対比較型検査の信頼性を確認するために、検査A・Bの8つの各尺度で、検査間の相関係数を求めた。また f_0 を導入することによるモデルの信頼性の変化を検証するために、 f_0 を推定しない場合における検査間の相関係数も同時に算出した。

TABLE 7には算出された相関係数が記載されている。 f_0 を推定する場合の検査間の相関係数は0.52~0.73の間の値をとる一方で、 f_0 を推定しない場合の相関係数は0.69~0.86の間の値をとっていた。 f_0 を推定しない場合の検査間の相関係数が高いという傾向が示された。

f_0 を推定しない場合では全ての尺度において相関係数が0.69以上であることから、慣習的基準を鑑みるならば本検査の代替検査信頼性は確認されたといえる。一方 f_0 を推定した場合では、 f_0 を推定しない場合と比較すると全体的な相関係数の値は低くなっている。これは f_0 を導入することによってモデルが複雑になり因子スコアが不安定になったために検査間の相関係数が低くなったと解釈することができる。しかし慣習的基準に照らし合わせると、 f_0 を推定するモデルの信頼性に関して致命的に問題があるとは言い難い。

標準化についての総合考察

モデルの適合度に関しては先行研究(豊田・真柳(2001)、豊田(2001))による考察から、特にRMSEAに基づく判断が適切であり、この観点からSEMによって構成された一対比較モデルの適合が良好であることが示された。また豊田・川端・松下(2005)に対する拡

TABLE 7 代替検査信頼性

尺度	f_0 推定	f_0 非推定
1 現状認識力	0.60	0.77
2 ビジョン創出力	0.59	0.69
3 達成行動力	0.60	0.69
4 自己管理力	0.73	0.72
5 メンタルタフネス	0.54	0.86
6 人間関係構築力	0.56	0.76
7 コミュニケーション力	0.52	0.71
8 組織適応力	0.67	0.69

張として導入された f_0 と a_{ik} の存在もモデルの適合という観点からは適切に機能することが明らかになった。

特に社会的望ましさの体系に対する個人傾向として導入された f_0 に関しては、その標準偏差の有意性検定の結果から、社会的望ましさの体系に対する考え方の個人差を適切に捉える指標であることが明らかになった。

またSEMによる一対比較型検査の信頼性を、A・B両検査の因子スコア間の相関係数を求ることによって確認した。代替検査信頼性の観点から、 f_0 の導入の有無にかかわらず本検査の信頼性は許容範囲内であることが示された。ただし f_0 を導入した場合のテスト得点の解釈において、導入しない場合からの信頼性の減少を考慮する必要がある。

妥当性研究

目的

最初に選抜的評価場面における一対比較型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し、豊田・川端・松下(2005)の結果が再現されるかを確認する。次に社会的望ましさへの個人差として導入された f_0 の分布を記述し、その妥当性を検証する。かつタウ等価測定の制約を外したことによるモデルの適合も検討する。

最後に評価のされ方に関する内省報告を被験者に求め、モデルの妥当性について質的に検討を行う。

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2×2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週間の間を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。同一被験者が2つの評価場面と2つの測定方法を経験する2要因被験者内計画であった。

教示：非選抜的評価場面では「素直に、正直に、思ったままにお答えください。」と教示した。一方、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るとします。そして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。」と教示した。

調査票：標準化された検査Aのみを用いた。

実施期間：2005年7月6日、13日の2日間で実施された。

被験者：心理学を専攻する都内の私立大学生に対して実施した（本調査の被験者とは異なる）。選抜的評価場面では212名（平均年齢、23.84歳）、非選抜的評価場面では208名（平均年齢、23.44歳）であった。また両評価場面においてもれなくリッカート測定法に回答した被験者数は199名であり、一対比較測定法に回答した被験者数は200名であった。

因子スコアの推定：テストモデルの標準化の過程で測定方程式中の a_{ik} (a_{jl}) と α_{ik} (α_{jl}) は求められているため、妥当性研究ではこれらを固定母数として因子スコアを算出した。ただし f_0 の期待値と分散は自由母数とし、標準化データにおける母数の推定値からの差を確認することで f_0 に関する妥当性について考察する。

分析プログラム：SAS for Windows V8 (SAS Institute Inc) における CALIS プロシジャーを用いた。

結果と考察

各測定法の平均値・標準偏差・効果量：TABLE 8には選抜的評価場面・非選抜的評価場面におけるリッカート測定法、一対比較法のそれぞれの尺度得点の平均値・標準偏差、そして標準化された平均値差（南風原、2002）である効果量が示されている。

まず評価場面の違いによる尺度得点の平均値の差に注目する。TABLE 8 を参照するとリッカート法では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値が高くなっている。一方、一対比較法ではリッカート法のような系統的な差はみられないという結果になった。

場面間における尺度得点の平均値に関する考察を補助する指標として、標準化した平均値差である効果量を算出した。その結果リッカート法では1.473～1.834の間の値をとる一方で、一対比較法では-0.48から0.56の間の値をとっていた。従ってリッカート法は全ての尺度において一対比較法よりも効果量が正に大き

いという結果となった。一方、一対比較法では効果量においても系統的な差はみられなかった。これらの結果からリッカート測定法に比較して一対比較法は場面間で被験者の反応を変化させないという性質を保持することが示されたといえる。

次に尺度得点の標準偏差に注目すると、一対比較法はリッカート測定法に比較して、非選抜的評価場面と選抜的評価場面との差が僅かであった。リッカート測定法において、選抜的評価場面で標準偏差が小さくなる傾向は豊田・川端・松下（2005）においても報告されており、選抜的評価場面で平均値が系統的に上昇していることを併せて考察するならば、被験者が社会的望ましさを考慮して、一様に良い自分を装って反応をしたため標準偏差が小さくなったと解釈することができる。

各尺度における場面間の相関係数：一対比較型検査の収束的妥当性を考察するために、各尺度における場面間の相関係数を算出した。またここでも f_0 を導入することによるモデルの妥当性の変化を検証するために、 f_0 を推定した場合の相関係数と、推定しない場合での相関係数を算出した。その結果は TABLE 9 に示されている。

TABLE 9 から f_0 を推定した場合の一対比較測定法においては場面間の相関係数が0.29～0.47の間の値をとっていたことがわかる。一方で f_0 を推定しない場合の一対比較法においては場面間の相関係数は0.39～0.52の値をとっていた。またリッカート測定法における場面間の相関係数は0.28～0.44の間の値をとっていた。従って f_0 を推定しない一対比較測定法は収束的妥当性の観点からリッカート測定法よりも優位であることが示された。一方 f_0 を推定する場合はリッカート測定法と同程度まで相関係数が低下する傾向がみられた。ここでも f_0 を導入することによってモデル

TABLE 8 一対比較法・リッカート法による尺度得点の分布・相関係数・効果量

尺度番号	リッカート法					一対比較法				
	選抜的状況		非選抜的状況			選抜的状況		非選抜的状況		
	平均	SD	平均	SD	効果量	平均	SD	平均	SD	効果量
1	42.62	4.53	34.87	5.51	1.533	-0.30	1.02	0.20	1.09	-0.48
2	43.08	4.43	33.01	6.42	1.825	0.24	0.79	-0.02	0.87	0.31
3	44.36	4.00	35.91	5.96	1.663	0.25	0.75	-0.18	0.79	0.56
4	39.56	5.64	28.33	7.07	1.754	-0.02	0.38	0.00	0.42	-0.05
5	40.50	5.49	28.50	7.42	1.834	-0.03	0.39	-0.01	0.44	-0.05
6	42.46	5.38	32.12	6.76	1.688	-0.16	0.94	-0.21	1.05	0.05
7	42.68	4.74	34.93	5.71	1.473	-0.10	0.98	0.19	1.03	-0.29
8	41.85	5.20	32.93	6.46	1.518	-0.21	0.70	-0.09	0.87	-0.15

TABLE 9 各尺度における場面間の相関係数

尺度	f_0 推定	f_0 非推定	リッカート
1 現状認識力	0.29	0.40	0.39
2 ビジョン創出力	0.38	0.39	0.37
3 達成行動力	0.39	0.44	0.28
4 自己管理力	0.47	0.47	0.37
5 メンタルタフネス	0.37	0.48	0.37
6 人間関係構築力	0.42	0.52	0.44
7 コミュニケーション力	0.44	0.46	0.44
8 組織適応力	0.41	0.47	0.43

が複雑になり、因子スコアが不安定になったために場面間の相関係数が低くなったと解釈することができる。

f_0 の分布： f_0 を導入することの妥当性を検討するために、 f_0 の平均と標準偏差を自由推定し、標準化モデルの解と比較した。TABLE 10は選抜的評価場面・非選抜的評価場面においてそれぞれ自由推定された f_0 の平均と標準偏差そして効果量である。TABLE 10をみると選抜的評価場面における平均値が0.790、非選抜的評価場面における平均値が0.988と表記されている。先に標準化された一対比較型検査では f_0 の期待値は1に固定されていたから、非選抜的評価場面における f_0 の平均の推定値がほぼ1であることは、 f_0 を含め

た一対比較モデルの交差妥当性を示す有力な情報であると考えられる。効果量においても0.802と各尺度における効果量に比較して大きく、評価場面の差異に f_0 の値が影響を受ける可能性が示唆されている。

冒頭では f_0 はその値が大きいほど a_{ik} による社会的望ましさの体系を尊重し、0に近いほど社会的望ましさの体系を気にしないという個人傾向であると定義した。TABLE 10は選抜的評価場面において f_0 の平均値が減少することを示している。この点に関しては次のように説明することができる。

選抜的評価場面では被験者は常に出題者の意図を推測し、その隠れた題意に添う形で反応をすることが予想される。従って選抜的評価場面では自然な社会的望ましさの体系が機能せず、自分が入社したい企業が求めるような能力・価値観が、特殊な社会的望ましさの体系として機能すると考えられる。

内省報告：妥当性研究の一環として、選抜的評価場

TABLE 10 両場面における f_0 の分布

記述統計量	非選抜評価場面	選抜的評価場面	効果量
平均値	0.988	0.790	0.802
標準偏差	0.234	0.258	—

TABLE 11 自由記述項目の内容

順位	度数	記述内容
1	30	自分が考える社会的望ましさにあわせて回答した
2	26	同じ質問が登場すると前の回答と矛盾しないよう注意した
3	22	自分のつきたい職種・企業に求められるであろう人物像にあわせて回答した
4	19	理想化した自分を想像してそれにあわせて回答した
5	19	質問の意図を考え回答に時間がかかって疲れた
6	18	向上心があり前向きに積極的に目標に向かって物事に取り組んでいるように演出して回答した
7	18	良い・元気・ポジティブな状態の自分を思い浮かべて、強調して回答した
8	13	真ん中につけたら優柔不断なやつだと思われると感じたので中心化傾向を排した
9	13	仲間を大事にする協調的な人間に思われるよう回答した
10	12	譲れない項目・後でばれる項目には正直に答えた
11	10	真ん中につける傾向があった
12	9	社会適応性を重視した
13	6	社交性が高いように演出して回答した
14	5	私人としてではなく職業人として自分がするであろう像を思い浮かべて回答した
15	4	理想化した自己と、社会的望ましさが一致せず苦労した
16	3	苦手なものがない、プレッシャーに強いという人物像を演出して回答した
17	3	ものごとに冷静に対処できるかのごとく装って回答した
18	3	選抜的評価場面のほうが回答し易く、時間が短かった

面における被験者の内省報告を自由記述という形で測定した。TABLE 11はそれらを18のカテゴリに大別し、度数の大きいものから降順に並べたものである。大別された18のカテゴリは主に選抜的評価場面における社会的望ましさを考慮したうえで、被験者が回答を行ったことを明確に示している。例えば最も度数の多かった“自分が考える社会的望ましさにあわせて回答した”という報告は、選抜的評価場面における社会的望ましさを各被験者が独自に推測・定義して回答したことと表している。また“同じ質問が登場すると前の回答と矛盾しないように注意した”，“質問の意図を考え回答に時間がかかって疲れた”といった比較的度数の多い報告からも、選抜的評価場面という特殊状況下では自然な社会的望ましさよりもむしろ、出題に対する適切な回答としての望ましさの体系に従って回答が行われていることが示されている。このような状況下では、実質的に自然な社会的望ましさが軽んじられており、従って選抜場面における f_0 の平均値が1からの乖離が生じたと考えることができる。

妥当性研究の総合考察

選抜的評価場面・非選抜的評価場面におけるリックート測定法、一対比較法のそれぞれの尺度得点の平均値・標準偏差・効果量の比較の結果、評価場面の変化に対して一対比較法における尺度得点は効果量、標準偏差の両方で、リックート法に比較して不变性を持つことが明らかになった。つまり評価場面における反応歪曲に対して頑健であることが明らかになった。

また選抜的評価場面と非選抜的評価場面の統計量における大きな差異により新検査の妥当性を評価することは、リックート測定法による検査と、一対比較法による検査において全く同一の項目を用いているという事実からも支持され、従って反応歪曲に抗する一対比較モデルとしての妥当性を本検査は保持している可能性が示された。

次に場面間での尺度間相関を求めたところ、 f_0 を推定しないモデルにおいて全ての尺度でリックート測定法の尺度間相関係数を上回っていた。従って f_0 を推定しないモデルの下では、反応歪曲に対して頑健であり、かつ高い収束的妥当性を保持する一対比較型検査を実施することが可能である。

また非選抜場面において、自由推定された f_0 の期待値が極めて1に近かったことで、モデルの交差妥当性が示唆された。ただモデルが複雑になるため因子スコアの推定値が不安定になる傾向が認められ、現時点では f_0 を含めたモデルを積極的に採用することは難し

い。

しかし選抜的評価場面における f_0 の期待値が1よりも低かったことで、 f_0 を利用する可能性が開けたといえ、モデルの安定性を含めて今後詳細に検討する余地があるといえる。

引用文献

- エドワース 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.)
- 狩野 裕・三浦麻子 2002 グラフィカル多変量解析
現代数学社
- 南風原朝和 2002 心理統計学の基礎 有斐閣アルマ
Pp.163-164.
- 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
- Thorndike, R. M. 1997 *Measurement and evalua-
tion in psychology and education* (6th ed.).
Englewood Cliffs, NJ : Prentice-Hall.
- 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・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.)
- 豊田秀樹・真柳麻薺美 2001 繰り返し測定を伴う実
験のための因子分析モデル—アイスクリームに關
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.)

(2005.10.17 受稿, '06.11.18 受理)

A Paired Comparison Test Model Using Structural Equation Modeling Resistant to Response Bias

HIDEKI TOYODA (WASEDA UNIVERSITY), IKKO KAWAHASHI (GRADUATE SCHOOL OF LETTERS, ARTS AND SCIENCE, WASEDA UNIVERSITY),

TOHRU WATANABE AND MIEKO HIDAKA (EI RESEARCH) JAPANESE JOURNAL OF EDUCATIONAL PSYCHOLOGY, 2007, 55, 325-334

As an extension of the model of Toyoda, Kawahashi, and Matsushita (2005), the present authors added to the model the parameter f_0 , which represents individual differences in social desirability. In addition, we removed the assumption of tau-equivalent measurement for items. A paired-comparison test was constructed using competency scales developed in a pretest. The test was split into half, and the two halves standardized. The data from 400 participants indicated that split-half validity and reliability was confirmed. In a validity study, in order to examine the robustness of the standardized test to response biases, 199 participants took the paired-comparison test and a Likert test. Participants were instructed to assume that they were in a 2-selection evaluation scene. The results indicated that the new method is more robust to response biases than the Likert method, irrespective of differences in the evaluation scene.

Key Words : paired comparison competency test, response biases, social desirability, structural equation modeling