

リッカートとイプサティブ回答形式を用いた性格特性尺度の検討

— 測定内容の違いに着目して —

今城志保¹ 酒匂志野¹ 坂本佑太郎¹

¹株式会社リクルートマネジメントソリューションズ

1. 背景

性格適性検査は、性格によって職務への適性を推測する検査である。入社試験、昇進昇格試験などの場面で活用されることが多いため、自分をよく見せようとする回答傾向(フェイキング)の影響が問題視される。

フェイキングを統制するための方法として、一対比較法やランキング法などのイプサティブ形式が知られている(Thurstone, 1927)。伝統的なイプサティブデータの分析法では、個人間の比較ができないという課題があったが、近年、多次元IRTの技術を応用することにより、イプサティブデータから個人間比較を可能にする方法が提案されている(Maydeu-Olivares & Brown, 2010)。また、川端(2007)は、個人間の比較が可能となるようなSEMのモデルを提案しており、「企業の採用試験」という場面の想起を求める条件とそうではない条件で性格検査への回答を求めたところ、リッカート形式での得点は前者において高まるがイプサティブ形式での得点は両者においてさほど大きな違いが見られないことを示している。

上記の研究では、「リッカート形式とイプサティブ形式では同一の概念が測定されている」と考えられている。しかし、たとえばJoubert, et al. (2015)では、両者の相関係数が.50~.84、Brown & Maydeu-Olivares (2013)では.52~.73、酒匂ら(2016)では.44~.71と報告されており、研究や尺度によって両者の相関には違いがある。酒匂ら(2016)では、両者の違いについてさらに検討を行った結果、リッカート形式のほうがイプサティブ型式よりも高い得点を示す回答者の虚偽尺度得点は、逆のパターンを示す回答者のそれよりも有意に高かった。しかし、両者の得点に違いの無い人でも、中程度の虚偽得点があるなど、虚偽尺度のみで両者の得点の違いが十分に説明できたわけではない。

そこで本研究では、二つの測定形式に対するフェイキングの影響を異なる方法を用いて分析し、測定結果の違いが何によって生じるかについて探索的な検討を進めることを目的とする。

性格適性検査とフェイキング

先行研究では、性格適性検査は職務でのパフ

フォーマンスを基準としたときに、ある程度の妥当性が示されている(Barrick, Mount, & Judge, 2001; Hurtz & Donovan, 2000; Salgado, 1997; 都澤ら, 2005)。しかし、前述のようにフェイキングが性格適性検査活用時の重要な懸念点との指摘がある(Hogen, Hogen, Roberts, 1996)。

フェイキングは、仮にあったとしても妥当性には実務上問題となるような影響を及ぼさない(Barrick & Mount, 1996; Ones, Viswesvaran, & Reiss, 1996)との議論もあれば、応募者而非応募者で得点が異なることを示す研究もある(Birkeland, Kisamore, Brannick, & Smith, 2006)。多くのフェイキング研究では、フェイキングするようにインストラクションを受けた群とそうでない群での比較検討を行うが、実際の、例えば採用場面で生じるフェイキングとは、異なる(Griffith & Peterson, 2011)。例えば Sackett(2011)は、性格検査には以下6つのシステマティックな分散が含まれるとしている;

1. Mean true score across situations
2. Situationally specific true departure from grand mean
3. Cross-domain erroneous self-perception
4. Situationally specific erroneous self-perception
5. Cross-domain consistent impression management
6. Situationally specific intentional distortion.

フェイキングのインストラクションは上記の6を操作していると考えられるし、応募者而非応募者の比較では、上記の5と6の両方の影響を見ている。しかし印象操作をしようという意図がない場合であっても、3や4に該当するような自己認知のゆがみが影響することも考えられる。自己認知を正確に測定するという意味では、3や4は測定の妥当性を損なうものではないが、適性検査の目的であるパフォーマンスの予測を考慮した場合は、3、4も問題になる可能性がある。

そこで本研究では、意図的な印象操作の影響にとどまらず、自己高揚的な傾向が、二つの測定形式にどのような影響を与えるかについても、検討を行う。

2. 方法

日本のビジネスパーソンを対象にインターネット

調査を実施した。調査は 2016 年に実施し、調査対象者は日本企業に勤めるホワイトカラー800名であった(男性のみ/年齢:30~49歳/勤務先企業の従業員規模:500名以上)。調査の冒頭で簡単な英語の問題に解答を求め、その結果、英語の理解度に問題がないと思われる644名のデータを分析に用いた。なお、本調査で実施した尺度は、以下の通りである。

イプサティブ形式の5尺度

性格を問うイプサティブ形式の5尺度(外向、統率、変革、強靱、大胆)について、4つの項目を1セットとして提示する。その中から最も当てはまる項目(most like me)と最も当てはまらない項目(least like me)を1つずつ選択する(27セット)。IRTモデルによる推定には Mplus(ver7.2)を用いた。分析時には2項目ずつ比較した場合への反応(2値データ)に変換した(例:外向が most、統率が least の場合、外向1、統率0)1セットにつき6ペア作られる。なお、各セットにつき mostにも leastにも選ばれなかったものどうしでは値が定まらない。よって、IRTモデルにより推定を行う際には、多重代入法により60個のデータセットを生成し、それぞれにおいて項目パラメータと個人特性値の推定を行い、それらの平均をとった。

なお、あてはめたモデルは以下の通りである。すなわち、特性 a (例:外向性)を測定する項目 i と特性 b (例:統率性)を測定する項目 k とを比べる比較対 l に対する反応を y_l とする。このとき、比較対 l への反応は、2項目の効用の差 y_l^* によって、

$$y_l = \begin{cases} 1, & y_l^* \geq 0 \text{ のとき} \\ 0, & y_l^* < 0 \text{ のとき} \end{cases}$$

と決まると考える。さらに、効用の差 y_l^* に対して以下のような因子分析モデルを考える。

$$y_l^* = -\gamma_l + (\lambda_{ia}\eta_a - \lambda_{kb}\eta_b) + (\epsilon_i - \epsilon_k) \quad (1)$$

ここで、 η_a, η_b は潜在特性 a, b の値を表しており、 $\lambda_{ia}, \lambda_{kb}$ はそれぞれの項目への各特性の因子負荷量を、 ϵ_i, ϵ_k は項目 i と k の独自性を、 γ_l は比較対に固有の切片を、それぞれ表している。

リッカート形式の5尺度

性格を問うリッカート形式の5尺度(外向、統率、変革、強靱、大胆)について、4件法(Strongly agree - Strongly disagree)での評定を求める。IRTモデルの母数の推定には段階反応モデルを用い、尺度ごとに推定を行った。推定には R(for

windows 3.0)を利用した。

虚偽傾向

7件法(Strongly agree - Strongly disagree)のリッカート形式で6項目に対して回答を求めた。Balanced Inventory of Desirable Responding (BIDR-J)を参考に作成したもので、例えば「I am fully in control of my own fate.」のような項目がある。

自己高揚傾向

リッカート形式で自分のパフォーマンスへの満足度を問うもので、7件法(Highly satisfied - Not satisfied at all)で5項目について回答を求めた。項目例として「to proactively build a relationship with many people」がある。

ただしこの尺度は、前述の Sackett(2011)の3, 4に該当するとは限らない。つまり実態と比べて自己評価が高ぶれするエラーであるとは言えず、あくまで高く自己評価を行う傾向であることを注意が必要である。

3. 結果

分析1;相関係数

イプサティブ形式の5尺度(IRT推定値)と、リッカート形式の5尺度(IRT推定値)で相関を求めた結果を Table1 に示す。対応する尺度間の相関は.544~.640と対応しない尺度との相関と比べると高く、弁別的妥当性が確認された。ただし、相関は前述した先行研究と比べるとやや低い水準にとどまった。また、リッカート尺度間の相関(.409~.626)に比べると、イプサティブ尺度間の相関(.365~.814)のほうが高かった。逆に Brown & Maydeu-Olivares(2013)では、リッカートによる性格特性推定値間の相関の平均値は .22 に対して、イプサティブの多次元 IRT による推定値間の相関は .12 とやや低めの値を報告している。

分析2;虚偽傾向と自己高揚傾向の影響

リッカートとイプサティブのそれぞれ5尺度に対して、虚偽傾向と自己高揚傾向がどのように影響するかを SEM を用いて検討した結果が、Figure1 である。

モデルの設定に際して、リッカートとイプサティブの回答傾向から各尺度へのパスは全て1に、また各因子から2種類の測定方法のIRT推定値へのパスも全て1に設定した。これらのパスについては、尺度ごとの違いは見られないという前提を置いた。モデルのあてはまりは、 $\chi^2 = .78 (df = 176)$,

CFI= .91, RMSEA= .07であった。

虚偽傾向からのパスは、リッカートの回答傾向に対してのみ有意となった。自己高揚傾向からのパスは、リッカートの回答傾向、イプサティブの回答傾向にともに有意な値を示した。

4. 考察と追加分析

リッカート形式の 5 尺度とイプサティブ形式の 5 尺度の相関の様子からは、両尺度が似た構成概念を測定しているものの、相関係数は .553 ~ .640 であることから、同一のものを測定しているとは言い難い。酒匂ら(2016)では、リッカート形式での得点とイプサティブ形式での得点の差に着

目して、2 つの測定形式間の違いが生じる原因について探索的な検討を行った。その結果、虚偽尺度の高低が影響している可能性が示されたが、得点差の無い人でも虚偽尺度得点がある程度高いことから、虚偽傾向による説明では不十分との結論であった。

今回 SEM で分析を行ったが、まず、虚偽傾向と自己高揚傾向には有意な相関があり(.43)、また自己高揚傾向が高い人は、リッカート形式でもイプサティブ形式でも得点が高くなる傾向が見られた。酒匂ら(2016)の際に、リッカート形式の 5 尺度とイプサティブ形式の 5 尺度に違いが無かった人に見られた虚偽尺度の得点は、自己高揚傾向によるも

Table1 尺度間相関 (N=644)

	リ_外向	リ_統率	リ_変革	リ_強靱	リ_大胆	イ_外向	イ_統率	イ_変革	イ_強靱	イ_大胆
リッカート_外向	1									
リッカート_統率	.553**	1								
リッカート_変革	.417**	.626**	1							
リッカート_強靱	.449**	.451**	.443**	1						
リッカート_大胆	.409**	.594**	.502**	.524**	1					
イプサティブ_外向	.640**	.377**	.252**	.282**	.230**	1				
イプサティブ_統率	.403**	.597**	.428**	.267**	.356**	.663**	1			
イプサティブ_変革	.317**	.475**	.561**	.286**	.369**	.523**	.795**	1		
イプサティブ_強靱	.289**	.280**	.307**	.544**	.415**	.471**	.455**	.566**	1	
イプサティブ_大胆	.225**	.391**	.362**	.423**	.553**	.364**	.573**	.685**	.815**	1

** p < .01

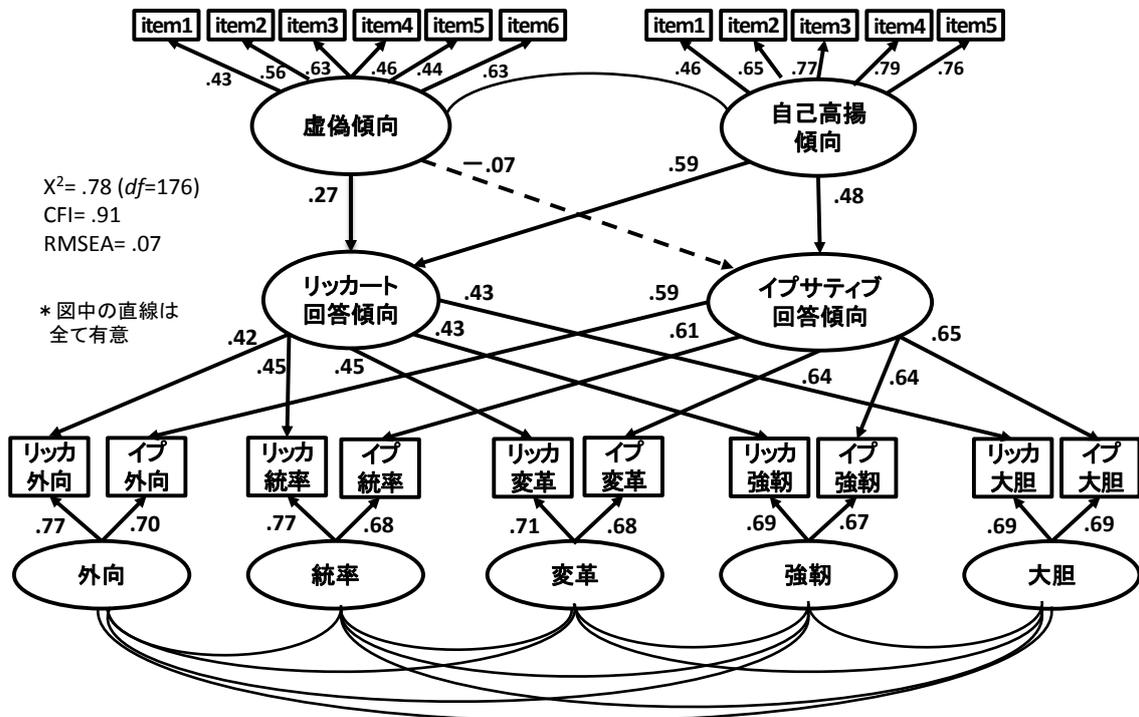


Figure1 2つの測定形式に対する虚偽傾向と自己高揚傾向の影響

のと考えられる。そして本分析でも虚偽傾向の高い人はリッカート形式のみで得点が上昇することが示された。虚偽傾向の違いによって、2つの測定形式間の尺度得点はずれを生じることになるため、尺度間相関が低くなったと考えられる。

また今回、イプサティブ形式での尺度間の相関が、リッカート尺度間の相関よりも高かったが、このような現象は先行研究では見られない。SEMの結果を見ても、イプサティブ回答傾向から各尺度へのパスは、リッカート回答傾向のそれと比べると若干高くなっている。今回SEMのモデル作成において「イプサティブ回答傾向」が、イプサティブ形式での尺度間相関の高さを表現した結果になったが、この「イプサティブ回答傾向」とは何だろうか。

イプサティブ回答形式での回答傾向を扱った研究はこれまで存在しない。そこで、高い負荷量((1)式の $\lambda_{ia}, \lambda_{kb}$)を示す項目をmost like meとして選択する傾向が高いことが、尺度間相関を高めている可能性を考え、項目の負荷量((1)式の $\lambda_{ia}, \lambda_{kb}$)とmost like meの選択率(反転項目の場合least like meの選択率)の間の相関をとったが、結果は無相関であった($n = 60, r = -.07, p = .61$)。

イプサティブ形式での尺度得点の合計を用いて、回答者を高群、中群、低群に3等分し、これを結果変数に、各項目の選択結果を説明変数として、一般化線形モデルを用いて分析を行った。説明変数の投入は尺度ごと、12項目ずつで、該当項目がmostにもleastにも選択されない場合を基準にして、most like me, least like meそれぞれの選択の効果を見た。この結果からも、結果に対する予測力のある項目のラムダは、必ずしも高くなかった。例えば外向性の「I am able to make friends with others quickly.」の λ は1.60と大きかったが、most, leastのいずれの選択も有意に得点の全体的な高さを予測しなかった(それぞれ $\beta = .16, .19$)。一方「I find it easy to keep conversations going.」の λ は0.80であったが、mostの選択は有意であった($\beta = .75$)。

SEMモデルで「イプサティブ回答傾向」と置いたものが何であったのかは、明らかにならなかった。Van der Linden, te Nijenhuis, & Bakker (2010)は、パーソナリティにも能力のような一般因子が存在し、これがパフォーマンスの予測に大きく寄与していることを示している。イプサティブの尺度間相関が単なる回答傾向か、あるいは高次の心理的な構成概念を測定しているのか、その構成概

念はパフォーマンスを有意に予測するのかなど、今後の検討課題としたい。

5. 文献

- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of applied psychology, 81*(3), 261.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and assessment, 9*(1-2), 9-30.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317-335.
- Brown, A. & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18* (1), 36-52.
- Griffith, R. L., & Peterson, M. H. (2011). One piece at a time: The puzzle of applicant faking and a call for theory. *Human Performance, 24*(4), 291-301.
- Joubert, T. et al. (2015) A comparison of the psychometric properties of the forced choice and likert scale versions of a personality instrument. *International Journal of Selection and Assessment, 23*(1), 92-97.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American psychologist, 51*(5), 469.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of applied psychology, 85*(6), 869-879.
- 川端一光 (2008) 構造方程式モデリングによる一対比較型尺度構成法の開発. 早稲田大学博士論文.
- Maydeu-Olivares, A., & Brown, A. (2010). Item Response Modeling of Paired Comparison and Ranking Data. *Multivariate Behavioral Research, 45*, 935-974.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660.
- Sackett, P. R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance, 24*(4), 379-385.
- 酒匂志野, 今城志保, 北原瑞穂, 荘島宏二郎 (2016). イプサティブ回答形式を用いた性格特性尺度分析に関する考察. テスト学会年次大会
- Salgado, J. F. (1997). The Five Factor Model of Personality and Job Performance in the European Community. *Journal of Applied Psychology, 82*(1), 30-43.
- 都澤真智子, 二村英幸, 今城志保, & 内藤淳. (2005). 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. *経営行動科学, 18*(1), 21-30.
- Thurstone, L. L. (1927) A law of comparative judgment. *Psychological Review, 34*(4), 273-286.
- Van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of research in personality, 44*(3), 315-327.