

採用場面における適性検査のフェイキング回答に関する研究

-検出手法の有効性及び質問形式による耐性の検討-

○阿部優・坂本亜沙美・内藤淳

㈱リクルート HR・斡旋カンパニー 新卒事業本部
インフローソリューションディビジョン
ソリューション開発部 測定技術研究所

1.問題意識と目的

我が国では、企業の採用選考場面において適性検査が実施されることが一般的であり、職務や組織風土に対する適応性を把握することを目的に、一般知的能力と合わせて性格検査が広く用いられている。また、米国においても1990年代前半に性格検査が入社後のパフォーマンスの予測に関して高い妥当性を有することがメタ分析によって示された（e.g., Barrick&Mount,1991）ことを契機に利用場面が拡大しつつあり、現在では採用、臨床、教育などの場面で幅広く利用されている。

活用場面の拡大に伴って、性格検査に対するフェイキング（受検者が意図的に回答を歪める行為）の問題が注目を集めるようになってきている。米国ではフェイキングによる妥当性への影響やフェイキングの検出、抑制の方法に関する様々な研究が行われているが（e.g.,Birkeland,Manson,Kisamore, Brannick,&Smith,2006; Ones,Viswesvaran&Reiss,1966; Li&Bagger,2006; Viswesvaran&Ones, 1999）、日本においては性格検査のフェイキングに関する研究はまだ少ないのが現状である。

性格検査のフェイキングの検出に最もよく用いられているのは「ライスケール」とも呼ばれる社会的望ましき尺度（Social desirability scale）であるが、開発者が意図した通りの反応を受検者はしないため、実際の場面では有効に機能しないという問題が指摘されている（Birkeland,Kuncel, Borneman,&Kiger,2012）。ライスケールで通常用いられるような極端な内容を表す項目では、受検者にそれと意識されてしまう懸念があるといえる。

他方、先行研究において、性格検査のフェイキング耐性を高める上で有力とされている方法として項目の中性化（Evaluation neutralization : Bäckström,Björklund,&Larsson,2009）と強制選択方式の出題（Forced-choice Item format）がある。項目の中性化とは、質問項目の表現からできるだけ社会的望ましさを減らすことによりフェイキングを抑えようとするものである。一方、強制選択方式の出題とは、社会的な望ましさがほぼ同等の複数の項目を提示し、どちらが自分に近いかを受検者に選ばせる手法である。ただし強制選択方式での出題では、通常、異なる尺度に帰属する項目を対置した上で一方の選択を受検者に求めるため、測定結果がイプサティブになり、採用選考場面等で求められるような受検者間の特性の比較には使用できないという問題がある（Guilford,1954; Saville&Willson,1991）。

本研究では研究1として実験的な環境でのデータに基づき、回答者に違和感を与えない項目から成る新たな検出尺度の作成を試みる。また研究2として、提示する項目の社会的望ましきのバランスに配慮して作成された対形式のノーマティブな質問形式が一般的なリッカート形式と比較してより高いフェイキングへの耐性を有するかどうかをBig5のフレームによる性格検査を用いて検証する。

2.方法

2.1.研究1.フェイキング回答検出尺度の作成および一般的なライスケールとの比較

新しいフェイキング尺度を構成するための実験Aと、その新尺度の有効性を検証するための実験Bを行った。

実験A 会場に集めた大学生から社会人までの211名の被験者(男性121名,女性90名;平均年齢24.4歳;大学生89名,大学院生21名,社会人101名)に対し、1回はフェイキング回答を促すインストラ

クシオン（表1）、もう1回は正直に回答するように促すインストラクションを行い、111項目の性格検査に249項目の日ごろの行動や考え方について問う、回答が極端に偏らないような項目を加えた同一の質問紙に2回回答させた。なお、被験者を半数ずつの2群に分け、どちらのインストラクションを先に行うかの順序を入れ替えて、順序効果の影響を排除している。

実験Aで収集したフェイキング回答時と非フェイキング回答時のデータを比較して回答傾向の差が大きい項目を14項目ピックアップし、フェイキング回答を検出するための新しい尺度を構成した。

表1 フェイキング回答を促すインストラクション

自分がとても就職したいと思っている会社の就職試験を受けているとします。
採用担当者が自分に対して良い印象をもち、採用されるチャンスが最大化されるように回答してください。自分をよく見せるように回答してください。
必ずしも自分のことを正直に答える必要はありません。

実験B 会場に集めた大学性から社会人までの144名の被験者（男性64名、女性76名；平均年齢22.2歳；大学生129名、大学院生9名、社会人2名）に対し、実験Aと同様にフェイキング回答と正直回答を促す2回の異なるインストラクションを用いて質問紙に回答させた。

質問紙は、実験Aで構成した検出尺度（14項目）と、「嘘をついたことは一度もない」など、回答が大きく偏る極端な内容を表す項目で構成された一般的なライスケール（14項目）と、394項目の性格検査から成る。

2.2. 研究2. 質問形式によるフェイキング回答の影響の違いの検証

実験Aの被験者211名に対してリッカート形式と対形式による性格検査を実施し、フェイキング回答時-非フェイキング回答時の得点上昇の傾向（変化量）に違いがあるかを検証した。

利用した性格検査はBig5に対応した外向性・勤勉性・親和性・情緒安定性・開放性の5つを測定する111項目から成る尺度を今回新たに構成した。リッカート形式（特定の記述に対して「あてはまるかどうか」を問うもの）と、対形式（「Aの記述とBの記述のどちらに近いか」を問うもの）の双方で構成されている。各尺度に対応するリッカート形式の尺度と対形式の尺度は、それぞれ、平均・SD・信頼性係数（ α 係数）ができるだけ同等になるように構成した（表2）。

表2 利用した性格検査（N=15492）

形式	外向性		勤勉性		親和性		情緒安定性		開放性	
	リッカート	対形式	リッカート	対形式	リッカート	対形式	リッカート	対形式	リッカート	対形式
項目数	11	11	16	18	13	11	8	10	6	8
平均	3.27	2.99	2.83	2.65	2.97	2.78	2.26	2.39	3.22	2.70
SD	.46	.44	.51	.60	.39	.50	.59	.50	.58	.60
α 係数	.84	.80	.90	.91	.75	.82	.81	.81	.84	.83
形式間相関	.76		.70		.65		.74		.68	

3. 結果

3.1. 研究1.

新たなフェイキング検出尺度と一般的なライスケールのそれぞれについて、正答項目数ごとのエラー率を算出した（図1, 図2）。新フェイキング検出尺度は11項目以上、一般的なライスケールは3項目以上正答した受検者をフェイキング回答として検出した際に、エラー率が最小となった。

エラー率とは、I. 見逃しのエラー数（フェイキング回答をしたにも関わらず検出されなかったデータ数）とII. 誤検出のエラー数（非フェイキング回答であるにも関わらず誤って検出されたデータ数）

の和を全体データ数で割った値である。

実験Bのデータを用いて新たなフェイキング検出尺度と一般的なライスケールとの有効性比較を行ったところ、新フェイキング検出尺度のほうがエラー率が低いことが示された（表3）。

$$\text{エラー率} = \frac{(\text{フェイキング回答だが検出されなかった数}) + (\text{非フェイキング回答だが検出された数})}{\text{全回答数}}$$

表3 新フェイキング検出尺度と一般的なライスケールの最大検出時のエラー率の比較

	I.見逃しのエラー率	II.誤検出のエラー率	エラー率（平均）
新フェイキング検出尺度	15.0%	10.8%	12.9%
一般的なライスケール	27.2%	12.8%	20.0%

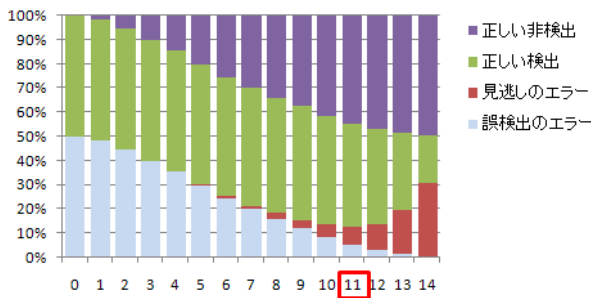


図1 新フェイキング検出尺度の検出状況

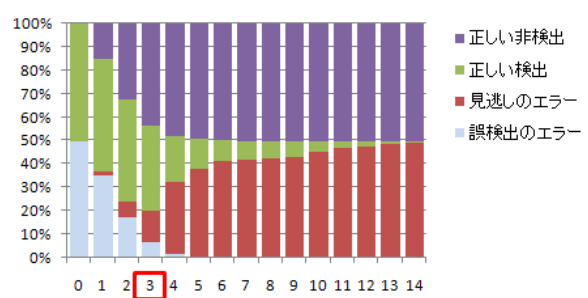


図2 一般的なライスケールの検出状況

3.2. 研究2.

Big5の5尺度について、リッカート形式、対形式ごとのフェイキング回答時と非フェイキング回答時の差を確認したところ、勤勉性、親和性、開放性の3尺度で対形式のほうがフェイキング回答の影響を受けにくいことが確認された（表4）。

表4 フェイキング回答時と非フェイキング回答時の変化量 d と形式ごとの変化量の差分

形式	外向性		勤勉性		親和性		情緒安定性		開放性	
	リッカート	対形式	リッカート	対形式	リッカート	対形式	リッカート	対形式	リッカート	対形式
変化量(d)	.81	.92	.45	.02	.40	.08	.82	.72	.51	.29
d の差分		-.12		.43**		.32**		.10		.22*

** $p < .01$, * $p < .05$

注1) 変化量 d は、尺度の標準偏差で除することで基準化された値で記述

4. 考察

4.1. 研究1.

一般的なライスケールよりも新フェイキング検出尺度のほうがエラー率が低く検出の精度が高い理由として、受検者がフェイキング回答を行うとき、一般的なライスケールに該当しないように回答するという行為がなされていることが推察される。

採用選考で用いる適性検査において公平性は重要な要素であるが、実際の場面における検出精度を上げるために、受検者にとってそれとわかりにくい自然な項目を用いた検出尺度を作成することの意味は大きい。

4.2. 研究2.

勤勉性、親和性、開放性で対形式のほうがフェイキング回答の影響を受けにくかった理由としては、

項目の中性化が機能しているためだと考えられる。例えば親和性の場合、対概念として「競争心」「自律性」など幅広い概念を置くことができる。仮に親和性と競争心とが同程度の社会的望ましさをもつとすれば、受検者のフェイキングを抑えることができる。

一方、外向性と情緒安定性の2尺度については形式による有意な差は認められなかった。これは、外向性、情緒安定性の場合には、同等に社会的望ましいような対概念を置くことが難しいためではないかと考えられる（外向性の対概念は内向性、情緒安定性の対概念は情緒不安定となり、ネガティブになりやすい）。測定内容によって、対形式による中性化が行いやすいものと行いにくいものがある可能性がある。

先行研究にも示されているように（Bäckström et al.,2009）、項目の中性化を行うと一般的に項目の示す内容がわかりにくくなり、信頼性が下がる傾向がある。本研究においても今回 Big 5 の尺度を作成するにあたり、対形式のほうが同レベルの信頼性を確保するためにより多くの項目数を必要とする傾向があった。

適性検査の品質を語る上で、信頼性・妥当性に加えフェイキング耐性がこれからは大きな要素となると考えられるが、そのためには従来よりも多数の質問項目が性格検査には求められるだろう。

5. 今後に向けた課題

本研究は、一般的なライスケールが実際の選考場面では十分には機能していない可能性があることを示し、新しい検出尺度の有効性を示唆するものである。また、質問形式の違いについては、項目を中性化できる場合には、対形式のほうがフェイキングを抑えやすいことが示された。

しかしながら、本研究はあくまで実験に基づくものであり、現実の場面での有効性を示しているとは断定できない。今後は実際の採用選考場面で研究を重ね、今回得られた知見の一般性を検証していくことが求められる。

参考文献

- Bäckström,M., Björklund,F., & Larsson,M.R. (2009). Five-factor inventories have a major higher order factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43, 335-344.
- Birkeland,S.A. ,Manson,T.M., Kisamore,J.L., Brannick,M.T., & Sumith,M.A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317-335.
- Barrick,M.R.,& Mount,M.K.(1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Dilchert,S., Ones,D.S., Viswesvaran,C., & Deller,J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, 48, 209-225.
- Guilford,J.P.(1954). *Psychometric methods*(2nd ed.). New York: McGraw-Hill.
- Li,A., & Bagger,J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131-141.
- Nathan,R.K., Matthew.B., & Thomas.K.(2011). Innovative Item Response Process and Bayesian Faking Detection Method: More Questions Than Answers . Matthias,Z., Carolyn, M., & Richard,D. (Eds). *New Perspectives on Faking in Personality Assessment*. Oxford University Press, pp102-112.
- 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, 660-679.
- Saville,P., & Willson,E. (1991).The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64,219-238.
- Viswesvaran,C., & Ones,D.S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210.
- Ziegler,M., & Buehner,M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, 69, 548-565.