

Do We Really Need Lie Scales to Detect Faking on Self-Report Measures?

Walter P Vispoel*, Murat Kilinc, and Wei S Schneider

University of Iowa, USA

*Corresponding author: Walter P Vispoel, University of Iowa, USA

ARTICLE INFO

Received: 📅 May 15, 2024

Published: 📅 June 04, 2024

Citation: Walter P Vispoel, Murat Kilinc, and Wei S Schneider. Do We Really Need Lie Scales to Detect Faking on Self-Report Measures?. Biomed J Sci & Tech Res 56(5)-2024. BJSTR. MS.ID.008917.

ABSTRACT

Possible faking on objectively scored self-report measures has been a serious concern to researchers and practitioners since the inception of such measures. Accordingly, various methods to discourage and detect faking have been developed over the years, with embedded validity or so called "lie" scales remaining a popular method to serve those purposes. In this brief article, we describe a rediscovered "total score-based" method for detecting faking that obviates the need for lie scales when using multidimensional self-report inventories and report results from an empirical study in which this rediscovered method was better at detecting both faked good and faked bad responses to a widely administered personality inventory than were validity scale scores from the Balanced Inventory of Desirable Responding.

Abbreviations: SDR: Socially Desirable Responding; IM: Impression Management; SDE: Self-Deceptive Enhancement; BIDR: Balanced Inventory of Desirable Responding; PDS: Paulhus Deception Scales

Overview

Objectively scored non-cognitive instruments such as Likert-Style self-report questionnaires are used routinely for information gathering, diagnosis, placement, theory building, prediction, classification, and selection within numerous disciplines including biomedicine. However, one of the most serious drawbacks to such measures is their susceptibility to response biases that can undermine valid interpretation of results. Pervasive among such biases is Socially Desirable Responding (SDR), which reflects tendencies to endorse (fake good) or deny (fake bad) socially acceptable behaviors when something can be gained by doing so. A common way to address such problems is to administer validity or so called "lie" scales intended to detect such response tendencies along with the targeted measures of interest. Examples of such validity scales include the K and L scales from the Minnesota Multiphasic Personality Inventory [1,2], Marlowe-Crowne Social Desirability Scale [3], Edwards Social Desirability Scale [4], Eysenck Lie Scale [5], Martin-Larsen Approval-Motivation Scale [6], Jacobson-Kellogg Social Desirability Inventory [7], Self- and Other-Deception Questionnaires [8], Balanced Inventory of Desirable Responding (BIDR; [9]), and Paulhus Deception Scales (PDS; [10]), among others

The Balanced Inventory of Desirable Responding (BIDR)

The BIDR and PDS are widely used companion measures that can be coupled with any self-report questionnaire to detect possible faking. They are essentially the same instruments, sharing 38 of 40 items in common and the same theoretical underpinning. Both instruments also are more comprehensive than most of the ones previously cited because they measure two distinct components of socially desirable responding: Impression Management (IM) and Self-Deceptive Enhancement (SDE; see, e.g., [11-13]). The SDE subscale consists of 20 items intended to measure honest but inflated self-presentation. High scores on SDE demonstrate exaggeration of skills and lack of self-awareness. The IM scale consists of 20 items that reflect uncommon but socially desirable behaviors. Higher scores on IM may reflect intentional attempts to present a socially approved but inaccurate image to others. Similarly, lower scores on IM and SDE can reflect a socially disapproved and oppositely distorted image to others. Each subscale consists of equal numbers of positively and negatively phrased items. In common applications, the BIDR includes a 7-point response metric, whereas the PDS includes a 5-point response metric,

although both inventories could include either metric. Within each instrument, the lowest scale point is labeled as “not true” and highest as “very true,” with negatively keyed items reverse scored. We report results for the BIDR here due to its widespread use, effectiveness in detecting faking (see, e.g., [14-16]), and availability from its author Delroy Paulhus at no cost (<https://www2.psych.ubc.ca/~dpaulhus/>).

Although both the BIDR and PDS have been effectively used to detect faked responses to questionnaires, they each are 40 items in length and thus are frequently impractical to administer. One possible way to avoid using validity scales altogether is to use responses to the targeted measures themselves to detect faking directly. Over the years, many such techniques, varying in complexity, have been proposed (see e.g., [17-19]). Unfortunately, in most cases, these techniques have not fared any better than external validity scales in detecting faking.

A Simple Way to Detect Faking Without Lie Scales

One extremely simple technique that our research team recently considered for detecting faking was to use total scores for instruments that have multiple subscales measuring weakly correlated constructs. For most self-report measures, respondents can recognize responses that are socially desirable or undesirable and fake accordingly (see [19] for a comprehensive review of studies into faking responses on personality inventories). However, with multidimensional inventories with weakly correlated subscale scores, a pattern of highly desirable or undesirable responses across all subscales would be very unusual when responding honestly, and therefore provide an alternative and potentially more effective way to detect faking. After coming up with this idea, we checked back over the research literature to determine whether others had used this technique in the past. We uncovered only one such study in which Comrey and Backer [20] found that the total score from the Comrey Personality Scales was more effective in detecting faking than were validity scale scores and other indices derived from item scores within the same instrument.

Purpose of our Recent Study

Given that the Comrey and Backer [20] study is nearly 50 years old, and that their approach has seemingly not been used much thereafter, we decided to put it to the test in a new study that we recently described at the annual meeting of the American Psychological Association [21]. In the remainder of this brief article, we will share results from that study in which we compared the effectiveness of the validity scales from the BIDR and total scores from the Big Five Inventory [22] in detecting instances of faking good and faking bad.

Methods

Participants and Measures

We assigned 448 college students at random to two research con-

ditions: (1) fake good ($n = 224$) and (2) fake bad ($n = 224$). In each condition, respondents completed web-based versions of the Big Five Inventory (BFI; [22], also see [19,23-26]) followed by the Balanced Inventory of Desirable Responding (BIDR; [9-10], also see [15,16,18,24,27-31]). The BFI has 44 items with five 8- to 10-item subscales that measure five superordinate dimensions of personality: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness, answered using a 5-point Likert-style response metric (1 = disagree strongly, 2 = disagree a little, 3 = neither agree nor disagree, 4 = agree a little, and 5 = agree strongly). As noted earlier, the BIDR has two 20-item subscales to measure Impression Management (IM) and Self-Deceptive Enhancement (SDE) using a 7-point response metric (1 = not true, 4 = somewhat true, 7 = very true).

Procedure

Participants in each research condition answered the measures honestly first, and then again to convey either the best (fake good) or worst (fake bad) impressions of themselves. To use all collected data to full advantage (that is, include 448 cases in each condition), we combined fake-good scores with honest scores for the fake-bad condition and combined fake-bad scores with honest scores from the fake-good condition. When computing total scores across subscales for the BFI, responses to the Neuroticism subscale were reverse scored so that higher scores would represent more socially desirable responses. Classification accuracy (correctly labeling honest and faked responses), false-positive-error rate (labeling honest as faked responses), and false-negative-error rate (labeling faked as honest responses) were calculated for all scoring points for each scale within the BFI and BIDR under fake-good and fake-bad conditions. The score that maximized overall classification accuracy was selected as the cut score for each scale, and overall classification accuracy was compared across subscales and instruments. Using large and equal size groups for honest and faked responding was intended to provide very strict and conservative tests of classification accuracy because it would rarely be the case in practice that half of the respondents would willfully fake responses.

Results

Cut scores and indices of classification accuracy are provided in Tables 1 & 2 for faking good and faking bad, respectively. Our analyses for individual subscales revealed that Neuroticism, when reversed scored, was best for detecting faking good (90.38% classification accuracy), and that Agreeableness was best for detecting faking bad (95.19% classification accuracy). At the overall instrument level, total BFI scores were better than total BIDR scores for detecting both faking good (93.75% versus 89.90% classification accuracy) and faking bad (95.67% versus 93.00% classification accuracy).

Table 1: Cut scores and Indices of Classification Accuracy for Detecting Faking Good.

Subscale	Cut Score	% Accurate	% False Positive	% False Negative
BIDR Total	207	89.90	1.92	8.17
Impression Management	109	87.95	1.34	10.71
Self-Deceptive Enhancement	99	85.27	4.24	10.49
BFI Total	183	93.75	1.92	4.33
Agreeableness	41	83.17	9.13	7.69
Conscientiousness	41	88.94	5.29	5.77
Extraversion	31	76.92	13.46	9.62
Neuroticism (reverse scored)	33	90.38	2.88	6.73
Openness	42	75.48	12.98	11.54

Note: BIDR = Balanced Inventory of Desirable Responding [9]; BFI = Big Five Inventory [22].

Table 2: Cut scores and Indices of Classification Accuracy for Detecting Faking Bad.

Subscale	Cut Score	% Accurate	% False Positive	% False Negative
BIDR Total	127	93.00	3.80	3.20
Impression Management	48	93.30	1.12	5.58
Self-Deceptive Enhancement	68	80.36	7.81	11.83
BFI Total	130	95.67	1.44	2.88
Agreeableness	25	95.19	0.48	4.33
Conscientiousness	18	94.23	0.00	5.77
Extraversion	18	79.81	3.85	16.35
Neuroticism (reverse scored)	12	88.94	1.44	9.62
Openness	22	93.75	0.48	5.77

Note: BIDR = Balanced Inventory of Desirable Responding [9]; BFI = Big Five Inventory [22].

Summary and Conclusions

Since the beginning of formal uses of objectively scored self-report measures, socially desirable responding and related tendencies to fake responses have been serious concerns for users of results from such measures. Accordingly, a wide variety of methods to discourage and detect such invalid response tendencies have been suggested and evaluated over the years. Administering validity or “lie” scales along with the targeted measures of interest is perhaps the most common technique used to detect faking but has the drawback of requiring administration of extra items. To address such inefficiencies, we revisited the technique of using total scores for measures that assess multiple, and ideally weakly correlated, constructs. To provide a strong test of the effectiveness of the “total score-based” method, we implemented it using a brief length but widely administered personality inventory along with one of most comprehensive direct measures of socially desirable responding available to large and equal size honest and faking groups. Results were very encouraging in showing that the “total score-based” method outperformed popular extended length validity scales while requiring no additional items beyond those from

the targeted instrument. The total score technique also would be expected to perform even better with longer measures, additional subscales, and lower correlations among subscale scores. We, therefore, encourage further research into and application of this promising “re-discovered” procedure for detecting faking on self-report measures sharing the characteristics considered here.

References

- Hathaway SR, McKinley JC (1943) The Minnesota Multiphasic Personality Inventory (2nd Edn.), University of Minnesota Press: Minneapolis MN USA.
- Meehl PE, Hathaway SR (1946) The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology* 30(5): 525-564.
- Crowne DP, Marlowe D (1960) A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4): 349-354.
- Edwards AL (1957) *The Social Desirability Variable in Personality Assessment and Research*. Dryden Press: Fort Worth TX, USA.
- Eysenck HJ, Eysenck SBG (1964) *Manual of the Eysenck Personality Inventory*; University of London: London, UK.

6. Larsen KS, Martin HJ, Ettinger RH, Nelson J (1976) Approval seeking, social cost, and aggression: A scale and some dynamics. *The Journal of Psychology* 94(1): 3-11.
7. Jacobson LI, Kellogg RW, Cauce AM, Slavin RS (1977) A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society* 9: 109-110.
8. Sackeim HA, Gur RC (1978) Self-deception, self-confrontation, and consciousness. In: GE Schwartz, SD Shapiro (Eds.), *Consciousness and Self-regulation: Advances in Research*. New York, NY, USA: Plenum, pp. 139-197.
9. Paulhus DL (1991) Measurement and control of response bias. In: JP Robinson, PR Shaver, LS Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes*. San Diego, CA, USA: Academic Press, p. 17-59.
10. Paulhus DL (1998) Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding-7, User's Manual. Toronto, Ontario, Canada: Multi-Health Systems, Inc.
11. Paulhus DL (1984) Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46(3): 598-609.
12. Paulhus DL (1986) Self-deception and impression management in test responses. In: A Angleitner, JS Wiggins (Eds.), *Personality Assessment via Questionnaire*. New York, NY, USA: Springer-Verlag, pp. 143-165.
13. Paulhus DL (2002) Socially desirable responding: The evolution of a construct. In: HI Braun, DN Jackson (Eds.), *Role of Constructs in Psychological and Educational Measurement*. Mahwah, NJ, USA: Lawrence Erlbaum, p. 49-69.
14. Paulhus DL, Bruce MN, Trapnell PD (1995) Effect of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin* 21: 100-108.
15. Vispoel WP, Kilinc M, Schneider WS (2023) Detecting faking on self-report measures using the Balanced Inventory of Desirable Responding. *Psych* 5(4): 1109-1121.
16. Kilinc M (2020) Psychometric properties of full and reduced length forms of the Balanced Inventory of Desirable Responding under honest and faking conditions. Doctoral dissertation University of Iowa.
17. Ziegler M, Maccann C, Roberts RD (2012) *New Perspectives on Faking in Personality Assessment*. New York, NY, USA: Oxford University Press.
18. Clough SJ (2008) Computerized versus paper-and-pencil assessment of socially desirable responding: Score congruence, completion time, and respondent preferences. Doctoral dissertation, University of Iowa.
19. Schneider WS, Using IRLS estimators to detect faking on personality inventories. Doctoral dissertation, University of Iowa.
20. Comrey AL, Backer TE (1975) Detection of faking on the Comrey Personality Scales. *Multivariate Behavioral Research* 10(3): 311-319.
21. Vispoel WP, Kilinc M (2020) Do We Really Need Lie Scales to Detect Faking on Personality Inventories? Paper presented at the annual meeting of the American Psychological Association.
22. John OP, Donahue EM, Kentle, RL (1991) *The Big Five Inventory--Versions 4a and 54*. Berkeley, CA, USA: University of California, Berkeley, Institute of Personality and Social Research.
23. Soto CJ, John OP (2017) The next Big Five Inventory (BFI-2): Developing and accessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology* 113(1): 117-143.
24. Vispoel WP, Morris CA, Kilinc M (2018) Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods* 23(1): 1-26.
25. Vispoel WP, Xu G, Kilinc M (2021) Expanding G-theory models to incorporate congeneric relationships: Illustrations using the Big Five Inventory. *Journal of Personality Assessment* 103(4): 429-442.
26. Vispoel WP, Xu G, Schneider WS (2022) Using parallel splits with self-report and other measures to enhance precision in generalizability theory analyses. *Journal of Personality Assessment* 104(3): 303-319.
27. Vispoel WP, Tao S (2013) A generalizability analysis of score consistency for the Balanced Inventory of Desirable Responding. *Psychological Assessment* 25(1): 94-104.
28. Vispoel WP, Kim HY (2014) Psychometric properties for the Balanced Inventory of Desirable Responding: Dichotomous versus polytomous conventional and IRT scoring. *Psychological Assessment* 26(3): 878-891.
29. Vispoel WP, Morris CA, Kilinc M (2018) Practical applications of generalizability theory for designing, evaluating, and improving psychological assessments. *Journal of Personality Assessment* 100(1): 53-67.
30. Vispoel WP, Morris CA, Clough SJ (2018) Interchangeability of results from computerized and traditional administration of the BIDR: Convenience can match reality. *Journal of Personality Assessment* 101(3): 237-252
31. Vispoel WP, Lee H, Chen T, Hong H (2023) Using structural equation modeling to reproduce and extend ANOVA-based generalizability theory analyses for psychological assessments. *Psych* 5(2): 249-272.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2024.56.008917

Walter P Vispoel, Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>**Assets of Publishing with us**

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>