Psychologische Beiträge, Band 44, 2002, S. 50-61

Faking bad in personality inventories: Consequences for the clinical context

GABRIELE HELGA FRANKE[1]

Summary

A testee's response pattern to personality inventories may not necessarily reflect the truth as desired by test developers and users. Biased response sets which have generated the most attention are the "faking" sets. The fake bad response style is defined as any conscious or unconscious attempt to produce item-endorsement representing a more exaggerated symptomatology than is actually experienced by the respondent, with components of unbelievability and an unrealistic negative impression. Many commonly used personality inventories and self-report measures of psychopathology have been shown to be susceptible to attempts by respondents to distort results. Therefore several methods of detecting deviant response sets were incorporated, additionally several separate faking bad indicators have been developed. Faking bad response behaviour is usually investigated using a specified description of the disorder which should be simulated (e.g.: psychosis, depression, cognitive impairment, cancer-phobia, problems with alcohol, coping poorly with chronic pain). Several experimental studies demonstrated measurable effects of a faking bad response style in clinical, medical, forensic and neuropsychology. From an empirical point of view faking bad response behaviour is expected especially in the forensic area. High false positive rates (e.g., *Structured Inventory of Malingered Symptomatology*, Edens et al., 1999) among clinical populations in detecting potential faking bad behaviour illustrated forthcoming ethical problems. The diagnosis of faking bad behaviour in the single case is a complex process which requires sound knowledge of clinical, medical, forensic and neuropsychological areas of psychodiagnostic. The gold standard should comprise flexible usage of specific assessment techniques (e.g., individual tests, individual deficit tests, tests specifically for malingerers, and symptom validity testing), analysis of single case data against the background of documented normative data of healthy and clinical populations (e.g., threshold models) and multi-modal psychodiagnostic approaches – using different psychodiagnostic techniques.

Key words: assessment, faking bad, personality inventories, clinical consequences

Introduction

A testee's response pattern to personality inventories may not necessarily reflect the truth as desired by test developers and users. This problem of responding "inappropriately" interferes with the validity of any interpretation of the test scores.

Among biased response sets the "faking" sets received utmost attention. Johnson (1981) characterised two poles of the response pattern (a) "self-disclosure" as a simple communication of facts and (b) self presentation as "a social act intended to instruct others about how one is to be regarded" (p. 761). Recent developments in the design of personality inventories might be regarded as the pursuit of recognising self-presentation and understanding the complexity it introduces into measured personality scores (Pinsoneault, 1996).

[1] Dr. Gabriele Helga Franke, University of Applied Sciences Magdeburg and Stendal, Psychology of Rehabilitation, Osterburger Straße 25, D-39576 Stendal, Germany

Social desirability, or the tendency to present what is perceived as socially desirable responses, was an early explanatory paradigm. Initially, it was thought to be an attribute of test items and should be avoidable through careful test construction (Crowne & Marlowe, 1960). It was soon clear though, that to deal with people who attempt to make themselves "look good" on personality inventories requires more than omitting certain items (Pinsoneault, 1996).

Some authors have interpreted faking as a conscious, intentional attempt to respond less than honestly to personality items (Gordon & Gross, 1978; Rotter, 1960; Peltier & Walsh, 1990). Others have asserted that some distortion is due to conscious motivations but that some is also due to unconscious personality characteristics (Cattell, Ebert & Tatsuoka, 1970; Dicken, 1959; Eysenck, Eysenck & Shaw, 1974; Jemail & LoPiccolo, 1982; Mills & Hogan, 1978). Faking can be done to look better or worse than one is, and authors have suggested that a "fake good" set can be due to a naive or uninsightful, but honest self-assessment (Lachar, 1974) as well as to deliberate distortion.

The faking bad set has been defined broadly as any "conscious or unconscious attempt to produce item-endorsement patterns representing a more exaggerated symptomatology than is actually experienced by the respondent" (Archer et al., 1987, p. 507). On the other hand it could be argued, that this specific response set includes a component of unbelievability. Graham (1990) described fake bad subjects as being "motivated to present an unrealistically negative impression" (p. 35) of themselves.

Assessment of faking bad

Personality inventories

Many commonly used personality inventories such as the *Minnesota Multiphasic Personality Inventory-2* (MMPI-2; Butcher et al., 1989), the *Sixteen Personality Factors Questionnaire* (16-PF; Cattell et al., 1970), the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1975), the *California Psychological Inventory* (CPI; Gough, 1957), and self-report measures of psychopathology like the *Millon Clinical Multiaxial Inventory* (MCMI; Millon, 1983) have been shown to be susceptible to attempts by respondents to distort results and therefore incorporated several methods to detect deviant response sets.

The most widely used self-report measure of psychopathology, the MMPI, includes a number of indicators and scales designed to assess faked responses. These indicators and scales have been the focus of the majority of studies examining faking response behaviour (Brems & Harris, 1996; Dannebaum & Lanyon, 1993; Grillo, Brown, Hilsabeck, Price & Lees-Haley, 1994; Lim & Butcher, 1996; Lucio & Valencia, 1997; McGrath et al., 2000; Mihura, Schlottmann & Scott, 2000; Stein, Graham & Williams, 1995).

Bagby, Nicholson and Buis (1995, 1998) and Bagby, Rogers and Buis (1994) demonstrated the usefulness of the *Infrequency (F) Scale* of the MMPI-2 in the detection of a faking bad response style. McGrath et al. (2000) presented a new *Infrequency-Psychopathology Scale* (Fp-A) based on the MMPI-A (Butcher et al, 1992) detecting faking-bad.

On the other hand, Dannenbaum and Lanyon (1993) demonstrated that faking subjects' scores on subtle scales on the MMPI tend to be lower than those of non-faking subjects. They argued that this paradox comes about because the subtle scales do not possess predictive validity, but their face validity for psychopathology is the opposite of the intended direction. Subjects who attempt to fake psychopathology do so on the basis of item content and thus achieve lower rather than higher scores. Mihura et al. (2000), however, replicated this result only partly.

Investigating 90 personal injury claimants (MMPI-2, MCMI-II), Grillo et al. (1994) argued, that personality disorders, rather than malingering, contribute to exaggerated results in this forensic setting.

Bagby and Pajouhandeh (1997) investigated the two validity scores (*Too Good* (TG) and *Suspect Questionnaire* (SQ)) of the *Personality Diagnostic Questionnaire-4*, a 99-items, self-report, true/false questionnaire designed to assess the ten personality disorders as classified and defined in the DSM-IV. The scales TG and SQ consist of four and two items, respectively. The TG scale is designed to assess fake good responding or under-reporting; the SQ scale assesses either lying, responding randomly, or not taking the questionnaire seriously. Bagby and Pajouhandeh (1997) concluded that the overall classification rates (detecting faking good) of the two validity scales were too low to warrant use in a clinical setting.

Others developed new validity scales for existing inventories, e.g. the *Validity Scales* (Schinka, Kinder & Kremer, 1997) for the NEO-PI-R (Costa & McCrae, 1992), because this inventory does not include objective validity scales. The three validity scales *Negative Presentation Management* measuring faking bad, *Positive Presentation Management* measuring faking good, and *Inconsistency* scale detecting random responding were developed by systematically selecting items from the NEO-PI-R item pool. Several iterative item and scale analyses were conducted using multiple criteria for item selection, resulting in 10-item scales. Caldwell-Andrews, Baer and Berry (2000) demonstrated the usefulness of the proposed validity scales.

Pinsoneault (1996) developed validity scales (nine-item faking good scale: *Lie*; twenty-item faking bad scale: *Overt Symptomatology*) for the *Jesness Inventory* (Jesness, 1983), the second most widely used personality inventory in juvenile court clinics in the USA, using a rational approach (examining the items of the whole test, the items being judged by a panel of experts).

In most cases, validity scales are not used to modify obtained scores on the test, although the *K* scale of the MMPI-2 is an exception. More often, validity scales indicate whether the test taker's responses reflect a response set such as positive or negative impression management (Caldwell-Andrews et al., 2000). Such validity scales are particularly useful in cases where personal contact between clinician and test-taker is likely to be difficult (e.g., large group testing) or where test-takers have a compelling reason for presenting an unrealistic negative impression of themselves.

Separate faking bad indicators

A number of instruments have been developed for the purpose of identifying persons who are exaggerating or fabricating psychopathology, e.g. the *M Test* measuring malingering in schizophrenic individuals (Beaber et al., 1985), the *Malignering Scale* detecting prison inmates who fake insanity or mental retardation (Schretlen & Arkowitz, 1990), the *Malingering Probability Scale* (MPS; Silverton & Gruber, 1990), the *Structured Interview of Reported Symptoms* (Rogers, Bagby & Dickens, 1992), and the *Structured Inventory of Malingered Symptomatology* (SIMS; Smith & Burger, 1997).

Edens, Otto and Dwyer (1999) investigated the utility of the SIMS to distinguish truly symptomatic persons from persons instructed to fabricate symptomatology. In a sample of 197 participants who completed the SIMS and the *Symptom Checklist* (SCL-90-R; Derogatis, 1992, Franke, 1997, 1999, Schmitz et al., 2000) under both honest and malingering instructional sets, sensitivity and specificity rates were generally high for the SIMS. However, moderate correlations with the SCL-90-R were obtained, and specificity rates were lowest among a subset of participants reporting clinically significant levels of distress. Both findings raise concerns regarding the potential for high false positive rates among clinical populations.

Experimental methods in detecting faking bad

Several authors investigated faking behaviour in experimental settings with different faking instructions (e.g., faking good, bad; answering honestly, at random) in different languages (e.g. Spanish in Mexico; Lucio & Valencia, 1997). The traditional approach investigating faking good, e.g., asking subjects to answer "as if they had no psychological problems" (Bagby & Pajouhandeh, 1997, p. 306) typically relies on the subject's own broad interpretation of what is required and might lead to equivocal results. Alternatively, subjects can be instructed to fake good responses according to particular stereotypes, e.g. faking towards the personality of a successful stockbroker or librarian (Elliot, Lawty-Jones & Jackson, 1996). Viswesvaran and Ones (1999) give a good overview.

Usually faking bad is investigated using a specified description of the disorder which should be simulated. Edens, Ott and Dwyer (1999) instructed students to malinger one of three disorders (psychosis, depression, cognitive impairment), because "they were litigants in a personal injury lawsuit and were attempting to fabricate specific symptomatology (...) in an attempt to receive a large monetary award. (...) When instructed to malinger, participants were warned that the tests were designed to detect faking and that their task was to successfully fake the disorder without being identified. (...) Participants were provided with a summary of the disorder to insure that they were knowledgeable about the symptomatology they were to malinger." (p.398-399).

Lees-Haley (1989) instructed 52 untrained volunteers to simulate psychological disturbance resulting from exposure to a toxic substance from a hazardous waste site, because of an involvement "in a lawsuit for damages associated with this experience (...). Their goal was to fake a psychological disturbance in a manner calculated to convince the psychologist that they were psychologically disturbed and 'cancerphobic' as a result of the litigated stressful experience. They were warned that their answers could come out in court, so they should use a response style which would deceive the psychologist and the jury." (p.1205).

Myerholtz and Rosenberg (1998, p. 442) instructed: "Pretend that you are in a situation where you want people to think you have an alcohol problem. You feel it would be to your benefit to look as though you do have problems with alcohol. The person reviewing the

questionnaire will be looking for clumsy attempts to 'look bad', so try not to appear as if you are a hard core alcoholic. Your goal when completing the questionnaire is to answer in a way that will make you seem as though you are being open and honest, but still try to convince the reviewer that you do have problems with alcohol."

Robinson et al. (1997, p. 76) instructed students to present themselves as "coping poorly": "You are to pretend that you have a chronic pain problem. That is, your problem with pain has lasted longer than 6 months and has not responded adequately to early treatment."

Comparing these different instructions led to questions about potentially equivocal results and the comparability to malingering in real life.

Empirical methods in detecting faking bad

From an empirical point of view, a study sample – instructed to answer honestly – was analysed regarding questions about faking bad. Grillo et al. (1994) argued that personality disorders, rather than malingering, contributed to exaggerated results in a forensic setting. Pinsoneault (1996) pointed out that juvenile delinquents are not a population known for co-operation and honesty. Kaplan (1998) and Kaplan et al. (1998) discussed the problem that specific diseases like having a brain tumour led them to some of the psychological symptoms. In this case over-interpretation of psychological distress as well as interpreting faking bad behaviour should be avoided.

Whereas Wade et al. (1995) discussed their findings that women with bulimia nervosa score significantly lower on the *Lie Scale* than women who do not have bulimia nervosa; they argued that women with this eating disorder probably have a specific style of thinking and self-perception.

Single case studies

Brahmham and Dalgeish (2000) discussed a forensic case where someone was suspected of faking bad their reading ability. The authors used a modified *Stroop test*: the subject was required to name the ink colour in which crime-related words (that he claimed to be unable to read) and also non-word control stimuli were written. There was a significant difference between the proportional retardation of the subject's performance on crime-related words relative to non-words, when compared with reading-age-matched controls. The subject's relatively retarded latency in colour-naming crime-related words suggested that he was able to read at least some of these words, contrary to his claims.

Because the number of medical expert opinions dealing with smell and taste disorders has continuously increased in recent years, Delank et al. (1999) discussed specific aspects of medical expert opinions of investigating malingering of smell and taste disorders. They concluded: "Malingering was diagnosed on a trial of multiple symptoms. Retronasal olfaction was tested using the Guettrich gustatory olfaction test" (p. 365) – an over-threshold test of smelling that naive patients thought would really only measure taste.

Faking bad behaviour (e.g., malingering reading or smelling disability) is detectable using diagnostic approaches not known and/or detectable to the client. In conclusion, detecting malingering reading or smelling disability could be regarded as playing hide-and-seek. A patient informed about the technological implications of the discussed specific diagnostic strategies might be able to successfully simulate the disorder.

Clinical implications

Clinical and medical psychology

Myerholtz and Rosenberg (1998) demonstrated the measurable effects of faking good and bad instructions on college students' answering behaviour on the *Substance Abuse Subtle Inventory-2* (SASSI-2; Miller, 1994), a relatively common screening instrument with university students in the US.

Students, instructed to fake a "coping poorly" response set were able to produce scores similar to those of symptomatic chronic pain patients in three common self-report pain assessment measures, the *Coping Strategies Questionnaire* (CPI; Rosenstiel & Keefe, 1983), the *Multidimensional Pain Inventory* (MPI; Kerns, Turk & Rudy, 1985), and the *Pain Beliefs and Perceptions Inventory* (PBPI; Williams & Thorn, 1989). Robinson et al. (1997) suggested that, without validity indicators, it is difficult, if not impossible, for clinicians or researchers to determine whether a given profile obtained from a patient or subject reflects a sincere approach to the testing situation. However, Wallis and Bogduk (1996) concluded that it was possible to identify the faking bad answers of students, faking neck pain following whiplash injury, on the SCL-90-R, because the students scored statistically significantly higher than those of real patients. They concluded that it is very difficult for an ingenious individual to fake a psychological profile typical for a whiplash patient.

In conclusion, tendencies of faking bad behaviour is detectable analysing data of a single case against the background of documented normative data of healthy and clinical populations.

Forensic psychology

Lees-Haley (1989) instructed 52 untrained volunteers to simulate psychological disturbance resulting from exposure to a toxic substance from a hazardous waste site, because of involvement in a lawsuit for damages associated with this experience. The resulting T-scores of the SCL-90-R were associated with a level of psychopathology characteristic of the upper end of the average range expected from psychiatric in- and outpatients.

In a study dealing with parenting measurements used for assessing risk for physical child abuse, faking good was detected in a higher percentage (>90%) than faking bad (58%; Milner & Crouch, 1997).

In forensic psychology only multi-modal psychodiagnostic approaches – using different psychodiagnostic techniques – should be regarded as standard.

Neuropsychology

Regarding neuropsychological assessment malingering, factitious disorders and neurotically determined response behaviour usually because of financial reasons are discussed. Bernard, McGrath and Houston (1996) investigated the effects of faking bad on the *Wisconsin Card Sorting Test* (Heaton, 1981) in a study comparing the performance of simulating malingerers (n=24) to controls (n=21), closed head-injured patients (n=70) and patients with mired CNS pathology other than closed head-injury alone. Stepwise discriminant analysis achieved good accuracy (91-96%), sensitivity (58-100%), and specificity (92-100%) in differentiating simulating malingerers from these groups. The *Categories Score* served as a consistent significant independent discriminating variable. *Categories* alone differentiated between the malingering group and the three other groups. However, discrimination between simulating malingerers and both clinical groups, produced more complex patterns of performance, that led to poorer performance ratios on *Categories* compared to Perseverative Errors. This supported the Pattern of Performance theory on the effects of the simulation of malingering within neuropsychological tests. This theory states that people who simulate malingering do more poorly on obvious versus subtle tasks compared to people with verified brain damage. Due to the difficulty of tracking one's performance on several different indices simultaneously, known patterns of performance for patients with verified brain dysfunction should be very difficult to fake, even with coaching.

Kaplan (1998) and Kaplan et al. (1998) discussed the problem of interpreting the psychological distress of patients with brain tumour, because a percentage of the reported distress could be attributable to the natural history of the basic disease (physical consequences of structural brain lesions, treatment effects and side-effects). They advised clinicians to be conservative when interpreting a SCL-90-R profile obtained from patients undergoing treatment for brain tumours.

In neuropsychology threshold models for the detection of malingering as well as specifically designed assessment techniques (e.g., deficit tests, tests specifically for malingerers, and symptom validity testing) are discussed.

Conclusion

Up until today, it has not been known how often subjects respond with fake good or fake bad response behaviour in clinical or other settings. Its prevalence is a function of the setting, the individual, and the particular situation of testing. Suggested figures for the fake good set have varied from 1-15% in clinical or counseling settings (Butcher et al. 1989; Krug, 1978) to 5-30% in adolescent populations (Archer, 1992; Butcher et al., 1992) to 26-61% in forensic settings (Wasyliw et al. 1988; Grossman & Wasyliw, 1988). Comparable data for the fake bad response set were not found. However, a meta-analysis of 51 studies concluded that fakability did not vary by personality dimensions, all the "Big Five" factors were equally fakable (Viswesvaran & Ones, 1999). Faking produced the largest distortions in social desirability scales. Instructions to fake good produced lower effect sizes compared with instructions to fake bad. Within-subjects designs produced more accurate estimates, whereas between-subjects designs may distort estimates due to Subject x Treatment interactions and low statistical power.

Ones, Viswesvaran and Reiss (1996) concluded in their extensive review of the literature dealing with social desirability that a faking good answering style is not much of a concern in terms of personality and integrity for personnel selection. It appears that bias by influence of social desirability does not destroy the convergent and discriminant validity of the "Big Five" dimensions of personality. On the other hand, social desirability is a complex phenomenon. Ones and Viswesvaran (1998) reported, as a result of a meta-analytic study, that respondents instructed to fake good were able to change their responses by almost .50 standard deviations on the "Big Five" factors.

In conclusion, references to the validity of validity scales of personality inventories as well as exclusively developed malingering inventories were found. On the other hand, paradoxical results of faking bad instructions (Dannenbaum & Lanyon, 1993; Mihura, 2000) as well as paradoxical results based upon potential personality factors (Grillo et a., 1994; Wade et al., 1995) were discussed. However, the ecological power of these references is unclear. Simulation of faking bad behaviour (e.g., simulating psychopathological disorders) potentially leads to different results than real faking bad behaviour. High false positive rates (e.g., SIMS, Edens et al., 1999) among clinical populations in detecting potential faking bad behaviour illustrated forthcoming ethical problems: detecting false faking bad is a methodological malpractice and destroys the client's confidence in the psychologist.

However, on a basic level of test construction, a forced speed instruction for personality inventories will not solve the faking problem: "With increasing scores of the social desirability scale, the validity of the personality measurement definitely decreased, whereas a moderating effect on the validity, as far as working speed was concerned, could not be identified. Results of a factor analysis showed that working speed in personality questionnaires can be related to the performance aspect of perceptual speed." (Maschke, 1989, p. 121).

The diagnosis of faking bad behaviour in the single case is a complex process which requires sound knowledge of the clinical, medical, forensic and neuropsychological areas of psychodiagnostic. The gold standard should comprise flexible usage of specific assessment techniques (e.g., individual tests, individual deficit tests, tests specifically for malingerers, and symptom validity testing), analysis of the single case data against the background of documented normative data of healthy and clinical populations (e.g., threshold models) and multi-modal psychodiagnostic approaches using different psychodiagnostic techniques.

References

1. Archer, R.P. (1992). MMPI-A: Assessing adolescent psychopathology. Hillsdale, NJ: Lawrence Erlbaum Associates.
2. Archer, R.P., Gordon, R.A. & Kirchner, F.H. (1987). MMPI response-set characteristics among adolescents. Journal of Personality Assessment, 51, 506-516.
3. Bagby, R.M. & Pajouhandeh, P. (1997). The detection of faking good on the Personality Diagnostic Questionnaire-4. Assessment, 4, 305-309.
4. Bagby, R.M., Nicholson, R. & Buis, T. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding – replication and extension. Psychological Assessment, 7, 84-92.

5. Bagby, R.M., Nicholson, R. & Buis, T. (1998). Utility of the deceptive-subtle items in the detection of malingering. Journal of Personality Assessment, 70, 405-415.
6. Bagby, R.M., Rogers, R. & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. Journal of Personality Assessment, 62, 191-203.
7. Beaber, R.J., Marston, A., Michelli, J. & Mills, M.J. (1985). A brief test for measuring malingering in schizophrenic individuals. American Journal of Psychiatry, 142, 1478-1481.
8. Bernard, L.C., McGrath, M.J. & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "pattern of performance" hypothesis. Archives of Clinical Neuropsychology, 11, 231-245.
9. Bramham, J. & Dalgleish, T. (2000). Assessment of report dissimulation of reading ability using a modified stroop test: A case report. Clinical Psychology and Psychotherapy, 7, 411-414.
10. Brems, C. & Harris, K. (1996). Faking the MMPI-2: Utility of the subtle-obvious scales. Journal of Clincial Psychology, 52, 525-533.
11. Butcher, J.N., Dahlstrom, W.G., Graham, J.T., Tellegen, A. & Kaemmer, B. (1989). MMPI-2: Manual for administration and scoring. Minneapolis: University of Minnesota Press.
12. Butcher, J., Williamns, C., Graham, J., Archer, R., Tellegen, A., Ben-Porath, Y. & Kaemmer, B. (1992). MMPI-A: Manual for administration, scoring, and interpretation. Minneapolis: University of Minnesota Press.
13. Caldwell-Andrews, A., Baer, R.A. & Berry, D.T.R. (2000). Effects of response sets on NEO-PI-R scores and their relations to external criteria. Journal of Personality Assessment, 74, 472-488.
14. Cattell, R.B., Ebert, H.W. & Tatsuoka, M.M. (1970). Handbook for the 16PF. Champaign, IL: Institute for Personality and Ability Testing.
15. Costa, P.T. & McCrae, R. (1992). Professional Manual for the NEO Personality Inventory. Odessa, FL: Psychological Assessment Resources.
16. Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 24, 349-354.
17. Dannenbaum, S.E. & Lanyon, R.I. (1993). The use of subtle items in detecting deception. Journal of Personality Assessment, 61, 501-510.
18. Delank, K.W., Nieschalk, M., Schmal, F. & Stoll, W. (1999). Specific aspects of medical expert opinions of smell and taste disorders (German). Laryngo-Rhino-Otologie, 78, 365-372.
19. Derogatis, L.R. (1992). SCL-90-R, administration, scoring & procedures manual-II for the R(evised) version and other instruments of the Psychopathology Rating Scale Series. Townson: Clinical Psychometric Research, Inc.
20. Dicken, C.F. (1959). Simulated patterns in the Edwards Personal Preference Schedule. Journal of Applied Psychology, 43, 372-377.
21. Edens, J.F., Otto, R.K. & Dwyer, T. (1999). Utility of the structured inventory of malingered symptomatology in identifying persons motivated to malinger psychopathology. Journal of the American Academy of Psychiatry and the Law, 27, 387-396.
22. Elliot, S., Lawty-Jones, M. & Jackson, C. (1996). Effect of dissimulation on self-report and objective measures of personality. Personality and Individual Differences, 21, 335-343.
23. Eysenck, H.J. & Eysenck, S.B.G. (1975). Manual for the Eysenck Personality Questionnaire. London: Hodder & Stoughton.

24. Eysenck, S.B., Eysenck, H.J. & Shaw, L. (1974). The modification of personality and lie scores by special "honestly" instructions. British Journal of Social and Clinical Psychology, 13, 41-50.
25. Franke, G.H. (1997). The whole is more than the sum of its parts: The effect of grouping and randomizing items on the reliability and validity of questionnaires. European Journal of Psychological Assessment, 13, 67-74.
26. Franke, G.H. (1999). Effects of computer administration on the Symptom Checklist (SCL-90-R) with a special focus on the item sequence (German). Diagnostica, 45, 147-153.
27. Grillo, J., Brown, R.S., Hilsabeck, R., Price, J.R. & Lees-Haley, P.R. (1994). Raising doubts about claims of malingering – Implications of relationships between MCMI-II and MMPI-2 performances. Journal of Clincial Psychology, 50, 651-655.
28. Gorssmann, L.S. & Wasyliw, O.E. (1988). A psychometric study of sterotypes: Assessment of malingering in a criminal forensic group. Journal of Personality Assessment, 52, 549-563.
29. Gordon, M.E. & Gross, R.H. (1978). A critique of methods of operationalizing the concept of fakability. Educational and Psychological Measurement, 38, 771-782.
30. Gough, H.G. (1957). Manual for the California Psychological Inventory. Palo Alto, CA: Consulting Psychologists Press.
31. Graham, J.R. (1990). MMPI-2: Assessing personality and psychopathology. New York: Oxford University Press.
32. Heaton, R.K. (1981). Wisconsin Card Sorting Test Manual. Odessa: Psychological Assessment Resources.
33. Jemail, J.A. & LoPiccolo, J. (1982). A sexual and a marital defensiveness scale for each sex. American Journal of Family Therapy, 10, 33-40.
34. Jesness, C.F. (1983). Manual: The Jesness Inventory. Palo Alto, CA: Consulting Psychologists Press.
35. Johnson, J.A. (1981). The "self-disclosure" and "self-presentation" views of item response dynamics and personality scale validity. Journal of Personality and Social Psychology, 40, 761-769.
36. Kaplan, C.P. (1998). SCL-90-R interpretation and brain tumor: A correction factor? Brain Injury, 12, 977-985.
37. Kaplan, C.P., Miner, M.E., Mervis, L., Newton, H., McGregor, J.M. & Goodman, J.H. (1998). Interpretive risks: The use of the Hopkins Symptom Checklist 90 Revised (SCL-90-R) with brain tumor patients. Brain Injury, 12, 199-205.
38. Kerns, R.D., Turk, D.C. & Rudy, T.E. (1985). The West Haven Yale Multidimensional Pain Inventory (WHYMPI). Pain, 23, 345-356.
39. Krug, S.E. (1978). Further evidence of 16PF distortion scales. Journal of Personality assessment, 42, 513-518.
40. Lachar, D. (1974). The MMPI: Clinical assessment and automated interpretation. Los Angeles: Western Psychological Services.
41. Lees-Haley, P.R. (1989). Malingering emotional distress on the SCL-90-R: Toxic exposure and cancerphobia. Psychological Reports, 65, 1203-1208.
42. Lim, J. & Butcher, J.N. (1996). Detection of faking on the MMPI-2: Differences among faking-bad, denial, and claiming extreme virtue. Journal of Personality Assessment, 67, 1-25.
43. Lucio, E. & Valencia, M.D. (1997). Detection of simulant and honest subjects through MMPI-2 scales (Spanish). Salud Mental, 20, 23-33.

44. Maschke, P. (1989). Response time in personality questionnaires used in personnel selection: Does it indicate faking? (German). Zeitschrift für Differentielle und Diagnostische Psychologie, 10, 121-127.
45. McGrath, R.E., Pogge, D.L., Stein, L.A.R., Graham, J.R., Zaccario, M. & Piacentini, T. (2000). Development of an infrequent-psychopathology scale for the MMPI-A: The Fp-A-scale. Journal of Personality Assessment, 74, 282-295.
46. Mihura, J.L., Schlottmann, R.S. & Scott, A.B. (2000). Are the MMPI subtle subscales subtle measures of their scales? Journal of Clincial Psychology, 56, 139-148.
47. Miller, G.A. (1994). The Substance Abuse Subtle Screening Inventory Manual: Adult SIASSI-2 Manual Supplement. Indiana: The Spencer Evening World.
48. Millon, T. (1983). Millon Clinical Multiaxial Inventory Manual (3rd edition). Minneapolis: Interpretive Scoring Systems.
49. Milner, J.S. & Crouch, J.L. (1997). Impact and detection of response distortions on parenting measures used to assess risk for physical child abuse. Journal of Personality Assessment, 69, 633-650.
50. Mills, C. & Hogan, R. (1978). A role theoretical interpretation of personality scale item response. Journal of Personality, 46, 778-785.
51. Myerholtz, L. & Rosenberg, H. (1998). Screening college students for alcohol problems: Psychometric assessment of the SASSI-2. Journal of Studies on Alcohol, 59, 439-446.
52. Ones, D.S. & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. Human Performance, 11, 245-269.
53. 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.
54. Peltier, B.D. & Walsh, J.A. (1990). An investigation of response bias in the Chapman Scales. Educational and Psychological Measurement, 50, 803-815.
55. Pinsoneault, T.B. (1996). Rationally developed fake-good and fake-bad scales for the Jesness Inventory. Journal of Psychopathology and Behavioral Assessment, 18, 255-273.
56. Robinson, M.E., Myers, C.D., Sadler, I.J., Riley, J.L., Kvaal, S.A. & Geisser, M.E. (1997). Bias effects in three common self-report pain assessment measures. Clinical Journal of Pain, 13, 74-81.
57. Rogers, R., Bagby, M. & Dickens, S. (1992). Manual for the Structured Interview of Reported Symptoms. Odessa, FL: Psychological Assessment Resources.
58. Rosenstiel, A.K. & Keefe, F.J. (1983). The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. Pain, 33, 33-44.
59. Rotter, J.B. (1960). Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. Psychological Review, 67, 301-316.
60. Schinka, J.A., Kinder, B.N. & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. Journal of Personality Assessment, 68, 127-138.
61. Schmitz, N., Hartkamp, N., Kruse, J., Franke, G.H., Reister, G. & Tress, W. (2000). The Symptom Checklist-90-R (SCL-90-R): A German validation study. Quality of Life Research, 9, 185-193.
62. Schretlen, D. & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. Behavioral Sciences and Law, 8, 75-84.
63. Silverton, L. & Gruber, C. (1990). Malingering Probability Scale Manual. Los Angeles: Western Psychological Services.

64. Smith, G.P. & Burger, G.K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered symptomatology (SIMS). Journal of the American Academy of Psychiatry and the Law, 25, 183-189.
65. Stein, L.A.R., Graham, J.R. & Williams, C.L. (1995). Detecting fake-bad MMPI-A profiles. Journal of Personality Assessment, 65, 415-427.
66. Viswesvaran, C. & Ones, D.S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. Educational and Psychological Measurement, 59, 197-210.
67. Wade, T., Tiggemann, M., Heath, A.C., Abraham, S. & Martin, N.G. (1995). EPQ-R personality correlates of bulimia nervosa in an Australian twin population. Personality and Individual Differences, 18, 283-285.
68. Wallis, B.J. & Bogduk, N. (1996). Faking a profile: Can naive subjects simulate whiplash responses? Pain, 66, 223-227.
69. Wasyliw, O.E., Grossman, L.S., Haywood, T.W. & Cavanaugh, J.L. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. Journal of Personality Assessment, 52, 321-333.
70. Williams, D.A. & Thorn, B.E. (1989). An empirical assessment of pain beliefs. Pain, 36, 351-358.