International Journal of Assessment and Evaluation in Education
Vol 1/Dec 2011 (65-76) ISSN 2232-1926

Identifying Socially Desirable Responses in Personality Inventory

Priyalatha Govindasamy
Universiti Pendidikan Sultan Idris

Ong Saw Lan
Universiti Sains Malaysia

Abstract

The application of self-reported personality test in a competitive environment is found to be susceptible to biased responding (Hirsh & Peterson, 2008). This biased response causes difficulty in assessing an applicant's true scores in a standard selection process (Ellingson, Sackett & Hough, 1999). Therefore, this study aimed to detect the socially desirable responses from artificially differential response condition using a personality inventory. An experimental repeated measure design involved 521 students responding twice to the shortened International Personality Item Pool (IPIP) under honest and socially desirable instructions. The analysis of mean, score distribution and Rasch model's outfit indices were examined in differentiating the socially desirable responses from the honest responses. The socially desirable condition was found to have a higher mean compared to the honest condition. The percentage of respondents in the socially desirable condition obtaining scores with two standard deviations (2SD) above the mean was two times the honest condition. Additionally, two-third of the respondents with outfit values greater than 2.0 logits were from the socially desirable condition. Based on the findings in this study, it can be concluded that the score deviation greater than 2SD above the mean and the outfit values exceeding 2.0 logit are indications of high possibility of socially desirable responses. Therefore, test users for high-stake decisions can examine extreme high scores and an individual's inconsistent response as an initial detection of the socially desirable responses. This would help to minimize the issue of social desirability in high-stake testing.

Keywords *personality test, socially desirable responses, honest responses*

INTRODUCTION

The means of gathering and organizing information about a person's individual attributes is called personality assessment (Lanyon & Goodstein, 1982). Personality

assessment comprises a series of items that describe an individual's personality (Lonnqvist, 2008). Due to the descriptive nature of personality assessment, it was used extensively in clinical, counseling, business, industrial, governmental, military, educational and school contexts (Aiken, 1999). Inference about an individual from the personality inventory provides extra information which would be utilized in the selection of appropriate candidates (Carrigan, 2007). Hence, the personality inventory is relied upon when screening job applicants in employment selections (Li & Bagger, 2007). The predictive capacity of personality tests emerges as an aid in decision making and has evolved in the Malaysian context. The significance of personality is noted as an element in employment and in the educational setting. This is seen in the application of personality elements in the University Sains Malaysia (USM) entrance examinations known as the Malaysian University Selection Inventory (MUnSyI) and in the entrance examinations for teacher trainees known as the Malaysian Educators Selection Inventory (MEdSI). In addition, personality elements are included in the recruitment of graduate employees into the Malaysian civil service. Even though the personality tests are used in different settings, they are all administered to predict an individual's attributes in aiding decision-making. However, the application of self-reported personality tests in competitive environments has shown susceptibility to biased responding (Hirsh & Peterson, 2008). Studies have shown that people like to demonstrate good impressions in psychological testing (Dunn, 2009).

In cases where the test scores become a determinant for a person's future, the individual has the tendency to manipulate and answer dishonestly on the test (Aiken and Marnat, 2006). Fox and Meijer (2008) added that people often respond untruthfully on personal or sensitive questions in psychological or educational assessments. The tendency to respond in a way that presents the individual in a good light rather than to report in an accurate and truthful manner is known as social desirability (Holtgraves, 2004). Socially desirable responding has been described as individuals projecting a socially desirable image of themselves (Cervellione, Lee & Bonnano, 2009). Richman, Kiester, Weisband and Drasgow (1999) has stated social desirability as respondents' propensity in manipulating responses in a socially desirable manner under different test conditions, modes and administration. The socially desirable responses become a great concern to many as they reduce the validity of the personality measured (Crowne & Marlowe, 1960). The emergence of untruthful responding in personality testing opens to debate over the use of it during the selection process (Dilchert, Ones, Viswesvaran & Deller, 2006). This is because individuals have the tendency to fake personality testing to avoid any personal disclosures (Dunn, 2009). This raises the question on the accuracy of prediction of an individual and becomes a threat to the validity of the personality inventory (Kroner, Mills, Yessine & Hemmati, 2004). Ellingson et al. (1999) mentioned that in a standard selection process, it is difficult to assess an applicant's true scores. In such circumstances, selection is done based on the observed scores which could be the true or fake responses. Therefore, it is essential to examine the score outputs as well as the scoring pattern to determine any aberrant scores. Therefore, this study intended to investigate the extent to which the socially desirable distorted responses could be differentiated from the honest responses.

Factors contributing to Social Desirability in Personality Assessment

Faking in a test is reported as a result of person and situation interaction (Morgeson, Campion, Diboye, Holleabach & Schmitt, 2007). This means the degree of faking differs between individual and the situational demands. Day (2008) in his study found that the job applicants were more motivated to fake their responses compared to the incumbents of the job. In a study by Harvey, Wilson and Hansen (2005), the instructions to fake the personality test to the Troopers officers found score elevations compared to the responses in honest condition. Individuals manipulate their responses with the intention to gain social approval from others (Stocke and Hunkler, 2007). In critical situations such as a job application will stimulate motivation for the test takers to manipulate the responses. Individuals distort their responses to increase the likelihood of getting the job that they are applying for and the distortion occurring in employment is referred to as employment related motivational distortions (Hakstian & Ng, 2005). It is believed that lower self-esteem individuals try to respond in a socially desirable manner with the intention of gaining social approval from others (Magnus, Viswesvaran, Deshpande & Joseph, 2006). Besides that, the obvious meaning of items clearly indicates the most positive answers to the candidate which then enables the candidate to fake the answers easily (Morgeson et al., 2007). The familiarity of items could easily determine the social values placed on the scale items which lead to the likelihood of socially desirable responding (Dilchert et al., 2006). Moreover, the lack of resources in checking and verifying the characteristics of the applicants also contributes to faking in personality measures (Isaacson, Griffith, Kung, Lawrance & Wilson, 2008).

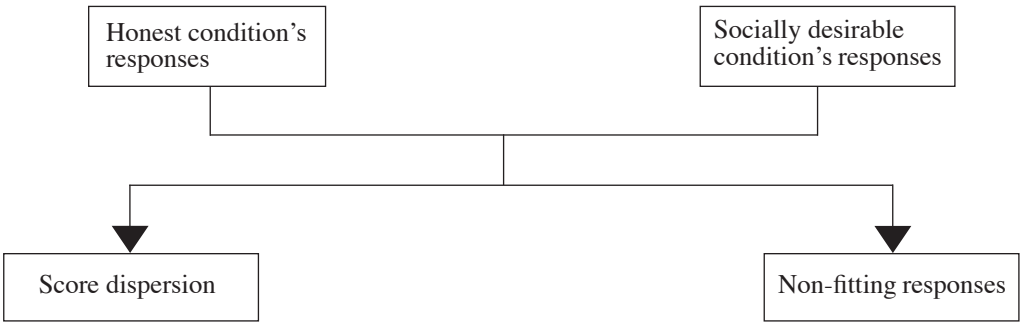
Underhill and Lords (2002) in their study found that the scaling format of an inventory is vulnerable to faking and less predictive in behavior. Jackson, Wroblewski and Asthon (2000) reported that the traditional single stimulus (Likert scale) items have scores that are higher than expected. The continuous items are more prone to faking because in continuous items the individual has the opportunity to fake to the maximum positive and minimum negative behaviors (Graham, McDaniel, Douglas & Snell, 2002). That shows that even the format of scale influences the individual's responses in an inventory. It is suggested that forced choice format items are fake resistant and better behavior predictors compared to the Likert scales (Underhill & Lords, 2006). Besides that, the commonly used personality inventory is rather general in context (Robie, Schmit, Ryan & Zickar, 2000). Problems arise when the general context personality items are administered for a specific purpose, for example for recruitment. When a general personality inventory is used for a specific purpose such as recruitment, it would fail to tap attributes required for employment but rather will provide a generalized understanding of an individual's characteristics which will be least useful for decision making. In addition, the paper and-pencil administered tests are found to provide opportunities for the test takers to distort their responses (Cruz & Dipboye, 2003). Research has found less social desirability when computerized testing is conducted (Richman et al., 1999). In conclusion, both person and personality inventory compliment each other in producing socially desirable responses.

Identifying socially desirable responses

Faking by applicants on personality measures are defined as deliberate alteration of responses (Griffith, Peterson, Quist, Benda and Evan, 2008). In socially desirable conditions, respondents have a tendency to inflate their scores in traits that respondents believed to portray a good image of themselves (Salgado, 2005). The score inflation can then result in incorrect decisions made about individuals. In the employment setting, studies have reported mean score differences in a test between job applicants and the incumbents of the job (Birkeland, Manson, Kisamore, Brannick & Smith, 2006; Griffith, Chmielowski & Yoshita, 2007). The mean score differences are reported as evidence of individuals' faking behaviors (Burkevisch, Jenkins & Griffith, 2007). This is because faking individuals' have the ability of inflating their scores a half standard deviation greater compared to the honest responses (Viswesvaran & Ones, 1999). A previous study has also found high scores in non-cognitive tests under high-stake situations as the cause of response distortions (Dilchert et al., 2006). Therefore, socially desirable responses have the tendency of score inflation and results in wrong reporting based on the score. In this study, the score distribution and dispersion from the mean was examined and compared between the honest and socially desirable responses.

In this study, the responses were analyzed by means of the Rasch model, which first establishes the expected response patterns to a particular set of items in a questionnaire and then provides a basis for identifying individuals whose responses are not consistent with the expected pattern. The differences between the observed responses and expected scores from the personality inventory are indications of a person misfit (Ferrando and Chico, 2001). This study adapted the approach of person fit to determine the non-fitting response patterns by comparing respondents observed scores to the expected responses from the model. In addition, scores or responses from non-cognitive measures can also be used in understanding the items and dimensions measured in an inventory. In high stake testing, respondents reported to have the tendency of suppressing their negative traits while enhancing their positive traits to portray their best image possible (Kirkcaldy, 2001). This shows the importance of selectiveness of items and traits to be responded in a socially desirable manner. Therefore, selectiveness in item responding and inconsistencies are studied parallel to their actual ability.

Conceptual Framework



METHODOLOGY

Sample

The sample consisted of 521 secondary school students aged between 14 to 16 years old. Of the total sample, 45.3% (n=236) were represented by males and 54.3% (n= 283) was represented by females. The composition according to ethnic background was 44.3% Malay, 47.4% Chinese, 6.5% Indian and 1.5% from other ethnic groups.

Design

An experimental repeated measure design was implemented in studying the socially desirable responses. In this study, the participants were asked to respond to the International Personality Item Pool (IPIP) inventory twice. During the first administration, the respondents were asked to respond honestly to the IPIP inventory. After three weeks interval, the same IPIP inventory was administered to the same sample under socially desirable responding instructions. The responses from the first administration were classified as the honest condition and responses from the second administration were grouped as the socially desirable condition. The responses from the honest and socially desirable conditions were compared to identify individuals whose responses were distorted in a socially desirable manner.

Measures

In this study, the International Personality Item Pool (IPIP) comprising 50 five-point Likert-type items was used. The IPIP measures the five-factor model of personality which taps an individual's level of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (Goldberg, 1999). Items in IPIP are scored using a five-point Likert scale, where "1" represents Strongly Disagree (SD) to "5" representing Strongly Agree (SA). However, the scores are reversed for negatively keyed items. The internal consistency reliability values of this study were reported as .74 (Openness), .78 (Conscientiousness), .69 (Extraversion), .75 (Agreeableness) and .69 (Neuroticism). According to George and Mallery (1995), reliability values of greater than .70 are reported as acceptable. Since all the dimensions reported values of closer and greater than .70, therefore the reliability values of this study is accepted.

Data Analysis

The personality scores from honest and socially desirable conditions were compared to identify the characteristics of the socially desirable responses. Firstly, the score deviations of honest and socially desirable conditions were examined. Scores from the honest and socially desirable conditions were combined and computed for the mean and standard deviations. Based on the mean, score dispersions were categorized accordingly by using the standard deviations intervals. Score distribution of the honest and socially desirable conditions was examined at each standard deviation interval

from the mean. The score distribution is used to examine the differences in score deviations between the honest and socially desirable conditions.

Next, the Rasch model was applied in studying the response patterns of the honest and socially desirable conditions. The individual response pattern was evaluated through person fit. Person fit examines the extent to which a respondent's response conforms to the model (Bell, 1982). The person fit approach evaluates the respondents abnormal response pattern through calculated standardized squared residual differences between the respondents observed and expected (from the model) scores after considering the person's ability and item difficulties (Bell, 1982). This study applied outfit statistics to detect the respondent's abnormal response patterns. The outfit statistics is a mean square fit statistic which is sensitive to unexpected response patterns on items (Linacre, 2006). The fit is deliberated through calculating the difference in each pair of observed and expected scores, squaring the differences, summing all the pairs and finally averaging and standardizing to a unit normal distribution. Therefore, it illustrates how much a person's observed score conforms to the expected scores from the model (Mueller, Bullock & Leierer, 2010). In order to detect socially desirable responses, the observed responses of honest and socially desirable conditions were combined and compared with the expected responses from the model. The acceptable outfit values ranged between 0.7 – 1.3 (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). Therefore, outfit values outside the acceptable range were classified as non-fitting responses. Then, the relative frequency of the non-fitting respondents in honest and socially desirable conditions was calculated.

RESULTS

Comparing the mean score deviations and differences between the honest and socially desirable conditions

Table 1 reports the mean and standard deviation descriptions for honest, socially desirable and combined data of honest and socially desirable conditions. The socially desirable condition reported a higher mean compared to the honest condition. However, the standard deviation was the same across all three conditions.

Table 1 Mean and standard deviation for honest, socially desirable and combined conditions

	Condition		
Description	Honest	Socially Desirable	Combined
Mean	162.1	163.3	162.7
Standard Deviation	18.18	18.18	18.19

The score distribution of honest and socially desirable conditions was studied by examining the condition's score deviations from the mean. Therefore, the combined mean was set as a base and the dispersion of the scores from the mean was categorized according to the standard deviation intervals. Table 2 reports the distribution of scores according to the standard deviation cutoff values.

Table 2 Score distribution for the honest and socially desirable conditions

Standard Deviation (Score range)	Condition				Total
	Honest		Socially Desirable		
	N	%	N	%	
2 SD below mean (below 127)	18	66.70%	9	33.30%	27
1 SD below mean (128 – 145)	70	48.95%	73	51.05%	143
1 SD within mean (146 – 181)	365	50.56%	357	49.45%	722
1 SD above mean (182 – 199)	57	47.11%	64	52.89%	121
2 SD above mean (above 200)	11	37.93%	18	62.07%	29
Total	521		521		1042

The percentage of responses at one and two standard deviation above the mean was higher for the socially desirable condition. The difference in percentage of score distribution for honest and socially desirable conditions was narrower at one standard deviation above the mean. At 1SD above the mean, honest and socially desirable conditions reported almost equal score distributions. The percentage of scores at 2SD above the mean indicating the socially desirable condition were two times more than the honest condition. On the other hand, 66.7% of scores at two standard deviation below the mean was represented by the honest responses. The score distributions at 2SD below and above the mean clearly discriminate between the two conditions of responding.

Distribution of non-fitting response of the Rasch model

Table 3 lists the summary of non-fitting outfit values. The outfit values revealed individuals' response patterns that do not conform to the Rasch model. The individuals' observed scores from the IPIP inventory differed from the estimated Rasch model score

patterns. In total, 255 out of 1042 responses reported non-fitting response patterns. The socially desirable condition constituted 69.4% of the non-fitting responses. Nonetheless, 83.5% of non-fitting responses fall within the outfit values of 1.3 to 2.0 and only 16.5% of the misfit responses were greater than 2.0 logits. A very high percentage, 78.6% of outfit values greater than 2.0 logits came from the socially desirable condition respondents. Compared to the honest condition, the socially desirable condition had two times more of students with outfit values bigger than 2.0 logits. The results showed that non-fitting responses with outfit values of more than 2.0 logits were much more likely to be produced by respondents in the socially desirable condition.

Table 3 Summary of non-fitting responses according to the conditions

Outfit t-values (logits)	Condition					
	Honest		Socially Desirable		Total	
	N	%	N	%	N	%
1.3 < t < 2.0	69	32.4	144	67.6	213	83.5
t > 2.0	9	21.4	33	78.6	42	16.5
Total	78	30.6	177	69.4	255	100

DISCUSSIONS AND CONCLUSION

The results showed differences in the score distribution between the honest and socially desirable conditions. The dominance and difference in the scores for the two conditions waere clear at two standard deviation away from the mean. The greater distribution of socially desirable responses at 2SD above the mean indicates the potentials of score inflation. This finding coheres with previous studies showing faking individuals have elevated their scores on the tests (Morgeson et al., 2007). Identification of faking individuals is important as Hirsh and Peterson (2008) described such high scores on traits as over self-presentation and biased responses rather than providing accurate personality description. In this study, the high percentage of socially desirable responses is those with high scores elevated to more than 2 SD above the mean. Therefore, scores deviation of more than 2SD above the mean is appropriately used to consider for initial screening for the presence of social desirability. Similar results have been reported in a study on job applicants who obtained high percentage scores in a test that dispersed at least two or three standard deviation above the mean (Rosse, Stecher, Miller & Levin, 1998).

Another piece of evidence in detecting the socially desirable responses was from the fit indices. The fit indice statistics is an internal mechanism for identifying inappropriate responses to the items (Green & Frantom, 2002). When individuals distort their responses in a socially desirable manner, it is expected that they will

provide inconsistent responses. In motivating settings, the discrepancy between the true nature and observed scores are thought to be due to the intentional response distortions (Dilchert et al., 2006). Therefore, the inconsistency between individuals' observed scores and expected scores of person were quantified and projected as the non-fitting responses. Findings from this study reported the majority of the non-fitting responses were from the socially desirable condition. Non-fitting responses from the socially desirable condition showed that socially desirable instruction did affect the response patterns and resulted in the discrepancy between their observed and expected responses. Two thirds of the responses with outfit values greater than 2.0 logits were from the socially desirable condition. The outfit values of more than 2.0 logits mean a greater inconsistency between observed and expected scores. Subsequently, an individual's response patterns that do not conform to the model with outfit values greater than 2.0 logits are a useful indicator to detect socially desirable responses in a personality inventory.

Results from this study supported earlier findings that showed score inflation and high dispersion from the mean as an indication of socially desirable responses. Besides, the non-fitting response patterns to the Rasch model were found to be indicating the presence of socially desirable responses. This illustrated preliminary evidence in detecting the socially desirable responses from a personality inventory. Therefore, the score dispersion and non-fitting response pattern can be applied in tests to identify individuals who might be responding in a socially desirable manner. The detection of socially desirable responses would help in screening out applicants in high-stake tests such as MEdSI and MUnSyI. It is also recommended that, universities, teacher training colleges, employers and even schools that use personality inventories should pay more attention to those with extreme high scores and inconsistent response pattern. This will reduce misrepresentation of the social desirability individuals in the selection with administration of personality inventory. By examining both the candidate's scores and personality inventory to minimize the social desirability issues can lead to more accurate predictions.

REFERENCES

Aiken, L.R. (1999). *Personality assessment methods and practices.* Third *Edition.* United States of America: Hografe and Huber Publishers.

Aiken, L. R., & Marnat, G. G. (2006). *Psychological testing and assessment*. United States of America: Pearson.

Bell, R. C. (1982). *Person fit and person reliability*. Retrieved from http://www.rasch.org/erp8.html.

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. Retrieved from http://www.olwnet.rice.edu/Birkeland.pdf

Burkevich, S. M., Jenkins, M., & Griffith, R. L. (2007). *Lying down on the job: Applicant faking and dependability*. Retrieved from http://www.cpla.fit.edu/io/documents/Burkevichetal2007.pdf

Carrigan, M. (2007). Pre-employment testing-prediction of employee success and legal issues: A revisitation of Griggs v. Duke Power. *Journal of Business and Economic*

Research, *5*(8), 35-44. Retrieved from http://www.cluteinstitute-onlinejournals.com/pdfs /172. pdf

Cervellione, K. L., Lee, Y. S., & Bonanno, G. A. (2009). Rasch modelling of the Self-Deception Scale of the Balanced Inventory of Desirable Responding (BIDR). *Educational and Psychological Measurement*, *69* (3), 438-458. doi: 10.1177/001316440

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Citation classic*, *24*(18), 349 – 354.

Cruz, P., & Dipboye, R. L. (2003, April).*Justice in personality testing: Influence of outcome, modality and elaboration*. Poster presented at the 18th Annual Conference of the Society of Industrial and Organizational Psychology, Orlando, FL. Retrieved from http://rcoes. rice. edu/docs.cruz&dipoye2003.pdf

Day, N. T. (2008). *Item and person characteristics as predictors of faking* (Unpublished master's thesis). Wright State University, Ohio. Retrieved from http://etd.ohiolink. edu/send-pdf.

Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessment. *Psychology Science*, *48*(3), 209 – 225. Retrieved from http://www.moityca. com.br/pdfs/ Response.pdf

Dunn, D. S. (2009).*Research methods for social psychology*. United Kingdom: Wiley Blackwell.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability correction in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*(2), 155–166.

Ferrando, P. J., & Chico, E. (2001). Detecting dissimulation in personality test scores: A comparison between person fit indices and detection scales. *Educational and Psychological Measurement*, *61*(6), 997 – 1012. Doi: 10.1177/00131640121971617

Fox, J. P., & Meijer, R. R. (2008). Using item response theory to obtain individual information from randomized response data: An application using cheating data. *Applied Psychological Measurement*, *32*(8), 595-610.doi:10.1177/0146621607312277

George, D., & Mallery, P. (1995). SPSS/PC+ step by step: A simple guide and reference. United States of America: Wedsworth Publishing Company.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several Five-Factor Model. *Personality Psychology in Europe*, *7*, 7 – 28.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and scores inflation with faking: Do items attributes explain variances across items? *Journal of Business and Psychology*, *16*(4), 573-592. Retrieved from http:// www.jstor. org/stable/250092794

Green, K., & Frantom, K. G. (2002, November).*Survey development and validation with the Rasch model*. Paper presented at International Conference on Questionnaire Development, Evaluation and Testing: Charleston, SC. Retrieved from http://www. jpsm.umd.edu/qdet/ final_pdf_papers/green.pdf

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*(3), 341 – 355. doi: 10.1108/00483480710731310

Griffith, R. L., Peterson, M. H., Quist, J., Benda, A., & Evans, A. L. (2008). *Faking the personality profile: Easier said than done*. Paper presented at the 23rd Annual conference for the Society of Industrial and Organizational Psychology: San Francisco, California Retrieved from http://cpla.fit.edu/ Griffth_Peterson_QuistBenda_Evan_ 2008_Final.pdf

Hakstian, A. R., & Ng, E. L. (2005). Employment-related motivational distortions: Its nature, measurement and prediction. *Educational and Psychological Measurement*, *65*(3), 405–441. doi: 10.1177/0012164404267293

Harvey, R. J., Wilson, M. A., & Hansen, R. L. (2005).*Detecting CPI faking in police sample: A cautionary note*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology. Los Angles, CA. Retrieved from http://www.harvey. psyc. vt.edu/ SIOP2005.handout. Harvey_Wilson_Hansen.pdf

Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with a "Fake-Proof" measure of the Big Five. *Journal of Research in Personality*, *42*, 1323-1333. doi: 10.1016/j-jrp.2008.04.006

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, *30*(2), 161-172. doi: 10.1177/0146167203259930

Issacson, J. A., Griffith, R. L., Kung, M. C., Lawrance, A., & Wilson, K. A. (2008).*Liar, liar: Examining background checks and applicants who fail them*. Paper presented at the 23rd annual conference for the Society for Industrial and Organizational Psychology, San Francisco, CA.

Jackson, D. N., Wroblewski, V. R., & Asthon, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution. *Human Performance*, *13*(4), 371 – 388.

Kirkcaldy, B. (2001). Individuals' ability to predict their own personality test scores. *Evaluation Journal of Australia*, *1*(2), 63-65.

Kroner, D. G., Mills.J. F., Yessine, A. K., & Hemmati, T. (2004). The generalized instructional set of Criminal Attribution Inventory (CRAI): Socially desirable responding and faking. *International Journal of Offender Therapy and Comparative Criminology*, *48*(3), 360-372. doi: 10.1177/0306624X04263452

Lanyon, R. L., & Goodstein, L. D. (1982). *Personality Assessment* . Second Edition. United States of America: Wiley-Interscience Publication.

Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement*, *67*(3), 525–544. doi: 10.1177/0013164406292087

Linacre, J. M. (2006). Winsteps (Version 3.65.0) [computer software]. Chicago: Winsteps.com

Lonnqvist, J. E. (2008). *Issues in socially desirable responding and personality research*. Helsinki: Helsinki University Printing House.

Magnus, J. M., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: the role of over-claiming, self-esteem and emotional intelligence. *Psychology Science*, 48(3), 336–356.

Morgeson, F. R., Campion, M. A., Diboye, R. L., Holleabach, J. R., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection context. *Personnel Psychology*, 683–729.

Mueller, C. E., Bullock, E. E., & Leierer, S. J. (2010). *Examining psychometric and measurement properties of the Career Thought Inventory: Demonstration and use of the Rasch measurement model in career assessment research*. Retrieved from http://www. career.fsu/techcentre/TR51.pdf

Richman, W. L., Kiesler, S., Weisband, S., Drasgow, F. (1999). A Meta-analytic study of social desirability distortions in Computer-Administered Questionnaire, traditional questionnaire and interviews. *Journal of Applied Psychology*, *84*(5), 754–775.

Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. *Organizational Research Methods*, *3*(4), 348–365. doi: 10.1177/109442810034003

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion in pre-employment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634–644.

Salgado, J. F. (2005). Personality and social desirability in organizational settings: Practical implications for work and organizational psychology. *Papeles del Psicologo*, *26*, 115–128. Retrieved from http://www.papelesdelpsicologo.es/emgish/1252.pdf

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). *Rasch fit statistics and sample size considerations for polytomous data*. Retrieved from http://www.biomedcentral.com/1471-2288/8/33/prepub.

Stocke, V., & Hunkler, C. (2007). Measures of desirability beliefs and their validity as indicators for socially desirable responding. *Field Methods*, *19*, 313–336. doi: 10.1177/1525822X07302102

Underhill, C. M., & Lords, A. O. (2002). *Fake resistance of forced-choice paired comparison personality measure*. Retrieved from http://www.internationalmta.org

Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis on fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*(2), 197–210.