International Journal of Arts Humanities and Social Sciences Studies
Volume 4 Issue 6 ǁ July 2019.
www.ijahss.com

Measurement and Control of Response Bias In Non-Cognitive Instrument

LONGJOHN, IBIENE TANDI (PhD), AJALA, IBRAHIM

Department of Educational Psychology, Guidance and Counseling, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria

Abstract*: Response bias continues to be a perennial challenges for social scientist, including educators, who depend considerably on non-cognitive instrument to generate data in their attempt to provide solution to educational challenges and knowledge advancement. Response bias, a generic term for the tendency to respond to questionnaire items on some basis other than the specific items in the questionnaire has generate concerns, yet few attempt has been made to measure and control it. In this paper, attempt was made in discussing two common response bias namely social desirability bias and extreme response biases. Also presented were some procedural steps to ameliorate its impact during the process of instrument development and some statistical techniques that have been developed to measure and control them in data analysis and interpretation. This work concludes by recommending the best possible approach for controlling response style within the context of a developing country like Nigeria.*

Keywords: Response style, response set, social desirability bias, extreme response style.

I. Introduction

Interest in non-cognitive assessment techniques such as interviews, observation, and rating scales can be stated as predating cognitive instrument because affective characteristics have often be assessed albert informally in homes, religious settings and other social gathering where emotional, personality and attitudinal traits are gauged for effective interaction and peaceful living. Today, the interest in these traits cuts across both educational and corporate environments (McCoach, Gable & Madura, 2013). In both educational and corporate settings, scholars, theorist and practitioners agree on the importance of personality and social characteristics such as self-efficacy, motivation, perceived satisfaction, teamwork, and leadership style for success and productivity or profitability. Although the above mentioned non-cognitive assessment techniques are employed both formally and informally in the assessment and development of these and other psychosocial characteristics, the one that has had the most widespread application in the formal setting including schools and business organisations starting from the beginning of the twentieth century are rating scales.

Along with this interest comes the need for the development of useful, valid and reliable instruments to tap into specific construct. Beyond the instrument development phase, it is equally vital that appropriate strategies be taken to ensure that during the process of responding, analysis and interpretation, only the construct of interest in measured and nothing else (Kpolovie, 2010). Instruments developed to assess affective characteristics have been referred to with various names such as questionnaire, inventory, scale etc. Irrespective of the nomenclature used, most psychological and affective construct depend on instruments in which respondents are expected to describe themselves (self-report) or whereby others are asked to comment about friends, family, leaders or colleagues (others-report) by responding to a set of items. Responses to these items are often provided on dichotomous formats (true or false, yes or no, present/not present) etc. Other types of response are provided on a fixed number of polytomous rating scales with wordings such as strongly agree, agree, disagree, agree or never, sometimes and always. It is a common assumption among researchers and practitioners, that scores from the response to questionnaire items only represent the underlying trait level and the random error. For instance a person who endorsed "strongly agree", on the item "I keep my room tidy"is assumed to be higher on the personality trait of conscientiousness than an individual who endorses agree on the some item (Wetzel, Bohnke & Brown, 2016). The validity of this approach is only possible, if and only if, other factors do not influence respondents on the option they endorse in an item of a questionnaire, which is often not possible.

In most research studies conducted using survey research methodology, four sources of error have been distinguished that confound the possible results obtained. According to Schwarz, Groves, Schuman (1998) these are coverage or frame errors, sampling error, non-response error and measurement error. Coverage or frame errors

arises when the members of the sample does not corresponds to the target population; sampling error occurs because it is usually impossible to collect measurement from members of the entire population as such the study is conducted with a sample of respondents. Non-response errors are those arising from respondents' inability to respond to some or all of the items in the instrument for data collection. Finally measurement error refers to differences between respondents' true value for the construct under study and the obtained value from measurements. According Baumgartner and Steenkamp (2005) measurement errors can be divided into those that arise from the method of data collection (e.g.interview, questionnaire, computer-based testing), contextual factors (e.g. time pressure, other respondents present, physical environment) and those from respondents.

The focus of this paper will be on measurement error arising from respondents although it possible for respondents' errors to be accentuated by data collection method or contextual factors. As most researchers can attest from previous testing experiences, respondents endorse items outside the content of the item itself. This is what researchers refer to as response bias (Paulhus, 1991; Nwankwo, 2013). Although scholars have used the terms response bias, response set and response style interchangeably (Nwankwo, 2013) some differences exist between them as identified by Paulhaus (1999:17). According to him, response set is a temporary reaction to a situational demand such as time pressure or public expectation in the way a person responds to an item in a questionnaire, while response style refers to an individual display of a response set consistently over time and across situations, while response bias which subsumes the above two terms refers to any tendency to a respond to a range of items on an instrument, questionnaire or inventory on some basis other than the specific item content (i.e. what the items were designed to measure).

Baumgartner and Steenkamp (2001) used the terms measurement error synonymously with response bias, as response errors arising from the respondents. Various forms of response bias have been gleaned by researchers but for this paper, only two of such bias will be discussed namely social desirability bias and extreme response bias and acquiescence bias. In the sections that follow, we firstly introduce the above mentioned bias and some strategies that have been developed aimed at minimizing them in the process of test construction. Then we highlight some methods developed for correcting the effects of response bias when they have been detected. We concluded by highlighting some limitations involved in assessing and controlling for responsebias. Before presenting the types of response bias, it is important we highlight dangers that could arise from failure to take them into consideration in social science research and practice.

II. Dangers of Response Bias

Scholars have found that response biases have some negative impact on research instruments and the inference made from it. According to Van Vaerenbergh and Thomas (2012) response bias affect means and variances, which are essential for further advanced statistics such as t-test and F-test. For example, Moors (2012) obtained results which shows that women are more likely to endorse the lowest and highest response categories of a rating scale than men in a study of gender differences and leadership styles. This is likely to results in spurious conclusion and invaded inference on gender differences and leadership style.

Most statistical techniques such as Cronbach alpha, regression analysis, factor analysis and structural equation modeling, rely on correlations between two or more variables. Therefore, empirical investigations examining relationships without measuring and controlling for the effect of response bias may obtain misleading result. An example might suffice in this context. Baumgartner and Steenkamp (2001) conducted a multivariate correlation between health consciousness (HCO) quality consciousness (QCO), environmental consciousness (ECO) and ethnocentrism (ETN) and obtained the following correlations: HCO - QCO 0.40; HCO-ECO = 0.33; HCO-ETN = 0.28; QCO-ECO = 0.31, QCO-ETN = 0.19 and ECO-ETN = 0.15. When improved upon by controlling for response style the following correlations were obtained: HCO-QCO = 0.20, HCO-ECO = 15, QCO-ECO = 0.13, HCO-ETN = 0.02, QCO-ETN = 0.00 and ECO-ETN = 0.01. A comparison of these values obtained before and after controlling for the response style shows that the magnitude of the relationship between variables were over bloated due to the effects of response style.

Finally, beyond wrong inference arising from response bias, it can equally lead to wrong decision regarding selection and admission process. Within the academic and corporate world, there is a greater emphasis on "soft skills", psychosocial qualities or characteristics that complements cognitive skills, as vital for truly productive academic and industrial success. The assessments of these skills have been integrated into the recruitment of employees and admission of students into learning institution. These skills include emotional and social intelligence, self- efficacy, multitasking abilities, attitude toward learning, leadership styles and value orientation (Partnership for 21st Century skills, 2008). When required to respond to questionnaires on such skills, as well as opinion on sensitive matters such as political ideology, religion, and abortion, people tend to present a favourable image of themselves (van de Mortel, 2008) or engaging in *satisficing*, whereby they engaging in less cognitive processing and making responding as easy as they can (Krosnick, 1991). These can lead to the selection

of wrong candidates for employment which in turn leads to poor productivity for corporate organization or the admission of students who lack "soft skills" needed to succeed on academic task. Some scholars have even found out that even large scale international studies such as the Organization for Economic Cooperation and Development's (OECD) Programme for International Students Assessment (PISA) which is a periodic cross-national assessment consisting of cognitive, attitudinal and socio demographic measures containing Likert-type scales also suffer from the problem of response bias (Buckley, 2009). As such it has been emphasized that practitioners take into consideration the effect of response bias when making decision about employment and recruitment, especially when Likert-type items are to be used.

III. Social Desirability Bias

Social desirability bias (SBA) is different from other types of response bias in that it requires reading and cognitively comprehending the item content for it to occur (Wetzel *et al*, 2016). According to Paulhus (2002) it is the tendency to "give positive self-description" on the basis of item content (p. 47). To Steenkamp, de Jong & Baumgartner (2009), socially desirable responses are answers that make the respondent look good, based on cultural norms about the desirability of certain values, traits, attitudes, interests, opinions, and behaviors. In the opinion of Johnson and Fendrich (2005), social desirability bias in the tendency for participant in a research activity to present themselves in positive light or present a favourable image of themselves. While for Mick (cited in Baumgartner & Steenkamp, 2005) social desirability responding (SDR) can be defined as the tendency for people to present themselves favourably according to cultural norms (p. 208). What constitute a socially desirable behaviour is dependent on a wide away if factors, and varies from one situation to another. The tendency to engage in SDR is higher when the questions are socially sensitive or personally gratifying. Researchers have found that respondent engage in socially desirable bias on topics such as domestic violence, racism, sexual practice, financial practice, (Babcook, Costa, Green & Eckhardt, 2004; Difranceisco, McAuliffe & Sikkema, 1998; Kelly 2015).

Palhaus (1991) stated that SDR is the most studied of all response bias by tracing its history to over 70 years ago. According to Baumgartner and Steenkamp (2005), initial conceptualization of SDR see it as a one-dimensional construct until Paulhus (1991) presented his two factor models of SDR consisting of impression management and self-deceptive enhancement. Impression management is an intentional component that is sometimes called faking. In this component, an individual knowingly reads the items in a questionnaire and deliberately endorse response categories that would present him in a positive image. Impression management is closely related to faking, dissimulation, lying and deceiving others (Baumgartner & Steenkamp, 2005). Self-deceptive enhancement is people's tendency to provide self-report that are honest but positively biased, which is a form of rigid over confidence. Other two-factor models of SDR exist such as Damerin and Messick, Sackein and Gur but Paulhus' conceptualization of SDR is currently the most influential and most utilized in SDR research.

Arguments abound as to the stability of SDR in individuals across time, situation and measurement instrument. Ones, Viswesvaran and Reiss (1996) argued that it's a relatively stable trait across domain, time and instruments while Chung and Monroe (2003) provide evidence that it varies depending on situation and religiousness. Studies on social desirability bias have been conducted along either of two strands, either as a characteristic of items or as an aspect of personality. As an aspect of personality, SDR bias has been found to be higher in females than males, in individualistic cultures than in collectivistic cultures (Lalwani, Shavit, Johnson & Zhang, 2005).

IV. Measurement and Control of Social Desirability Response

A basic goal of every research instrument is to ensure that the discriminating validity of an instrument is maximized, which social desirability bias and other measurement errors threatens. As such various approaches have been advanced for the measurement of social desirable bias. Some approaches recommended by Paulhaus (1991) for the control of SDR includes

The administration of various instrument measuring SDR simultaneously with instruments measuring the construct and correcting the score obtained using covariates and target rotation techniques. One might decide to discard a respondent's total score, if it is found out that the person engaged in a high level of SDR on the basis of a predetermined cut off point.

In this direction, various measures of SDR have been developed to assess its confounding effect in non-cognitive instrument, but the two most popular scales according to Baumgartner & Steenkamp (2005) are the Marlowe-Crowne Social Desirability Scale (MCSDS) and the Paulhus Balanced Inventory of Desirable Responding (BIDR). The MCSDS is a 33-item scale which describes either desirable but uncommon behavior (e.g. admitting mistakes) or undesirable but common behaviour (e.g. gossiping). Respondents are asked to

respond true or false to 18 items keyed in the true direction and 15 items in the false direction, although it has been recommended that one can use rating scales of end points with "true" and "false". The scale has shown good reliability with as high 0.70s to 0.80s. A short form version of the instrument made up of 13 items has also been developed by Raynolds (1982) with lower reliability of 0.60s and 0.70s. The BIDR scale is a 40 item instrument which measures the two SDR factors of impression management and deceptive enhancement, with 10 items positively keyed and 10 items negatively keyed for each factor. Responses to the 40 items are provided on a seven-point scale, but the developer also recommended that five-point scale. Paulhus recommends to dichotomize the scores into 0 (scores 1-5) and 1 (6-7) and to add the dichotomize scores to arrive at a respondents score on impression management and self-deception respectively. The two components of the scale reported average (0.60) and high (0.70) reliabilities for self-deceptive enhancement scale and impression management respectively. The total scale of the 40 items reported a considerable high reliability of 0.80 (Baumgartner & Skccnhamp, 2005).

The first step for controlling SDR during the process of developing the instrument is to assure the respondents of utmost anonymity. When respondents are assured that their identity would not be revealed or made public, they are most likely to give response that are free from the constraints of social pressure. This implies that researchers do not provide spaces for respondents to give their names, or other personal informational which can be traced to them such as registration number, matriculation number or home address.

Also researchers should refrain from captioning the items in the questionnaire to reflect the attributes or characteristics to be measure especially if the items are developed around a socially desirable construct. If the instrument has sections, as in a multivariate instrument, the various sections should not be named or captioned (Nwankwo, 2013). Another suggestion has been to shuffle the items of the instrument in a manner that no two items assessing a possible socially desirable traits are presented consecutively. It has been also suggested that researchers and test developers construct the items in a test in the "third form" where respondents are asked to give their assessment of what people in general, their neighbourhood,-or school, think about the topic.

Although questionnaire design may reduce SDR, it is quite unlikely to eliminate the tendency for respondents to present a positive image of themselves. To control for this, Baumgartner and Steekamp (2005:212) recommended three strategies on the basis that an SDR scale is attached. These approaches are (i) researchers can conduct a regression analysis where respondents' response to each individual item is regressed against the focal construct on their scores on the SDR measure. The residual of the regression analysis are the scores purged of a person's generalized tendency to respond in a socially desirable manner; (ii) by regressing a respondent total score on the focal construct on his or her score on the SDR instrument with the residual of the regression analysis being the person true score of the constructs free from SDR tendency; (iii) by adding a respondents score on the instrument as a covariates in analyses involving the focal constructs to other construct of interest.

V. Extreme Response Style Bias

As stated previously, while dichotomously and polytomous response categories have been routinely used in self-report and non-cognitive instrument, there is a greater utilization of dichotomous formats (right/wrong) for cognitive assessment, while non-cognitive instrument such as psychological tests and attitudinal survey questionnaires often adopts polytomous response categories (Cho, 2013). Research by Wetzel, Carstensen, and Bdhnke (2013) shows that response style are relatively higher in non-cognitive instruments than in cognitive instruments. Considerable debate exist on the menace or usefulness of extreme response style which we shall briefly discuss, but before addressing this debate it is vital we understand what extreme response style is.

According to Morren, Gelissen and Vermunt (2009) ,extreme response style is defined as the tendency of respondents to express themselves to an item by choosing the end points on rating scale or instrument, independent of the extremity of their opinion. Similarly, Paulhus (1991) stated that extreme response bias is the tendency to use the extreme choice on a rating scale (e.g. 1 and 7 on a seven-point scale, strongly agree or strongly disagree, always or never, very true of me or not true of me). This tendency is assumed to exist irrespective of the item content, but to consistently take the most extreme positive or negative categories of an item responses.

The problem of extreme response style has a long history, with Cronbach (1946) stating that response style always reduce logical validity because they contribute to effect of people with equal level of the construct under investigation to have different scores. Also Liu, Wu and Zambo (2009) found out that response style negatively impact on test reliability. Another issue on the dangers of response style is the differences .that exist among groups. According to Cho (2013) certain response style tend to be more prevalent in a particular group than in other groups. This variability in response style is likely to result in the violation of structural invariance and any observed difference between-group. Using multi-group confirmatory factor analysis, Cheng and Renvold (2000) provided evidence that extreme response style bias is responsible for certain measurement non-variance. Similarly, Morren et al (2009) using latent class factor analysis demonstrated that ERS influences group differences when the attitude of four culturally diverse individual was measured. Other studies in the United States

have shown that ERS is more prevalent among African-Americans than among Caucasians (Clarke III, 2000; Bachman, O'Meilley, Freedman-Doan, 2010), while East Asians (e.g. Chinese, Japanese and Koreans) in the US endorse more of extreme response on positive feelings than Americans (Lee, Jones, Mineyama, & Shanng, 2002; Lee & Green. 1991). Similarly, ERS have been found to be more common among Mediterranean nations (Spain, Italy and Greece) than Western Europe nations such UK, Germany and France (Vann Herk, Poortinga & Verhattcn, 2004). On gender basis, De Jong, Steenkamp, Fox & Baumgartner (2008) reported that more women display ERS bias than males. Educationally, Marin, Gamba and Marin (1992) and Meisenberg and Williams (2008) provided evidence that lower educated people have a greater tendency for ERS than those with higher educational qualification. The dangers in interpreting scores from various groups without taking into consideration their response style was aptly captured by Bolt and Johnson (2009: 336) thus:

> Variability in ERS scores across groups can influence effort to understand an instrument's measurement equivalence. Specifically, items can appear to function differentially in their measurement of the intended traits when respondents groups vary only in response. In this respect, response style can be viewed as a type of nuisance dimension that interferes with measurement of the intended trait

Not all researchers agree that ERS is a nuisance source confounding the measurement of psychological and personality trait. Some studies have shown the relationship between certain psychological construct and ERS. For instance, ERS appeared to be positively related to test anxiety, extraversión and conscientiousness (Norman in Cho, 2013; Austin Deary & Egan, 2006; Harzing, 2006).

VI. Measurement and Controlling of ERS

After recognizing that extreme response can be a statistical nuisance to be controlled for or a construct of interest that could be related with other variables, Cho (2013) suggested that any data analysis must first create a distinction of cases that are influenced by certain responses style. Thereafter, appropriate statistical methods can be employed to control for it or related with other variables.

The first method of detecting and controlling for ERS according to Morren et al (2009) is the ERS sum-score index whereby the original items on a survey arc scored with a score of one (1) referring to an extreme response choice and zero (0) assigned to any other choice between the extremes. Thereafter,the sum of all the extreme answers (1s) are added together. Due to the fact that some respondents may truly respond and endorse extreme options on an instrument based on the degree they possess the latent trait, this method has been questioned.

To correct this anomaly, Greenleaf (1992) developed a 16-item instrument, specifically designed to access ERS. The instrument consist of 16 items that are unrelated conceptually and theoretically, with little similarity. If a respondents consistently favours a particular response category (e.g. extreme response) across the items in the scale, this can be assumed as evidence of an extreme response style. This scale has been found useful in detecting respondents with ERS and also to ameliorate the effect of ERS is subsequent statistical analysis.

Another suggestion has been to use multiple choice format in a Likert-scale (Baungarten and Steenkemp (2005). Instead of using the conventional formats where respondents are required to respond to items on number or wording scales, researchers should use other formats such as forced choice formats, anchoring vignettes etc. Another suggestion has been to administer the ERS measure directly to the respondents or subject population especially those composed of rigid people or those who lack experience. This is equally recommended for individuals with low educational level, older individuals and those from minority population.

The use of balanced scales have also been recommended by Cho (2013), Paulhus (1991) and Nwankwo (2013). A balanced scale consist of pairs of logically reversed items whereby one item of a pair states a construct positively, while the other of the pair states the equivalent construct negatively. Although theoretically different from a balanced scale, Nwankwo (2013:119) recommended that in writing items on a scale, equal number of positive and negative items should be presented. Ajala (2018) further presented evidence that the use of extended format of a scale can aid in the control of extreme response style arising from careless responding.

Due to theoretical shortfall (e.g. loss of vital information due to dichotomization of the scale, Morren et al, 2009) and practical consideration (e.g. test-taker motivation, Cho, 2013), it has not been feasible to develop or integrate into the testing process new instrument to assess and enable the control of ERS. It was on this basis that a multiplicity of statistical technique has been employed in the measurement and control of ERS. Some of these advance statistical approaches in the study of ERS includes item response theory (1RT) models (Cho, 2013; Bolt & Johnson, 2009; De Jong et al, 2008), latent class factor analysis (LCFA, Moors, 2003; Morren et al, 2009), structural equation modelling approach (Cheung & Rensvold, 2000) and latent class transitional analysis

(Alchhlozcr, 2013).

Although these statistical techniques are robust and empirical evidence shows that they allow for robust assessment of ERS and other response style bias (see Van Vaerenbergh & Thomas, 2012 for review), modelling of ERs on the basis of item, respondent groups etc., they are quite difficult to execute and demand considerable expertise in computer programming. In addition, statistical programmes and software available for the execution of these models are usually in the commercial domain and exorbitantly expensive for researchers in a developing country like Nigeria where there is a paucity of funds for basic researches like ERS.

VII. Conclusion and Recommendations for Practice

Rating scales such as questionnaires and inventories have played critical roles in improving our understanding of the attitudinal, social, emotional and personality aspect of human behaviour, as well as aid in making projection and prediction about future performance. Despite this extensive utilization of this strand of research instruments, it is not without its flaws. Among the multiplicity of flaws rating scales are vulnerable to, these paper critically analyzed the problems of response bias under two broad types of social desirability response and extreme response style. Theoretical and practical considerations on the need to understand the problems of these response bias were identified as well as approaches suggested by scholars to measure and control their negative impact. Although the measurement and correction of these response bias may be cumbersome to some researchers, some relatively easy but powerful methods for minimizing its impacts were highlighted. For researchers with understanding of advanced statistical tool, these relatively novel area for research offers an exciting opportunity for employing such methods as CFA, IRT modelling, SFM etc.

Therefore, instrument developers and those involve in assessment for learning should acquaint themselves more with the some of the error and response biases outlined here and others not mentioned, as well as equip themselves with the knowledge for ameliorating their negative effect.

REFERENCES

[1]. Ajala, I. (2018). Comparative psychometric analysis of Likert and expanded scale formats using graded response model. University of Port Harcourt : Unpublished Masters Dissertation

[2]. Aichholzer, J. (2013) Intra-individual variation of extreme response style in mixed-mode panel studies. *Social Science Research*, 42, 957-970.

[3]. Austin, E. J., Deary, 1. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modeling of responses to NEO-FFI items. *Personality and Individual Differences, 40,* 1235-1245.

[4]. Babcock, J., Costa, D., Green, C. and Eckhardt, C. 2004. What situations induce intimate partner violence? A reliability and validity study of the Proximal Antecedents to Violent Episodes (PAVE) scale. *Journal of Family Psychology,* 18(3):433-442.

[5]. Bachman, J.G., O'Malley, P.M., & Freedman-Doan, P. (2010). *Response styles revisited: Racial/ethnic and gender differences in extreme responding* (Monitoring the Future Occasional Paper No. 72). Ann Arbor, MI: Institute of Social Research.

[6]. Baumgartner, I. F., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research,* 38, 143-156.

[7]. Baumgartner, F. I., & Steenkamp, J. B. E. M. (2005). Response Biases n Marketing Research. In *The Handbook of Market Research: Do's and Don'ts* (pp. 204-237).

[8]. Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied*Psychological Measurement, 33, 335-352.

[9]. Buckley, J. (2009). *Cross-national response styles in international educational assessments: evidence from pisa 2006.* Retrieved from*https://edsurvcys.rti.org/PISA/documents/BuckleyPlSAresponsestyle.pdf*

[10]. Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology,* 31, 187-212.

[11]. Cho, Y. (2013). The mixture distribution polytomous rasch model used to account for response styles on rating scales: A simulation study of parameter recovery and classification accuracy. University of Maryland, College Park. Unpublished Doctoral Thesis.

[12]. Cronbach, L. J. (1946). Response set and test validity. *Educational and Psychological Measurement, 6,* 475-494.

[13]. Chung, J. & Monroe, G. S. (2003). Exploring social desirability bias. *Journal of Business Ethics,* 44 (4), 291-302.

[14]. Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior and Personality, 15,* 137-152.

[15]. De Jong, M. G, Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research, 45,* 104-115.

[16]. DiFranceisco, W., McAuliffe, T. & Sikkema, K. (1998). Influences of survey instrument format and social desirability on the reliability of self-reported high risk sexual behaviour. *AIDS and Behaviour,* 2(4), 329-337.

[17]. Greenleaf, E. A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research,* 29, 176-188.

[18]. Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-contry study. *International Journal of Cross Cultural Psychology, 20,* 296-309.

[19]. Johnson, T. & Fendrich, M. (2002). *A Validation of the Crowne-Marlowe Social Desirability Scale.* Paper Presented at the American Association for Public Opinion Research, St Petersburg,FL. Retrieved fromhttp://www.srl.inc.edu/publist/conference/crownmarlowe.pdf

[20]. Kelly, N. L. (2015). Examining social desirability bias in measures of financial behaviour. *Retrieved from http://* ir.library.illinoisstate.edu/cgi/viewcontent.cgi?article=1463&context=etd

[21]. Kpolovie P. J. (2010). *Advanced research methods.* Owerri: Spring Field Publishers Ltd.

[22]. Krosnick, J. A., (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5, 213 236

[23]. Lalwani, A. K., Shavitt, S., Johnson, T., & Zhang, J. (2005). *What is the relation between culture and socially desirable responding?* Working paper, University of Illinois at Urbana-Champaign.

[24]. Lee. C., & Green, R. T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies,* 22, 289-305.

[25]. Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: ordinal/rating scale item responses. *Educational and Psychological Measurement, 70,* 5-21.

[26]. Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology, 23,* 498-509.

[27]. McCoach, D., Gable, R., & Madura, J. (2013). Instrument development in the affective domain. New York, NY: Springer.

[28]. Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44,* 1539 1550. doi: 10.1016/j.paid.2008.01.010.

[29]. Moors, G. (2012). The effect of response style bias on the measurement of transformational, transaction, and laissez-faire leadership. *European Journal of Work andOrganizational Psychology,* 21, 271-298.

[30]. Nwankwo, O. C. (2013). A practical guide to research writing. Port Harcourt: M and J Grand Orbit and Communication Limited

[31]. Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology,* 81 (6), 660-679.

[32]. Partnership for 21st Century Skill (2008). 21st century skills, education and competitiveness: A resource and policy guide. Retrieved fromhttp://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf

[33]. Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah NJ: Erlbaum. Retrieved from https://edsurveys.rti.org/PlSA/'documents/'Buckley_PISAresponstyle.pdf

[34]. Paulhus, D. L. (1991). Measurement and control of response bias. In.J. P. Robinson, P. R. Shaver, & L. S. Wright (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA:

Academic Press.

[35]. Schwatz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.).*The handbook of social psychology* (pp. 143-179), Vol. 1. Boston, MA: McGraw-Hill.

[36]. Van de Mortel, F. (2008). Faking it: social desirability response bias in self-report. Australian *Journal of Advanced Nursing, 25* (4), 40-48.

[37]. Van Herk, H., Poortinga, Y. H, & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from 6 EU countries. *Journal of Cross-Cultural Psychology, 35,* 346-360.

[38]. Van Vaerenbergh, Y & Thomas, T. D. (2012). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research* doi: 10.1093/ijpor/eds021.

[39]. Wetzel, L., Bohnke, J. R., Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Hiesu (Eds.), *The ITC International handbook of testing and assessment.* Oxford: Oxford University Press.

[40]. Wetzel. E., Carstensen, C. FL, & Bohnlce, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality, 47,* 178-189.