Social Desirability Bias in Likert Scale Design: Concerns and Suggestions

Quoc Tran - Nam*

Ho Chi Minh University of Banking, Vietnam
** Corresponding author. Email: quoctn@hub.edu.vn*

ABSTRACT
The Likert scale is used commonly by researchers to obtain data of a human's internal state. However, the Likert scale also faces the problem of bias (i.e., social desirability bias) in the respondent's response. Factors leading to the social desirability bias are wording style in the questions, the culture, and the heuristic process. As controlling the social desirability bias, statistical techniques, special techniques to elicit truthful answers from sensitive questions, or changing the designs of the questionnaires are applied. This review article covers information about the social desirability bias and suggests an idea to develop a time controlling method for the researchers to develop in future research.

Keywords: *Bias, Likert scale, Solutions, Social sesirability, Respondents*

1. INTRODUCTION

The Likert scale is used to obtain data from the participant's attitude and internal states (Edmondson, 2005; Likert, 1932). In this technique, the attitude of a person is defined as preferential ways of reacting to specific events or people. Likert scale offers help to quantify subjective thoughts, actions, and feelings in a reliable and validated way (Schwarz and Bohner, 2001). Although the Likert scale is convenient and helpful, researchers confront it with concerns such as the optimal number of the items in the scale (Joshi et al., 2015), its scales of measurement (Brown, 2011), culture differences (Lee et al., 2002), and method bias (Bernardi and Nash, 2023; Podsakoff et al., 2003; Westland, 2022). For instance, social desirability bias is one of the method biases affecting the reliability of the Likert scale data and results (Paulhus, 1991). In this paper, we discuss social desirability as a method bias by reviewing the evidence that bears upon it. We will close the discussion by suggesting a promising method to deal with the social desirability to the researchers who are considering using the Likert Scales.

2. DISCUSSIONS

Method bias is problematic when there are factors in the survey or task that: (a) make the task too difficult to accurately respond; (b) decrease respondent's motivation to answer accurately; (c) undermine the capabilities of the respondent (Podsakoff and Mackenzie, 2012). Consequently, method biases lead to errors in measurement both in random and systematic ways (Podsakoff et al., 2003; Westland, 2022). Precisely, the systematic error component seems to be a severe case where it provides an alternative explanation for the study rather than the true one and is independent of the hypothesis. According to Richardson et al. (2009), common method variance (CMV) is one of the main sources for the systematic measurement error in method bias. CMV is the variance accounting for the measurement method instead of the construct of interest. For example, methods effects refer to response biases such as social desirability, halo effects, acquiescence, leniency effects in the self-reported test (Podsakoff et al., 2003). The CMV could either increase or decrease the estimated relationship between variables of the research by becoming the third-party variable. Thus, researchers expect to control the CMV as much as possible so that the researchers can reflect the result as accurate as possible (Podsakoff et al., 2003).

Indeed, while doing the questionnaire it is essential to acknowledge the subjective components from the participants affecting their response to control the CMV (Podsakoff et al., 2003). Social desirability (SD) has been found to be one of response styles bias (Bernardi and Nash, 2023; Paulhus, 1991). SD refers to the tendency of some people to respond to items more because of their social acceptability but not their true feelings (Podsakoff et al, 2003). For example, participants with SD would agree to attend voting, go to the church (i.e., activities that is more socially acceptable) and disagree with things such as tattooing, substance abuse (i.e., activities that is not socially acceptable). Also, participants with SD bias would underreport their undesirable behaviour and overreport their desirable

behaviours, or attitudes. Participants might think the SD answer makes them look good as it fits the cultural norms of appropriate traits, attitudes, opinions, behaviours, and the desirability of a specific thing (Fisher, 1993). It is more likely for a person to report positively about themselves in comparison to when they have to report about others (Jurgensen; 1978). It suggests that their perceptions about themselves are more socially desirable than their peers. However, the respondents did not consciously realize the contradiction.

Paulhus (1984) proposes two concepts of SD which are impression management (i.e., conscious control of one's public image in favour of society's norms), and self-deception enhancement (i.e., an unconscious tendency to portray oneself positively but honestly believed self-description). Paulhus's concept of SD has been supported by a neuroimaging study that shows people with SD engage in neural activity in the medial prefrontal cortex and left ventrolateral prefrontal cortex (Farrow, Burgees, Wilkinson, & Hunter, 2015). Evidence from this study also shows that there is a dissociation between impression management and self-deception in SD. Therefore, SD is a real phenomenon in the survey that researchers must take it seriously. This bias can create false effects or obscure relationships between variables (i.e., SD suppresses or increases or moderates the relationship) and change the mean of the answers (Bernardi and Nash, 2023; Fisher, 1993; King & Bruner, 2000; Tan et al., 2021). For instance, if we do questionnaires with entrepreneurs about their competencies, we must consider that they might overestimate their competencies due to social desirability. It will be problematic if the same entrepreneurs are asked to do evaluations of their company's performance. In such a case, social desirability can be a factor affecting data of both the competency questionnaire and the company's performance questionnaire. Therefore, SD is problematic to the researcher while obtaining and analysing data.

Acknowledging the problems that SD brings about, raises the question "why does SD happen?". It suggests that the questionnaire design might activate social desirability (Foddy and Foddy, 1994; Lietz, 2009). Firstly, the wording design in the questionnaire can cause a question threat to the respondent. Studies show that questionnaires misreporting on sensitive topics would motivate a process in which participants edit the information to avoid making their image bad, the reaction from the third party, or due to self-deception (Holtgrave, 2004; Tourangeau and Yan, 2007). The sensitive question means that it is a threat of disclosure and intrusiveness to the respondents while answering it in a survey (Tourangeau et al., 2000). For example, questions about religion or income or the risk of giving truthful information and opinion about the third party. The risk of them might be a broken relationship or job loss. The sensitivity also means the extent to which the question might elicit an undesirable/desirable answer from the respondents (i.e., when a question asks for a socially undesirable behaviour or attitude from respondents, the respondents might have to admit that they have violated the social norm). Factors are affecting responses to sensitive questions such as mode of administration (i.e., face-to-face; telephone interviews, and mail-survey), third-party presence, interview settings, wording (Tourangeau and Yan, 2007). Consequently, with those factors, respondents tend to adopt SD answers rather than their truths. Thus, survey makers should be more cautious in selecting appropriate data collecting questions that reduce respondents' discomfort to obtain more valid data.

Culture plays such a role in affecting the SD bias in survey responses (Johnson and Van de Vijver, 2003). Because the culture decides whether the behaviour or thought is appropriate, it is crucial to consider the cultural variations while conducting a survey. Indeed, it is difficult to generalize the findings from the cross-cultural questionnaires. Thus, the social desirability bias rate could vary between the studies across cultures using the same method of research. For example, a question expects to have a similar level of SD content among respondents from different cultural groups. However, some people from a certain background might be more likely to edit their answers than others. Specifically, collectivist cultures are more likely to give misleading information than individualistic cultures (Triandis, 1995). Another factor that can lead to SD bias in answering the survey is religiousness (Rallapalli, 1994; Chung and Monroe, 2003). A more religious person would report stronger beliefs and idealism compared to a less religious person (Rallapalli, 1994). Idealism is related to the desire to impress others favourably. Such a need to make a favourable impression reflects the social desirability of a person. Therefore, SD could be affected by religiousness.

Also, a socially desirable answer is created due to heuristic processing (Holgraves, 2004). He proposes that respondents use heuristic in the retrieval stage. He/she would only retrieve positive information and neglect the negative information. Therefore, the information is biased in the first place. However, a study shows that SD operates in the editing stage (i.e., the last stage in the survey responding process; Holtgraves, 2004; Podsakoff, 2003) rather than the retrieval stage. Specifically, when the participants face the statement in the questionnaires item, they might have to retrieve information from their long-term information. After that, they need to evaluate if that piece of information is socially appropriate or not. If it does not meet the social norm then they will edit that piece of information in the response. With this kind of process, participants take longer to respond to the question as they need to edit it before responding. Much empirical evidence has been found to support this theory. Holtgrave (2004) proposes 3 experiments to study how and when the SD operates during the survey responding process. The results from the 3 experiments show that it takes a longer time for participants in enhanced SD conditions to respond to

survey questions than the controlled group. Thus, this reflects that SD happens during editing time rather than retrieval.

Attempts have been made to control social desirability. Firstly, measurement conditions effect on SD could be diminished by anonymity guarantee from researchers, assuring respondents that there were right or wrong answers, or telling respondents that other people have very varied answers about the issued in the questionnaires. This method is believed to reduce the editing rate in respondent's answers (Podsakoff et al., 2003).

Secondly, researchers often use strategies for eliciting sensitive information (i.e., specialized questioning techniques and the bogus pipeline) and reducing SD bias (Tourangeau and Yan, 2007; Cerri et al., 2021). Specialized Questioning Technique (SQTs) functions to encourage truthful answers by creating a design to protect respondent's privacy (Cerri et al., 2021). There are multiple techniques in the SQTs, however, the most well-known one is the randomized response technique (Cerri et al., 2021). A meta-analysis examined the effectiveness of the randomized response technique (RRT) compared with standard face-to-face interviewing in obtaining the same data. RRT method uses some noise to mask the answers from the respondent with a known probability distribution (Nuno and St John, 2015). Specifically, noise may come from non-sensitive questions or the randomizing device. The more amount of noise, the higher the privacy protection level. RRT method is efficient in reducing the underreport of undesirable behaviours relatively to a face-to-face interview. However, the RRT makes it difficult to identify the relationship between the characteristic of the respondents and the sensitive behaviour as RRT only generates estimations rather than individual scores (Tourangeau and Yan, 2007). Furthermore, RRT requires experienced interviewers to successfully implement this method (Coutts and Jann, 2011). Overall, this method brought about both advantages and disadvantages to the researcher. Another method from the SQTs is the parallel model (Tian, 2014). The parallel model is not a randomized technique; however, it combines a sensitive question of interest and 2 non-sensitive questions with known prevalence. Tian (2014) assures that the parallel model is more statistically reliable, higher privacy protection. Nevertheless, this technique is applied once in the sexual habit (Tian et al., 2019), thus, it does not show validity and generalization to the other conditions.

Another method used to obtain sensitive data is a bogus pipeline (Tourangeau and Yan, 2007). The principle of the bogus pipeline is that the respondents believe that interviewers would find the respondent's truth on the variable in the questions regardless of any contents he or she reports (Roese and Jamieson, 1993). With the bogus pipeline method, the researchers would find any means to convince the respondents that they can detect lies and false reports such as graph-like data (e.g., heart rates, brain waves, etc; Tourangeau, Smith, & Rasinski, 1997). This method is believed to increase reports of socially undesirable attitudes, therefore social desirability would decrease. This method requires special machines to operate, thus, it is difficult for researchers to do the questionnaires on wide scales.

Regarding the verbal and language issue in the survey, item characteristic effect which is also known as item social desirability could be taken into consideration (Podsakoff et al., 2003). Item social desirability refers to the issue that items in the survey may be written in the way that they reflect more or less social desirability behaviours, attitudes, or perceptions. As such, these items could relate to the person's SD (Podsakoff et al., 2003). Thus, this suggests a method of eliminating item social desirability by rewording or replacing (Nederhof, 1985). Until now, there is no significant disadvantage of using this method apart from the fact that this method is sometimes undesirable to researchers to vary the scales formats and anchors (Podsakoff et al., 2003). However, changing the scale anchors might lead to sacrificing the scale validity and meanings. Researchers should be careful while using this method.

Lastly, there are statistical remedies for social desirability apart from the controlling techniques mentioned above. For example, partial correlation procedures are designed to control social desirability. Specifically, researchers measure these variables directly and then they partial their effects out of the criterion and predictor variables (Podsakoff et al., 2003; Bernardi and Nash, 2023). This method is straightforward to carry out. Researchers only need to obtain the data for the social desirability in advance and then compare the differences between the criterion and predictors variables in the partial correlation with zero-order correlation (Spector et al., 2000). Despite the benefits, the limitation of this method is that it does not differentiate the measure of a construct from the construct itself. This method is supposed to have effects at the item level rather than the construct level. Therefore, it is difficult for researchers to study the relative impact of these two types of effects (Williams et al., 1996). Other statistical remedies to control the social desirability are (1) multiple method factors; (2) controlling for the effects of an unmeasured, latent, methods-factors; (3) controlling for the effects of a directly measured, latent, methods-factors (Mishra, 2016). Of these, directly measured, latent, methods-factors is the more suitable technique than others for controlling and estimation of social desirability with the use of confirmation factor analysis (Mishra, 2016). Confirmation factors analysis has various advantages regarding the trait variance, error variance (scale specific biases and random error) and variance due to the method used etc (Mishra, 2016). However, the associated disadvantage of CFA is still considered as it might have an absence of a valid scale measuring the cause of method variance. Techniques and

statistical remedies both have advantages and disadvantages. Thus, it requires researchers to wisely choose the appropriate method.

A study from Holtgraves (2004) and Tan et al. (2021) suggests that SD does not always affect the particular answer but the responding time. Thus, it raises the question of whether the SD effects might be decreased or even eliminated if there are time constraints during questionnaires? However, the time pressure method should be carefully considered as giving such a shortage of time (i.e. 2 seconds or 10 seconds) might lead to higher social desirability bias in responses (Sutherland, 1964). Whilst this paper suggests that future research can try to control different points of time when the respondent starts to read the question until they answer it. If researchers can find a reliable effect of time on the SD, they will be able to develop a technique to control the SD better.

In conclusion, the social desirability responding problem has been a problem to the researchers using the Likert scale. Social desirability can be caused by sensitive questions, heuristic processing, and culture. There are multiple ways to deal with social desirability such as statistical techniques, special techniques to elicit truthful answers from sensitive questions, or changing the designs of the questionnaires. However, there exists disadvantages to each of the techniques. Thus, the authors suggest an idea for finding a solution for the social desirability of the respondent by using appropriate time pressure while the participants answer the questions. Future research can take this suggestion to develop a method using the time to elicit truthful answers from the respondents. This will contribute to improving the final data obtained by the other researchers by decreasing or eliminating the effect of the social desirability bias.

REFERENCES

Bernardi, R. A., & Nash, J. (2023). The importance and efficacy of controlling for social desirability response bias. Ethics & Behavior, 33(5), 413-429. https://doi.org/10.1080/10508422.2022.2093201

Brown, J. D. (2011). Likert items and scales of measurement. Statistics, 15(1), 10-14.

Cerri, J., Davis, E. O., Verissimo, D., & Glikman, J. A. (2021). Specialized questioning techniques and their use in conversation: A review of available tools, with a focus on methodological advances. Biological Conversation, 257, 109089. http://dx.doi.org/10.1016/j.biocon.2021.109089

Chung, J. & Monroe, G. S. (2003). Exploring social desirability bias. Journal of Business Ethics, 44, 291-302. https://doi.org/10.1023/A:1023648703356.

Coutts, E. & Jann, B. (2011). Sensitive questions in online survey: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Sociological Methods & Research, 40(1), 169-193. https://psycnet.apa.org/doi/10.1177/0049124110390768

Edmondson, D. (2005, May). Likert scales: A history. In Proceedings of the conference on historical analysis and research in marketing (Vol. 12, pp. 127-133).

Farrow, T. F., Burgess, J., Wilkinson, I. D., & Hunter, M. D. (2015). Neural correlates of self-deception and impression-management. Neuropsychologia, 67, 159-174. DOI: 10.1016/j.neuropsychologia.2014.12.016

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. Journal of Consumer Research, 20(2), 303-315. https://doi.org/10.1086/209351

Foddy, W. & Foddy, W. H. (1994). Constructing questions for interviews and questionnaires: Theory and practice in social research. Cambridge university press. https://doi.org/10.1017/CBO9780511518201

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

Johnson, T. P., & Van de Vijver, F. J. (2003). Social desirability in cross-cultural research. Cross-cultural survey methods, 325, 195-204.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. Current Journal of Applied Science and Technology, 396-403. https://doi.org/10.9734/BJAST/2015/14975

Jurgensen, C. E. (1978). Job preferences (What makes a job good or bad?). Journal of Applied psychology, 63(3), 267. https://psycnet.apa.org/doi/10.1037/0021-9010.63.3.267

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. Psychology & Marketing, 17(2), 79-103. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2%3C79::AID-MAR2%3E3.0.CO;2-0

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. Research in Nursing and Health, 25(4), 295-306. https://doi.org/10.1002/nur.10041

Lietz, P. (2010). Research into questionnaire design: A summary of the literature. International journal of market research, 52(2), 249-272. https://doi.org/10.2501/S147078530920120X

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140, 5-53.

Mishra, M. (2016). Confirmatory Factor Analysis (CFA) as an analytical technique to assess measurement error in survey research: A review. Paradigm, 20(2), 97-112. https://doi.org/10.1177/0971890716672933

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. European journal of social psychology, 15(3), 263-280. https://doi.org/10.1002/ejsp.2420150303

Nuno, A., & St John, F. A. (2015). How to ask sensitive questions in conversation: a review of specialized questioning techniques. Biological Conversation, 189, 5-15. http://dx.doi.org/10.1016/j.biocon.2014.09.047

Paulhus, D. L. (1984). Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 46(3), 598. http://dx.doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. Annual review of psychology, 63, 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Rallapalli, K. C. (1994). Rallapalli, K. C. (1994). Marketers' norms: An empirical investigation of the influence of personal and situational factors on marketers' deontological norms, ethical judgments and intentions. The University of Mississippi.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. Organizational Research Methods, 12(4), 762-800. https://doi.org/10.1177/1094428109332834

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. Psychological bulletin, 114(2), 363. http://dx.doi.org/10.1037/0033-2909.114.2.363

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. Blackwell handbook of social psychology: Intraindividual processes, 1, 436-457. http://dx.doi.org/10.1002/9780470998519.ch20

Spector, P. E., Chen, P. Y., & O'Connell, B. J. (2000). A longitudinal study of relations between job stressors and job strains while controlling for prior negative affectivity and strains. Journal of Applied psychology, 85(2), 211. http://dx.doi.org/10.1037//0021-9010.85.2.211

Sutherland, B. V. (1964). Social desirability, item-response time, and item significance. Journal of Consulting Psychology, 28(5), 447. http://dx.doi.org/10.1037/h0047898

Tan, H. C., Ho, J. A., Teoh, G. C., & Ng, S. I. (2021). Is social desirability bias important for effective ethics research? A review of literature. Asian Journal of Business Ethics, 10(2), 205-243. https://link.springer.com/article/10.1007/s13520-021-00128-9

Tian, G. L. (2014). A new non-randomized response model: the parallel model. Statistica Neerlandica, 68(4), 293-323. http://dx.doi.org/10.1111/stan.12034

Tian, G. L., Liu, Y., & Tang, M. L. (2019). Logistic regression analysis of non-randomized response data collected by the parallel model in sensitive survey. Australian & New Zealand Journal of Statistics, 61(2), 134-151. http://dx.doi.org/10.1111/anzs.12258

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. Psychological bulletin, 133(5), 859. http://dx.doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R., Rips, L. J., & Rasinski, K. (Eds.). (2000). The psychology of survey response. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Tourangeau, R., Smith, T. W., & Rasinski, K. A. (1997). Motivation to report sensitive behaviors on surveys: Evidence from a bogus pipeline experiment 1. Journal of Applied Social Psychology, 27(3), 209-222. http://dx.doi.org/10.1111/j.1559-1816.1997.tb00629.x

Triandis, H. C. (1995). A theoretical framework for the study of diversity. In M. M. Chemers, S. Oskamp, & M. A. Costanzo (Eds.), Diversity in organizations: New perspectives for a changing workplace (pp. 11–36). Sage Publications, Inc. https://doi.org/10.4135/9781452243405.n2

Westland, J. C. (2022). Information loss and bias in likert survey responses. PloS one, 17(7), e0271949. http://dx.doi.org/10.1371/journal.pone.0271949

Williams, L. J., Gavin, M. B., & Williams, M. L. (1996). Measurement and nonmeasurement processes with negative affectivity and employee attitudes. Journal of Applied Psychology, 81, 88-101. http://dx.doi.org/10.1037/0021-9010.81.1.88