

Self-reports and Observer Reports as Data Generation Methods: An Assessment of Issues of Both Methods

Mitchell Abernethy

University of Guelph, Canada

Copyright © 2015 Horizon Research Publishing All rights reserved.

Abstract Data collection through the self-report method can allow one to acquire both a different type and quality of information when compared to acquiring the information through the observer report method; however nor the observation or self-report processes are perfect data collection methods. When designing a study it is important to know when to use one over the other. By understanding how the data is collected in both of these processes and the problems associated with either process one can make an informed decision as to when to use one over the other. Additionally, though the use of only one method can be a practical choice the use of both methods is the better choice and should be used when it can. It should be noted that there are many problems and biases that need to be controlled for in these methods and there are other ways in which they can be controlled for beyond what is covered, however, this paper will be limited to the stated biases and control methods. Throughout my involvement in ‘The Fire Starter Study’ I have come to understand the process of how these methods are conducted and why they are conducted the way they are.

Keywords Self-report Method, Social Desirability Response Bias

Research Internship - Psych*3900

1. Introduction

Data collection is an integral component of research. Without raw data later inference would be impossible or at the very least incomplete. Currently, Psychological research use of a variety of data collection methods with each serving a certain purpose. Knowing the type of knowledge you wish to extract will influence the data you wish to collect as well as the method of data collection. Two broad data collection methods within Psychology include the self-report method and the observer report method. These methods are able to attain a certain type and quality of information not obtained through the other. Furthermore, both methods are hold limitations as well as strengths and thus nor the observation

or self-report processes are perfect data collection methods. When designing a study it is important to know when to use one over the other. By understanding how the data is collected in both of these methods as well as the problems associated with either process one can make an informed decision as to when to use one over the other. Additionally, the use of a single method can be a practical choice however the use of both methods should be used when available. It should be noted that there are many problems and biases that need to be controlled for in these methods and there are other ways in which they can be controlled for beyond what is covered, however, this paper will be limited to the stated biases and control methods. Throughout my involvement in ‘The Fire Starter Study’ I have come to understand the process of how these methods are conducted and why they are conducted the way they are.

2. Self-reports as a Data Generation Method

The most common method of data collection in psychology is the self-report method [20]. This method was first implemented under the idea that “no one knows you like your self” and has been proven beneficial at accessing certain types of information only available to the individual such as intensions, motivations and past experiences [16] leading to its continued use in studies today. However, due to certain problems found in this method, such as the social desirability response bias and the respondents’ ability to recall past information at the time of survey, some have questioned its use when researching areas that ask respondents for personal information. Such fields of study include assessing personality disorders and job performance. Due to these concerns methods have been developed to minimize the effects of these biases in hopes of increasing both the reliability and validity of this method. These methods include the Marlowe-Crowne Social Desirability Scale (MCSDS) and the use of multiple data sources.

Social Desirability Response Bias

One of the well-known problems associated with

self-reports is the social desirability response bias. Social desirability bias is a response bias that refers to the tendency for respondents to present themselves in a way that is untrue and projects a favorable image to the researcher [7]. Some believe this is a conscious effort. Holtgraves [7] found a positive relationship between the degree to which one is concerned about the impression their response presents to researchers and the level of consideration taken when presenting their answers; suggesting that respondents, when asked a question, selectively retrieve information which will present them in a positive light to the researcher. A study conducted by Adams, Soumerai, Lomas and Ross-Degnam [1] give support to the idea that this is a conscious decision. This study analyzed self-report bias in adherence to guidelines in the workplace. It was found that, “the increasing reliance on self-reports as a measure of quality of care appears to produce gross overestimations of performance,” attributing this “gross overestimation” to the, “social pressures that may promote socially desirable responses that do not reflect actual practices,” [1 p.190]. Some examples of such social pressures include fear of inadequacy, malpractice litigation and an over branching fear of losing one’s job. When participants are aware of such pressures they give the socially desirable answer in order to avoid the possible punishments associated with giving an accurate account of their behavior.

Because of this researchers need to be aware of such a factor in order to attain accurate information from their respondents. There are ways to quantify this bias such as the MCSDS and the use of this scale in tandem with a focal scale has been proven effective at controlling for this bias, as it will be discussed later on.

Errors in Memory

Not only can information given by a respondent on a self-report measure be distorted voluntarily, such was the case for social desirability bias, self-reported information can be distorted involuntarily through memory errors. This should be taken into account when assessing the validity of retrospective data.

Benefits of surveying a respondent’s past is that it allows researchers to trace, in a relatively fast and inexpensive way compared to longitudinal studies, long distances of time [13]. However some believe that memory is reconstructions of past events and this process is not immune to error [23]. Hence memory errors can be seen as the inability to properly reconstruct a memory. This may result in the creation and presentation of an artificial sequence of events when one is asked to recall events from their past [13]. As a result researchers should take into consideration the time elapsed between the event and the time of questioning when asking respondents to report on past events. Additionally, even when respondents are trying to accurately report on information from their past, this self-report is still subject to inaccuracy [20]. When applying this knowledge to ‘The Fire Starter Study’ it is easily understood that by having the

participants fill out the self-report survey in between trials, as opposed to reporting on all trials when the study was completed, was to decrease possible memory errors that may have resulted from having the respondents completing the self-report fill out at a later time.

In summary, by inquiring into a respondents’ past via the use of self-reports researchers are able to save resources such as time and money when compared to a conventional longitudinal study. However, doing this exposes your study to the possibility of collecting inaccurate data since respondents are not immune from making memory errors. Therefore the quality of your data becomes dependent on the respondent’s ability to correctly recall information from their past.

Marlowe-Crowne Social Desirability Scale

Though self-report surveys can be distorted consciously and unconsciously it is still a widely used method to attain personal information. Thus the data received and the claims made with this data should be done with caution. What is not debated however is that distortions are a real possibility and therefore, must be taken into account in the construction of questionnaires.

To help control for the social desirability response bias researchers make use of the Marlowe-Crowne Social Desirability Scale. By the early 1960s, more than a dozen scales had been developed to measure social desirability [29]. The scale contains a set of 33 true-false items which describe acceptable yet improbable and unacceptable yet probable behaviors [10]. Those who present themselves as having the desirable yet improbable traits and deny having the unacceptable yet probable trait are identified as someone who would respond in a socially desirable manor [27]. The scales ability to identify when respondents give socially desirable answers allows it to be a good comparison tool for other studies. For example, if the focal scale and the MCSDS have a low correlation then the researchers would conclude that the respondents’ scores on their scale are not biased in a socially desirable manor [12]. A study conducted by Johnson and Fredrich [10] testing the ability of the MCSDS found that, “respondents who under-reported cocaine use (i.e., tested positive but reported no use in the past year) scored higher, on average, on the CM [CMSDS] scale.” [10 p.1663] supporting the claim that the MCSDS’s is capable of identifying respondents who give the socially desirable answer.

To summarize, the Marlowe-Crowne Social Desirability Scale is capable of identifying respondents who respond in a socially desirable manor. By comparing the scores on this tool to a focal survey, researchers are capable of controlling for the social desirability response bias. Therefore the MCSDS should be used in self-reports when the probability for respondents to respond in a social desirability manor is high.

Multisource Method

A respondent's introspective ability can distort the results of the data obtained through self-reports and therefore should be controlled for in studies. To increase a study's accuracy and minimize the rate of false positives researchers can make use of the multisource method [2]. The multisource method makes use of information attained through more than one source. For example, when assessing personality effects on job performance in a workplace, researchers will draw upon information attained not only through self-reports but peers, subordinates and supervisors who have contact with that individual being assessed. It should be noted that these additional sources are a form of observer reports, a topic which will be expanded upon later in this paper.

A study conducted by Bohl [3] found that the use of multiple informants when assessing job performance in the workplace has superior results when compared to the more traditional method of information attained through only the supervisor. He supports this claim with statistics finding that 55% of respondents who use traditional systems which collect information only through the supervisor, compared to the 68% of respondents who use the 360 degree assessments which makes use of multiple sources of information, believe that their system make a difference when measuring performance results. By having multiple sources of information, researchers have an increased ability of attaining a more comprehensive account than what would have been covered by a single report [18]. Other benefits to using multiple sources methodologies include combating problems of introspection and self-report bias [19] and reducing the impact of biases and random errors [28]. The use of multisource methodologies are also found useful in the field of child psychiatric epidemiology where it is known that young children tend to provide unreliable information on their cognitive state [9].

In summary there are problems with using the self-report method as a sole study method. Through the use of multisource methodologies researchers can attain a more comprehensive account of events than using a self-report [18], increase a study's accuracy and minimize the rate of false positives [2], combating problems of introspection and self-report bias [19] and reduce the impact of biases and random errors [28]. This had led to its advocated use in job performance assessments and psychiatric epidemiology and has many applications in other fields. Therefore the multisource method is a useful tool in scientific research.

Self-report Summary

The self-report method consists of an individual divulging personal information, whether that be thoughts, feelings, basic information or occurrences in their past, to a researcher who will use it for later analysis [16]. This method was founded on the idea that no one knows you like yourself and has been proven useful in many fields of study such as job performance, personality and psychiatric assessment. Information obtained through this method should be taken with caution when asking for peoples personal opinions and other personal information due to the tendency for

respondents to respond in a way which adheres to what is socially desirable instead of responding in a true and honest fashion [7]. This socially desirable responding can be identified using the Marlowe-Crowne Social Desirability Scale that, when used in tandem with the study's focal survey, can help control for respondents who respond in a socially desirable manor [10]. Additionally, even when a respondent wishes to respond in an accurate manor memory errors may consequently produce an artificial sequence of episodes, leading to a distortion in the results [13]. In order to attain a more comprehensive account of events than using the self-report method alone, researchers may gather information not only from the individual but friends, family and peers of that individual [18]. Because of the self-reports ability to collect personal (internal) information from many people in a relatively quick fashion and its ability it is no wonder this method has come to be the most common method for data collection [20].

3. Observer Reports as a Data Generation Method

As previously noted another form of data collection is the observer report method. The collection of data through observations is a widely used method used by researchers to attain information, in both organizational and scientific fields, which they may not be able to receive through self-reports. There are, however, problems found within this method. Some of these problems include problems associated with inter-rater reliability and observer expectancy bias. In order to control for these problems researchers can make use of multiple coders as well as blind studies.

Observer reports have been found to attain different information than self-report studies; a difference resulting from the information available to those in either perspective. The Observers perspective is denied access to information such as motives, intention and past behaviors of an individual however its utility is based in attaining information pertaining to the performance and personality of those being observed (Strauss, 1994). Though observer reports don't consider the internal processes, it is a valid form of research. It has been used to predict of job performance in an organizational setting. A study conducted Mount, Barrick and Strauss found that, "each observer perspective accounted for significant variance beyond self-ratings for both conscientiousness and extraversion," [16 p. 277] when assessing observer ratings and self-report measures of personality factors that predict job performance. This point is echoed in a study published in the Journal of Applied Psychology assessing the construct validity of self and peer evaluations. It was found that, "peer nominations tap unique information relative to self and assessor evaluations," concluding, "peers may in fact be the most able to predict subsequent job advancement," [26 p.50]. Therefore even though observer reports may not be able to

attain the type of information that self-reports can about an individual, observer reports have found to be valid research method for assessing performance and personality.

Inter-rater reliability and multiple observers

In addition to an organizational workplace observer reports are used in lab oriented settings, such was the case in 'The Fire Starter Study'. Those observing watched videos of participants doing a task then recorded the observed behaviour into a form to be later used for statistical analysis. This process is known as "coding"; and is done through the use of "coders". It is common for studies that make use of coders to have multiple people coding in order to increase statistical power.

It is possible for those coding to record the observed in a way different from one another so it is important to know if those coding did so in a same or similar enough fashion. To assess this, a statistical analysis must occur to calculate inter-rater reliability. Inter-rater reliability quantifies the degree of agreement between two or more coders who make independent judgments by determining how much of the variance in the observed score is due to variance in the true scores after the variance due to measurement error between coders has been removed [6]. This is not an uncommon practice in scientific research. "Many research designs require the assessment of inter-rater reliability (IRR) to demonstrate consistency among observational ratings provided by multiple coders," [6 p.1].

As it is seen, the use of multiple observers in studies is done to increase the study's statistical power. To avoid coder bias inter-rater reliability needs to be calculated to assess the degree of agreement among the raters. This process is an important control that should be implemented in observational studies that make use of multiple coders to ensuring that those coding are doing so in a similar enough manner and can therefore control for the coders' (observer's) individual biases.

Observer Expectancy effects

The coding of behavior in 'The Fire Starter Study' was done through prerecorded videos of the participants performing the task. This allowed for the controlling of possible biases that result from the interactions between those involved in the study and the participants.

A journal entry published by the Association for Psychological Science on experimenter effects and the demand characteristic found that, "experimenters demonstrably exert a variety of different, unwarranted effects on the outcome of their experiments, thus exhibiting a form of unconscious bias," [11 p.573]. Furthermore, a book by Folger and Greenburg [5] about the controversial issues in social research echoed this point when they state, "It appears experimenters can influence subject behaviours in ways so subtle (e.g. paralinguistic and kinesics cues) that the experimenters themselves may be unaware of the bias introduced by their own expectations," [5 p.130]. This type of bias is understood as experimenter bias and can be

demonstrated through the observer expectancy effect. The observer expectancy effect, as defined by the American Psychological Association, is the effect in which the researcher's belief or expectations unconsciously affects the behaviour of those which are being observed [4].

Robert Rosenthal [24] has demonstrated this effect in a laboratory setting. In his study he had two conditions. In the first condition, the maze-bright condition, a group of psychology students were told the rats they were working with were bred for intelligence and could solve a maze quickly. In the second condition, the maze-dull condition, a group of psychology students were told the rats they were working with were bred for dullness and were slow at maze solving. It was found that, "the students who had been assigned the maze-bright rats reported significantly faster learning times than those reported by the students with the maze-dull rats," [24 p.1]. In reality, the rats were standard lab rats and had not been bred for either condition. The students had not intentionally slanted the results and what was concluded from this study was that experimenters can unintentionally influence their expectancies onto their test subjects [24].

Since observer expectancy effects are apparent in studies where those observing are aware of the test condition, it should be taken into account when designing an observational study. To mitigate this bias the use of a blind procedure, which withholds critical information from those involved and/or participating in a study, can be implemented.

Blind Procedures

Since Scott (the experimenter conducting the study) was in contact with the participants and was aware of the hypotheses of the study and the conditions in which the participants were in, it could be argued that he was not immune from experimenter expectancy bias and furthermore could have affected participant behaviour unknowingly. I however, as a coder for the study, was not given the participants' condition and had no interaction with the participants allowing me to not be affected by possible experimenter expectancy effects. Additionally the participants were unaware of their condition. Being that information was withheld from those participating in the study it can be seen that a blind procedure was in place.

There are three types of blind procedures: triple blind procedures, double blind procedures and single blind procedures; all of which try to control for expectancy effects. Blind procedures work by hiding critical information from certain people in the study (i.e. participants and/or researchers conducting the study) and the degree to which those people know this critical information indicates whether it is a triple, double or single blind study.

In a single blind experiment information that could introduce bias is withheld from the participant but the experimenter and observers are in possession of all information [25]. This type of experiment is still open to the observer expectancy bias since the experimenter is aware of the hypotheses of the study and the conditions in which the

participants are in. However the benefit to this method is that hypotheses are withheld from the participant [14]. It is understood that if the manipulation of the experiment is known by the participant they may alter their behaviour in accordance with the desired response therefore biasing the data [25]. Such an idea prompted the use of blind studies as they were implemented when researchers came to understand the role that participant knowledge of the study's purpose or hypotheses may play in influencing the outcome, [11]. This effect of the participant knowing key information about the manipulations of the study and acting upon this information is known as the participant expectancy bias. Therefore by implementing the single blind study one is capable of control for the participant expectancy bias but not the observer expectancy bias. This was the blind study used in the 'Fire Starter Study'.

The double blind method, in addition to withholding critical information from the participant, withholds critical information the researchers and those who interact with the participants and is therefore capable of controlling not only for participant expectancy effects but the observer expectancy bias. If this procedure is capable of controlling for both biases and achieve a higher standard of scientific rigor [14], why was it not implemented in 'The Fire Starter Study'? Simply, this procedure was avoided because it involves a more complex procedure than the single-blind method and even though those coding the participant behaviour were withheld information about the participants' condition, it was still easily identifiable. So implementing a double blind experiment may have worked in theory however it would mislead the readers on the internal validity of the study conducted. Additionally when the single-blind study is objective enough to withstand any additional biasing effects, then running a double blind study would be considered excessive and recourse consuming [25]. Nonetheless, when the potential for experimenter bias is high, then the double-blind method should be used.

Being that Scott, the experimenter conducting 'The Fire Starter Study', kept his interactions with participants to minimal level and the possibility of experimenter bias was low, implementing the single-blind study was the proper choice in order to not mislead the reader to the study's internal validity and because of its not as complex design, resource effective procedure, and its ability to control for participant expectancy effects.

Observer Report Summary

To summarize, though observers may not be able to identify the internal process such as goals, intentions and past experiences of whom they observe, observer reports have proven to be an effective tool for assessing personality and performance (Strauss, 1994). This has sanctioned its use in job performance assessment and scientific research however there are problems that need to be addressed in this method. When observers interact with their subjects unwarranted effects have been found to occur which distort the outcome of their experiments, such was demonstrated by

Rosenthal [24]. This has led to the use of blind studies, which withhold sensitive information from those conducting and/or participating in a study, to mitigate such a bias [11]. There is advocated use for multiple observers to code a study however when multiple observers are used, it is possible for those observing to report what they have observed differently from one another meaning inter-rater reliability can also pose a problem for observational studies [6]. Thus inter-rater reliability should be calculated in observational studies to assess to what degree the observers are reporting the same information. So it can be seen observer reports are a valid research method and though there may be problem in this method, as long as the proper procedures are in place, researchers will continue its use.

4. Final Conclusions

To conclude, both the self-report method and the observer report method are valid collection methods used by researchers however the self-report method is more common as a result of its effective use of resources. Each method is uniquely engineered to obtain a certain type and quality of information and thus lack the ability to acquire all types of data. Having a depth of understanding pertaining to the problems associated with either method as well as how to control for such problems researchers maintain the ability to make an informed decisions as to when to use one over the other. Furthermore, it should be seen that though the use of only one method can be a practical choice, the use of both self-report and observer reports should be used in tandem. I do believe my participation in 'The Fire Starter Study' has helped me come to understand these methods on a higher level then I previously did by leaning, in depth, the problems which need to be accounted for, how to control for those problems and the process of when and how to conduct either method.

REFERENCES

- [1] Adams AS, Soumerai SB, Lomas J, Ross-Degnan D. (1999). Evidence of self-report bias in assessing adherence to guidelines. *Int J Qual Health Care*. 1999 Jun;11(3):187-92.
- [2] Blackman, G.L., Ostrander, R., Herman, K.C. (2005). Children with ADHD and Depression: A Multisource, Multimethod Assessment of Clinical, Social, and Academic Functioning. *Journal of Attention Disorders* 8: 195.
- [3] Bohl, D. L. (1996). Mini Survey: 360 Degree Appraisals Yield Superior Results Survey Shows. *Compensation and Benefits Review*, 16.
- [4] Gerrig, R.J., Zimbardo, P.G. (2002). *Psychology and Life*, 16/e. Pearson Education. <http://www.apa.org/research/action/glossary.aspx>.
- [5] Greenburg, J., Folger, R. (1988). Ch.7 Experimenter Bias.

Springer Series in Social Psychology, (pp. 121-137).

- [6] Hallgren, K.A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psycho* 8(1): 23-34.
- [7] Holtgraves, T. (2004). Social Desirability and Self Reports: Testing Models of Social Desirable Responding. Society for personality and Social Psychology. *PSPB, Vol. 30 No. 2, February 2004* 161-172.
- [8] Hoskin, R. (2012). The Dangers of Self-Reports. *British Science Association*.
<http://www.sciencebrainwaves.com/uncategorized/the-dangers-of-self-report>.
- [9] Horton, N.J., Laird, N.M., Zahner, G.E.P. (1999). Use of Multiple Informant Data as a Predictor in Psychiatric Epidemiology. *International Journal of Methods in Psychiatric Research*.
- [10] Johnson, P.T., Fendrich, M. (2002). A Validation of the Marlowe-Crowne Social Desirability Scale. *American Association for Public Opinion Research - Section on Survey Research Methods*.
- [11] Klein, O., Doyen, S., Leys, C., Magalhães de Saldanha da Gama, P.A., Miller, S., Questienne, L., Cleeremans, A. (2012). Low Hopes High Expectations: Expectancy Effects and the Replication of Behavioral Experiments. *Perspectives on Psychological Science* 2012 7: 572.
- [12] Leite, W.L., Beretvas, N. S. (2004). Validation of Scores on the Marlowe- Crowne Social Desirability Scale and the Balanced Inventory of Desirability Responding. *Educational and Psychological Measurement* 2005 65: 140.
- [13] Mazoni, A., Vermunt, J.K., Luijckx, R., Muffles, R. (2010). Memory Bias in Retrospectively Collected Employment Careers: A Methodological Approach to Correct for Memory Error. *Sociological Methodology, vol. 40 no. 1* 39-73.
- [14] McShane, J.J. (2010). Blind and Open Experiments and Observer Bias. <http://www.thetruthaboutforensicscience.com>.
- [15] Miller, J.D., Pilkonis, P.A., Morse, J.Q. (2004). Five Factor Model Prototypes for Personality Disorders: The Utility of Self-Reports and Observer Ratings. *Assessment*. 11(2):127-38.
- [16] Mount, M.K., Barrick, M.R., Strauss, P.J. (1994). Validity of Observer Ratings of the Big Five Personality Factors. *Journal of Applied Psychology, Vol 79(2), Apr 1994*, 272-280.
- [17] Nerman, G.R., Streiner, D.L. (2008). Biostatistics: The Essentials. *B.C. Decker Inc*.
- [18] Okada, M., Oltmanns, T.F. (2009). Comparison of Three Self-Report Measures of Personality Pathology. *J Psychopathol Behav Assess*, 31(4): 358-367.
- [19] Pappas, J., Madison, J. (2013). Multisource Feedback for STEM Students Improves Academic Performance. *American Society for Engineering Education*.
- [20] Paulhus, D.L., Vazire, S. (2007). Ch.13 Self-report Method. *Handbook of Research Methods in Personality Psychology (pp.224-239)*.
- [21] Perry, J.C. (1992). Problems and Considerations in the Valid Assessment of Personality Disorders. *American Journal of Psychiatry*, 149:1645-1653.
- [22] Razavi, S. (2007). The Political and Social Economy of Care in a Development Context: Conceptual Issues, Research Questions and Policy Options. *United Nations Research Institute for Social Development*.
- [23] Roediger, H. L., & DeSoto, K. A. (in press). The Psychology of Reconstructive Memory. In J. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences*, 2e. Oxford, UK: Elsevier.
- [24] Rosenthal, R., SL Jacobson, L. (1966). What You Expect is What You Get. *Teachers' expectancies: Determinates of pupils' IQ gains. Psychological Reports*, 19, 115-118.
- [25] Salkind, N., Stewart-Hamilton, I. (2010). Single Blind Study. *Encyclopedia of Research Design*.
- [26] Shore, T.H., Shore, L.M., Thorton III, G.C. (1992). Construct Validity of Self and Peer Evaluations of Performance Dimensions in an Assessment Center. *Journal of Applied Psychology, Vol. 77, No. 1*, 42-54.
- [27] Tatman, A. W., Swogger, M.T., Love, K., Cook, M.D. (2002). Psychometric Properties of the Marlowe-Crowne Social Desirability Scale with Adult Male Sexual Offenders. *Bulletin*, 27, 151-161.
- [28] Wagner, S.M., Rau, C., Lindemann, E. (2010). Multiple Informant Methodology: A Critical Review and Recommendations. *Sociological Methods and Research*, 38: 582-618.
- [29] Wiggins, J.S. (1964). Convergences Among Stylistic Response Measures From Objective Personality Tests. *Educational and Psychological Measurement*, 24, 551-562.