



# Measuring Social Desirability Bias With Person-Items Map Analysis: Incorporating SEM and Person-Items Map Analysis Part 4

Hermansyah Andi Wibowo<sup>1\*</sup>  and Nafiudin Nafiudin<sup>1</sup> 

<sup>1</sup> Universitas Serang Raya, Kota Serang 42162, Indonesia

hermansyahandiwibowo@gmail.com\*, nafiuddin08@gmail.com

**Abstract.** In present day, the issue of social desirability bias (SDB) has only acted as a savior explanation when behavioral research results are considered anomalous. The well-known SDB scale, namely MCSD, has 33 items that are impractical to include SDB as a variable in the proposed model. Therefore, we try to offer an alternative way to measure SDB other than the existing ones. Novelty: This research provides evidence and formulates a way to measure SDB using the person-items map (PIM). Research Methods: We conducted a simulation of SDB calculations with PIM analysis on 367 Muslim women regarding their intentions to wear the hijab. This sample came from published SEM-based research that produced anomalous results. Findings: We found that respondents overestimated their religiosity to an impossible level. Statistically, an indication of high values can be seen from the comparison of M (mean), S (standard deviation) and T (2 times the standard deviation) values between respondents' abilities and the level of difficulty of the questions. There is serendipity in the form of considerations for determining the type of selected independent variables used as the location for measuring SDB. Conclusion: This article shows that SDB measurement can be done using PIM. Although it does not use the popular SDB scale, the results are still reliable because they meet strict requirements.

**Keywords:** Social Desirability Bias, Person-Items Map, Religiosity

## 1 Introduction

### 1.1 Background

This research is part of the big theme "the importance of incorporating Person-Items Map (PIM) analysis to Structural Equation Modeling (SEM) research result" in an effort to enrich the understanding of research results. In the current work, the topic of Social Desirability Bias (SDB) was chosen as the entry point to the big theme. Why? Because SDB infects research involving ethical/social variables in survey design research [1]. In fact, if ignored, SDB can cause the validity of research results to be seriously injured. Ganster, Hennessey, & Luthans[2] found the types of distortion

© The Author(s) 2025

W. Paramita et al. (eds.), *Proceedings of the 12th Gadjah Mada International Conference on Economics and Business (GAMAICEB 2024)*, Advances in Economics, Business and Management Research 319, [https://doi.org/10.2991/978-94-6463-692-5\\_16](https://doi.org/10.2991/978-94-6463-692-5_16)

by SDB on organizational behavior research are, 1) Spuriousness, misleading evaluation of entire measurement, 2) Suppression, masking the relationships between variables, 3) Moderation, altering the relationships between other variables. Next, we will show why SDB can be mitigated with this idea of combining methods.

Many behavioral studies do not mitigate the occurrence of SDB but instead use it as a way out when research results are "less than" satisfactory. While satisfactory empirical findings do not guarantee the absence of SDB in research, existing publication articles in 2008 [3], 2012 [4], 2012 [5], 2018 [6], 2022 [7], and 2023 [8], rarely or even not specifically describe mitigation done to reduce or eliminate SDB, rather than focusing the discussion on its satisfactory research results. This is most likely due to the impracticality of the commonly known SDB scale, namely the Marlow-Crowne Social Desirability Bias/MCSD (1960)[9] which consists of 33 measurement items; or it could also be due to the researcher's ignorance of the impact of SDB on research results. Many studies have successfully summarized or created shorter versions of the MCSD (nine of them analyzed by Barger (2002) [10] but there is no consensus among researchers regarding which scale is best. In addition, various contexts, including field of study/locus of research, prevailing culture, and study population, provide varying results of the short version of the MCSD scale.

## **1.2 Problem statement**

The severe impact of SDB in research, the failure to resolve the impracticality of MCSD, the contextuality of short-form MCSD performance, and the fact that researchers do not include SDB issues in research, of course raise doubts about the results of ethical/social-based research. Therefore, a breakthrough is needed that makes it easier for researchers to mitigate the occurrence of SDB in their research. Before entering into the research questions, we assume SDB as a variable whose essence is inherent in ethical/social variables. Therefore, we view the person-items map (PIM) as being able to determine the response patterns of research samples related to variables that are suspected of having the potential to be responded to with SDB.

## **1.3 Research question**

Based on a number of phenomena that have been explained, the assumptions used, and for the purpose of operationalizing into research, we formulate the following research questions. "Can PIM be used as a tool to reveal SDB in research variables?"

## **1.4 Objectives**

This study aims to demonstrate the measurement of SDB with the assumption that ethical/social variables have the potential to be over-responded. Specifically, to measure this overresponse, we chose a PIM with person ability properties (left side) separated by logit bars, and items with different difficulty levels (right side), to reveal this over-response pattern.

### 1.5 Method and analysis

The research methodology employed is laboratory simulation, as described by Buckley, Buckley, & Chiang (1976)[12], which has been tailored to suit the specific objectives of this study. Overall, this study falls under the quantitative genre and aims to reinforce the notion of "the importance of incorporating PIM analysis to SEM research result" to enhance comprehension of research findings. Arguments are formulated to determine which variables are potentially over-responsive indicating SDB. Conditions for understanding research findings are described to observe differences before and after treatment.

### 1.6 Conclusion

Based on PIM, it can be ascertained that the research respondents reported their condition higher than the normal condition. PIM successfully revealed this and finally it can be concluded that SDB has occurred in the religiosity variable. Wibowo & Masitoh's [6] confusion regarding the results of their research (high religiosity but attitudes and intentions are not related), can be resolved by conducting PIM analysis on the selected ethical/social variable, namely religiosity.

## 2 Literature review

### 2.1 Short history of the MCSD scale

Self-reporting is a widely employed technique for gathering data in various fields of behavioral science research due to its convenience, user-friendliness, and effectiveness in measuring constructs that cannot be directly observed.[13]. It is often the only practical option available when other methods are not feasible[14]. This method involves participants responding to questions that may reveal details about their personalities, attitudes, values, lifestyles, beliefs, sentiments, and behavior[15]. This method relies on the premise that respondents provide truthful responses. Nevertheless, the method frequently faces criticism due to the susceptibility of self-reports to different types of response biases. These biases include deviant responding, careless responding, consistent responding, omitting items, acquiescence (the tendency to agree), extremity bias (the tendency to use extreme ratings), and socially desirable responding [16][17]. These biases ultimately diminish the quality of the data.

Among the biases found in self-report surveys, SDB has received the most attention because of its obscure and intricate nature[18]. SDB on the other hand, represents systematic measurement error caused by the need for social approval. The presence of SDB threatens the validity of ethics research finding [19]. Therefore, the Marlowe–Crowne Scales were developed to specifically measure social desirability, defined as “a person’s need to gain approval by responding in culturally appropriate and acceptable ways” (Crowne & Marlowe, 1960)[9]. Crown and Marlowe selected a set of items from a number of personality inventories that were thought to represent the domains of

SDB, with the caveat that item responses do not reflect psychopathology or disorder[10].

Nevertheless, Edwards (1957)[21] is credited as the original creator of the SDB concept, as he expressed his fascination with the issue of social desirability in 1951, purely by chance, in his book. Additionally, he presents the SDB scale in the same book. Nevertheless, the MCSD scale was extensively embraced and remains widely used in the present day. Crowne and Marlowe meticulously curated collections of questions from various personality surveys that they deemed to be a faithful representation of that particular domains[10]. Nevertheless, they underscored that the participants' replies to the items did not indicate any indications of psychopathology or abnormal conduct.

Crowne and Marlowe's distinguished the new scale from e.g., Edwards' SDB Scale (1957) that contained maladaptation questions and whose validation involved clinical diagnostic criteria. These things exemplify behavior that is deemed suitable, but it is deemed improbable that the behavior will actually be executed (e.g., I consistently verify the safety of my car prior to embarking on a lengthy journey; Prior to casting my vote, I meticulously investigate the qualities of all candidates). Hence, Crowne & Marlowe (1960) stated that main objective of this scale is to offer a quantifiable assessment of social desirability that excludes any content connected to diseases. We found the pronunciation MCSD is used more often even though based on the author's order it should be CMSD.

## 2.2 Research about the MCSD scale

We grouped research related to SDB measurement and MCSD scales into 4 research objectives. First, researchers conducted empirical research on SDB measurements. Second, researchers offer ways to mitigate SDB. Third, because of the impracticality of the MCSD, researchers offer abbreviated versions of the MCSD scale or developments thereof. Fourth, researchers examined the psychometric properties of the MCSD scale

**Empirical researchs of SDB measurements.** Many studies in a variety of contexts – locations, methods and subjects– find the presence of SDB in research. The existence of SDB is empirically proven in the context of negative alcohol-related behavior[22] [23][24], [25], [26] terror-induced political participation[27], reported work but not announced [28], cases of obesity [29]. Sociodemographically, elderly subjects, Indian residents, people without income, and those who are/are still married, are more likely to experience SDB[30]; age and gender were determining factors for SDB[31] and Education level, job position and survey method have a significant effect on SDB, whereas age, gender and ethnicity have no effect [19]. The polarization of the Republican Party vs. the Democratic Party regarding the immigration policy of building a barrier appears lower than it actually is because of SDB[32]. SDB ruled out anti-bias based disciplinary decisions on Black students (punitive vs. rehabilitative)[33]. SDB bias and marriage cynicism have a moderating effect on reported purchase intentions. Older, well-traveled generations appear to be particularly susceptible to bias, and such bias negatively im-

pacts overall survey results[34][35]. A successful list experiment revealed 37% conventional financing or twice the actual amount in a sample of poor Muslims in Pakistan [36][37]. However, SDB has not been proven to exist in terms of antenatal care [38].

**How to mitigate or deal with SDB in research.** Ganster et al.(1983)[2] found the types of distortion caused by SDB in organizational behavior research are 1) Spuriousness, misleading evaluation of the entire measurement, 2) Suppression, masking the relationships between variables, 3) Moderation, altering the relationships between other variables. Therefore, researchers must use a mixture of prevention, minimization, measurement, and adjustment techniques [15]. Krumpal (2013)[39] stated that in addition to psychological factors (social approval and avoiding embarrassing interactions), SDB is also influenced by aspects of survey design (eg. anxiety about sensitive statements, or online survey[19] ), interviewer characteristics, and survey situations. Researchers have attempted to address SDB using three different approaches: administering a social desirability scale along with a primary interest scale, reducing social desirability responses by manipulating testing conditions, and using statistical methods to correct scores[40]. However, these approaches increase respondent burden or are costly to researchers in terms of research design or analysis complexity. Therefore, low-cost methods are needed to address SDB.

**Short form of MCSD or developments thereof.** Comparisons made between the short forms examined in this investigation suggest the 13-item form as a viable substitute for the regular 33-item Marlowe-Crowne scale [41]. Some, but not all, of the short forms considered by Reynolds were indeed appreciably better than random. [42]. Confirmatory Factor Analysis (CFA) was employed to establish the adequacy of subscales in measuring social desirability [43]. The original CrowneMarlow(CM)-scale consists of 33 items. Since this is a rather large instrument for a general survey most authors have used a reduced version of the CM scale [44]. Hays, Hayashi, and Stewart Hays, Hayashi, & Stewart (1989) proposed a 5-item scale to measure SDB named SDRS-5 for clinical investigators needing a very brief measure of socially desirable response set. Also 10 item Norwegian short-form of the Marlowe-Crowne Scale was produced.[46]. Some research suggested that Clancy and Gove (1974) short version performs best. But none of their phantom questions or any combination of them is able to pick up social desirability. Instead over-claiming is associated with a lack of knowledge (Franzen, et al., 2019) [44]. But none of these concise scales are robust enough in different contexts.

**Research on psychometric properties of the MCSD scale.** Psychometric analysis is lenses used to assess the quality and meaning of psychological test results. The techniques used are divided into item analysis which aims to measure the quality of each item in the test; reliability analysis which aims to measure the consistency and stability of test scores; validity analysis aimed at assessing whether the test measures what it is supposed to measure; factor analysis aimed at identifying latent structures in tests and grouping items based on similar characteristics; Lastly, norm analysis which aims to compare individual scores with reference group scores.

Currently there are two approaches to psychological measurement, namely Classical Test Theory (CTT) and Item Response Theory (IRT). Both are used to analyze test scores, but they differ in their approach. CTT focuses on the test itself and the total score achieved by an examinee while IRT focuses on the individual items within the test and the underlying ability (latent trait) of the examinee. CTT assumes the standard error for each examinee is the same, whereas the standard error in IRT depends on individual ability, which is partly reflected in the theta ( $\theta$ ) value. IRT consists of model 1 Logistic/PL parameter (item difficulty level) eg. the Rasch model, 2 PL model (difficulty level and item discrimination power), and 3 PL model (item difficulty level, item discrimination, guess element). The following is research related to the psychometric properties of the SDB scale.

Reynolds (1982)[47] used Factor Analysis (FA) and concluded that the 13-item form was a viable substitute for the traditional 33-item Marlowe-Crowne scale. Paulhus (1984)[48] added CFA and recommended that impression management, not self-deception, be controlled for in self-reports. Some researchers have concluded that further psychometric evaluation of the MCSD is needed (Beretvas et al., 2002)[49]. The MCSD scale lacks factor structure and internal consistency reliability ([50]; [51];[52]). Barger (2002) employ CFA on 9 short forms of the MCSD and found that there was evidence of adequacy of the short form of the MCSD, but inconsistent modeling across samples.

However, from the results of our literature review, SDB is a construct that has high contextuality. Why? Because of SDB originates from the standard values and norms that apply in the physical environment and mental condition of the research subjects. A number of contradictory results regarding the demographics of respondents who contributed to SDB such as Teh et al. (2023)[30] vs. Fraboni & Cooper (1989)[31], the psychometric properties of the 33-item MCSDS were sometimes better [53], [54] or worse [43], [55], [56] from its development version. The MCSDS has cross-country applicability [57] and [58] and researchers remain convinced that scale adaptation is necessary. The MCSDS is unidimensional [59]; [60] but is also suspected of measuring other constructs (Fernandez et al., 2019).

Finally, SDB itself can be seen as a separate construct so that a scale is created, but there are also researchers who question the measurability of SDB (Ganster et al., 1983) In this paper, the authors present an SDB measurement technique derived from the worldview/pragmatism paradigm (Creswell & Creswell, 2023[61]) in business research. We propose this technique with humility. This paradigm focuses on providing practical solutions to the challenges encountered by researchers in the business world (the motivation behind my research). This innovative technique is entirely novel, thereby facilitating open discourse. Next, allow me to provide a brief explanation of the tools we utilize at a later point.

### 2.3 Person-Items Map

Person-items map (PIM) is a feature of the Rasch model. This Rasch model is a new measurement paradigm used in the field of marketing (Salzberger & Koller, 2013[62]). The Rasch is a precise example of IRT 1PL. Within the broader framework of measuring latent variables in other fields, this assertion remains valid. For almost six decades, latent variables such as attitudes, opinions, and perceptions have been evaluated based

on the assumption that their values can be calculated as interval data types and exist on a continuous scale. Despite the use of a Likert scale, the measurements obtained still produce ordinal data. However, we often disregard this and mistakenly categorize the data as interval data. Hence, the difference between strongly disagree and agree (2 minus 1) is equal to the difference between agree and strongly disagree (5 minus 4). The Rasch model can be utilized to return data to its initial state (Sumintono & Widhiarso, 2016). For example, when human sentiments, which naturally exist on a continuous spectrum, are converted into discrete values using Likert scales, continuous probabilities are utilized.

The Rasch model has the ability to recover data accurately based on its original conditions (Sumintono & Widhiarso, 2016)[63]. For example, human attitudes naturally exist on a continuum but when measured using a Likert scale they produce ordinal data. To restore the quality of the data, it is converted into an odds ratio. The odds ratio formula in Rasch is defined as the comparison between the probability (P) of a participant answering correctly and the probability of a participant answering incorrectly. Mathematically, the odds ratio is expressed by a formula.

$$\text{Odds ratio} = P/(1-P) \quad (1)$$

After obtaining the OR value, the value is then transformed into a logit scale using the natural logarithm:

$$\text{Logit} = \ln(\text{OR}) \quad (2)$$

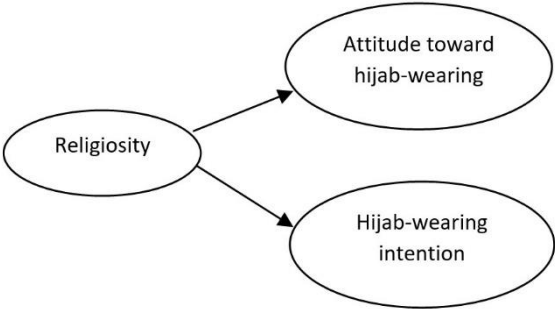
Where Logit is logarithm odds unit and ln is the natural logarithm.

The logit scale has several advantages over the odds ratio. The logit scale has equal intervals, so that differences between individual logits or items have consistent interpretation across all levels of ability and difficulty. This makes it easier to compare the test taker's abilities and the difficulty of the items. Rasch modeling will create a hierarchical relationship between respondents and the items used [64]. The following are the assumptions of the Rasch model [65]. 1) each person is characterized by an ability, 2) and each item by a difficulty that, 3) can be expressed by numbers along one line, 4) from the difference between the numbers (and nothing else), the probability of observing any particular scored response can be computed.

### 3 Method and analysis

#### 3.1 Research sample

Our sample is the same sample as Wibowo & Masitoh research as an implementation of incorporating dual quantitative modeling. This sample is collected by purposive sampling and produced 367 valid respondents from Serang City, Serang Regency, Cilegon City, Pandeglang Regency, Lebak Regency, Tangerang City, Tangerang Regency and South Tangerang City. The main analysis tool is SEM with the addition of the Latent Variable Score (LVS) technique from Joreskog [66] which is used to simplify the religiosity variable from 22 indicators to 5 indicators originating from 5 dimensions. Data was collected using a questionnaire as the main instrument, observations were made via offline (Tangerang area mall) and online (facebook.com), and a literature review. Analysis by Wibowo & Masitoh (2018)[6] produced a verified model which is the main material for this research.



**Figure 1** Main material that provides study material (religiosity variable)

The religiosity variable has 3 indicators, namely basic central duties, experience, and knowledge, which used to be dimensions. Because there are no theoretical changes related to the construct of religiosity other than increasing measurement detail via the LVS technique when calculating SEM, for the purposes of SDB detection, the three indicators of the religiosity variable are further dissected into dimensions containing the following items.

**Table 1.** Items of religiosity variable.

| Codes | Items                                                           | Dimensions   |
|-------|-----------------------------------------------------------------|--------------|
| R4    | I always start my actions with bismillah                        | Central duty |
| R7    | I pray five times a day                                         | Central duty |
| R8    | I'm always fasting on Ramadan                                   | Central duty |
| R12   | I always pay zakat fitrah                                       | Experience   |
| R13   | I once felt that I was given guidance when there was a problem. | Experience   |
| R14   | I have felt blessed                                             | Experience   |
| R17   | I know the Qur'an in general                                    | Knowledge    |
| R18   | I know the life and actions of the Prophet                      | Knowledge    |
| R19   | I often rean the Qur'an                                         | Knowledge    |
| R20   | I once cried because I read the Qor'an                          | Knowledge    |

The **figure 1** shows that attitudes and behavior (intention) are not related, and both are directly influenced by religiosity. This finding was quite surprising at the time, but Wibowo and Masitoh (2018) tried to resolve it by explaining the characteristics of the sample.

#### 4.1 Demography

All respondents were female and based on World Health Organization, 68.05% of them were young women whose ages ranged from 17 to 24 years old. This age group is characterized by: transactional orientation; seeking an identity; emotionally developing; and have a high imagination. Our additional interview confirmed this characterization. Women of this age are still looking for fun rather than conforming to their religion. There is a need to do some further research to obtain more reasonable explanations.

**Figure 2** Screenshot of Wibowo & Masitoh (2018) respondents demography

The prevailing understanding is the conclusion of Wibowo & Masitoh who assessed the sample's religiosity as low. This can be seen from the sentence "Women of this age are still looking for pleasure rather than following their religion". In fact, the standard average value for each religiosity item in the model is 4.35. **This is contradictory and requires further explanation.**

The phenomenon that occurs is that the characteristics of the study are closely related to a person's self-image. Where this self-image is related to the values and norms that apply in the physical and mental situations of the respondent's daily life. The religiosity scale is also very sensitive in describing the respondent's profile. Religiosity, attitudes and behaviors regarding the hijab, are three variables that can put someone in an unpleasant psychological situation. The authors suspect SDB occurred in my research with Martina in 2018. At that time we only focused on testing the model without carrying out the necessary SDB mitigation.

Technically, the simplification of the model with LVS by Wibowo & Masitoh (2018) is suspected of being able to reduce the quality of the variation of indicators formed due to dimension reduction. However, this suspicion was refuted based on the results of research by Wibowo & Indarti (2020)[67] which found that the grouping of research indicators in CFA and PIM visualizations was similar. Furthermore, in this article religiosity is the variable chosen because of the position of the model as an independent variable. There is ample evidence to suggest that those with stronger religious views, including business practitioners and consumers, tend to have more prominent ethical norms and judgments compared to individuals with weaker religious beliefs [48], while attitudes and intentions to wear the hijab are consequences of religiosity.

As an object of measurement, religiosity is measured by creating items that indicate its existence in the research subject. The items in table 1 have the potential to make respondents less comfortable in reporting their condition. Therefore, this premise will be tested by making a quantification using the IRT approach that considers a person's ability as seen in their response, with the level of difficulty of the items that will be agreed to measure the quality of a person's religiosity.

### 3.2 Rasch model stages

The analysis stages using Rasch modeling which will later be visualized by PIM, are shown in the following.

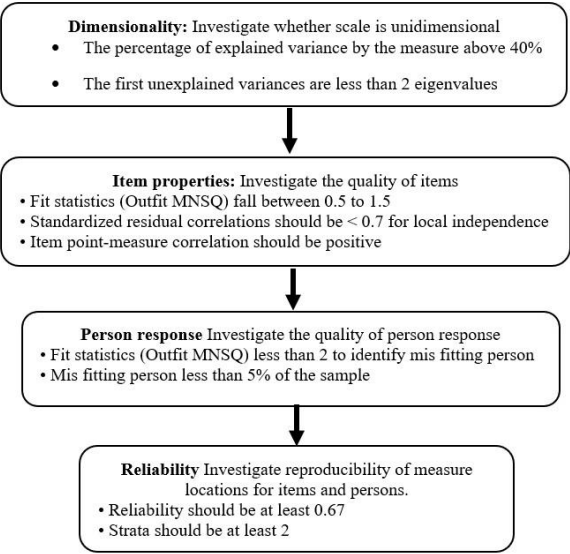


Figure 3 Rasch model stages

### 3.3 Analysis with PIM

In simple terms, this research carries out Rasch model analysis using the PIM feature on published research results. PIM displays logit measure of abilities and item difficulty by put it in a ruler. The respondent's factual abilities are reflected in his/her responses to the religiosity scale items. PIM analysis is carried out after the Rasch modeling stage is carried out.

## 4 Results and discussion

### 4.1 Dimensionality

Dimensionality analysis examined whether all scales are one-dimensional. Principal Component Analysis is utilized to assess the unidimensionality of an auction by analyzing Rasch measures and residuals. A scale is considered unidimensional when the measure can explain more than 40% of the variance and the first unexplained variance is less than 2 eigenvalues. (Linacre, 2012)[68]. In this study, the first condition was met, namely raw variance explained by measures 44.6%, while for the second condition the eigenvalues obtained were 0.1 greater, namely 2.1.

## 4.2 Item properties

Item properties analysis evaluates the characteristics and attributes of items to assess their quality. An item is deemed to have a good fit if the fit statistics for outfit MNSQ are between the range of 0.5 to 1.5. Standardized residual correlations indicate the presence of item local dependency. If the correlation is more than 0.7, it means that two items share more than half of their random variation, and only one of them needs to be kept [69]. Furthermore, the item point-measure correlation index should have a positive value, demonstrating the absence of polarity. Negative correlations suggest that there may be errors in the coding of reverse-coded items, while correlations close to zero suggest that the items are either very simple or difficult to agree with, or that they are measuring a different concept altogether. A correlation coefficient below 0.4 might be utilized to highlight items that require further examination regarding their phrasing. [70]

| ENTRY  | TOTAL  | TOTAL |         | MODEL | INFIT | OUTFIT | PT-MEASURE | EXACT MATCH |       |      |      |      |      |
|--------|--------|-------|---------|-------|-------|--------|------------|-------------|-------|------|------|------|------|
| NUMBER | SCORE  | COUNT | MEASURE | S.E.  | MNSQ  | ZSTD   | MNSQ       | ZSTD        | CORR. | EXP. | OBS% | EXP% | Item |
| 4      | 1694   | 367   | -1.50   | .12   | .97   | -.5    | 1.55       | 3.8         | A .48 | .53  | 77.8 | 74.1 | R12  |
| 1      | 1620   | 367   | -.46    | .12   | 1.16  | 2.2    | 1.16       | 1.8         | B .55 | .58  | 66.9 | 70.9 | R4   |
| 10     | 1500   | 367   | 1.07    | .11   | 1.15  | 1.9    | 1.16       | 1.9         | C .61 | .62  | 70.8 | 70.9 | R20  |
| 2      | 1595   | 367   | -.13    | .11   | 1.14  | 1.9    | 1.12       | 1.4         | D .60 | .59  | 68.5 | 71.1 | R7   |
| 8      | 1467   | 367   | 1.48    | .11   | 1.01  | .1     | 1.00       | .1          | E .57 | .63  | 72.5 | 70.7 | R18  |
| 7      | 1438   | 367   | 1.83    | .11   | .97   | -.3    | 1.01       | .1          | e .60 | .64  | 77.0 | 69.9 | R17  |
| 3      | 1693   | 367   | -1.48   | .12   | .95   | -.6    | .85        | -1.2        | d .58 | .53  | 77.5 | 74.0 | R8   |
| 9      | 1520   | 367   | .82     | .11   | .86   | -1.9   | .85        | -1.9        | c .66 | .62  | 76.7 | 71.1 | R19  |
| 6      | 1651   | 367   | -.88    | .12   | .84   | -2.5   | .83        | -1.8        | b .60 | .56  | 78.1 | 71.4 | R14  |
| 5      | 1642   | 367   | -.76    | .12   | .82   | -2.7   | .81        | -2.2        | a .61 | .57  | 77.5 | 71.2 | R13  |
| MEAN   | 1582.0 | 367.0 | .00     | .12   | .99   | -.2    | 1.03       | .2          |       |      | 74.3 | 71.5 |      |
| S.D.   | 89.1   | .0    | 1.16    | .00   | .12   | 1.7    | .22        | 1.9         |       |      | 4.1  | 1.3  |      |

**Figure 4** Screenshot of Winstep 3.73 output

In this study, item R12 has an MNSQ value outside the upper limit of tolerance, namely 1.55. However, other fit measurements are not violated. Based on the point-measure correlation index, all items have a positive value which indicates the absence of polarization. In aggregate, the items measuring religiosity perform quite well.

## 4.3 Person responses

Person response analysis evaluates the proficiency of an individual's answer. Person response was assessed using person fit statistics. [71] suggested utilizing an outfit MNSQ value of 2 as a criterion for detecting misfits, with an expected occurrence rate of fewer than 5% in the sample. Individuals that do not fit in with the norm may suggest that the self-reported data is of poor quality, possibly due to thoughtless or misunderstood responses to the questions (Kottorp, et al., 2003). In this study, the first condition was met but the numbers did not meet the second condition. The authors put the winstep calculation results table in the attachment. Winstep also informed that there were 50 respondents who gave unexpected responses. There are 3 possibilities when this happens:

- 1) Test takers may misunderstand the test items. This can happen because the test items are written ambiguously or because the test takers do not have enough knowledge to understand the test items.
- 2) Test takers may deliberately give wrong answers. This can happen because the test taker is trying to lie or because they don't want to spend time answering the test items.
- 3) Test takers may have different opinions about the test items. This can happen because test takers have different backgrounds or experiences from other test takers.

4.4 Reliability

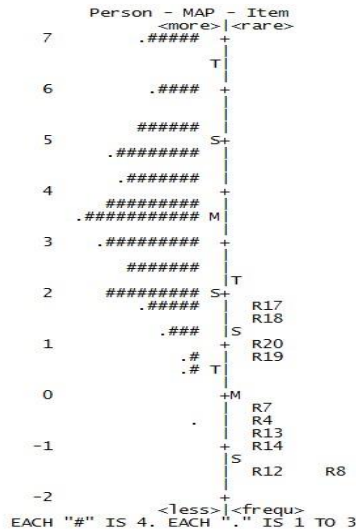
Summary statistics are summaries that show the quality of respondents, the quality of items, and also the quality of the results of the interaction between the two.

Table 2 . Reliability.

| No | Groups      | Reliability  | Strata |
|----|-------------|--------------|--------|
| 1  | Respondents | 0.73         | 2.55   |
| 2  | Items       | 0.99         | 13.33  |
| 3  | Both        | 0.80 (KR-20) |        |

From table 2 above, we can conclude that research respondents answered consistently. The quality of the items used to measure has reliable performance. Overall, the internal consistency value of the interaction between respondents and items is .8 and this is also very good. The second reliability requirement, namely strata, is also met, that is, it must be above 2. The authors have put the complete statistical summary table of Winstep's calculation results in the attachment. At this point, it is still unclear why some respondents gave unexpected responses.

4.5 Person-items map



**Figure 5.** Screenshot of PIM taken from Winsteps output

In Figure 5 above, it is very clear that the average of respondent response (M left side) is far above the average level of item difficulty (M right side). The majority of respondents rated themselves as very positive regarding the items in the statement measuring religiosity. For the most difficult item to agree on, R17 (I know the Qur'an in general), remains below the standard deviation value (S on the left side) of the respondent's response. In fact, there were 3 respondents (codes 085g4e, 116g8e, 317e6i) whose logit measure value was negative (their ability was below the average level of item difficulty).

Regarding the 50 respondents who gave unexpected responses, it is very possible that the cause was that the respondents deliberately gave the wrong answers. Respondents were trying to lie or because they do not want to spend time answering test items. Respondents who deliberately answered incorrectly were not because they didn't want to lose their time. But it is deliberately covering up the real conditions. Why? because the responses they gave varied. This means they read and easily understand the meaning of the items measuring religiosity. Respondents who don't want to waste their time will fill in quickly without reading and choose only 1 alternative response. There were only 11 people who gave a maximum score of 5 (strongly agree) and 36 people chose agree or a score of 4. Again, this shows that the 47 people who filled in this quickly rated themselves very positively.

Based on the univariate response pattern, we believe that the number of respondents who filled in seriously was 320 people (367 varying scores minus 47 scores with a variance of 0). Both those who deliberately filled it in and those who lied, both reported high levels of religiosity to me and Martina in 2018. If they read carefully, then the variance will not be zero (0). Why? Because the items in this study were of very high quality based on a high reliability score (.99). It seems that the respondent delib-

erately lied to maintain her social image as a devout Muslim woman. Observation results at [www.facebook.com](http://www.facebook.com) and mails support this explanation. (the authors used to search via typing responden email on Facebook desktop)

#### 4.6 Understanding after PIM

Based on the results of the previous PIM analysis, we can confirm that Wibowo & Martina [6] research respondents, responded very positively to religiosity indicators. The respondent's ability compared to the level of item difficulty is in a very unequal position. Research respondents rated their own religiosity very high. Thus, the results of the study by Wibowo & Masitoh (2018) which verified the influence of religiosity on attitudes and intentions to wear the hijab and there was no relationship between attitudes and intention to wear the hijab, were explained by them that the cause of the loss of this relationship was the lack of religious practice of Muslim teenage girls. In our research, the explanation for the loss of the relationship between attitudes and intentions to wear the hijab was caused by Muslim teenage girls overestimating their religiosity. This is evidenced by the results of the PIM analysis. We believe there is SDB in the study by Wibowo & Martina (2018) because there is a suppression effect (covering up the relationship between variables, Ganster et al., 1983)

#### 4.7 Suggestions for mitigating SDB

This section is made to meet the reproducibility requirements, we propose steps for replicating the mitigation method against SDB.

1. Determine variables that have SDB potential based on their equivalence to the SDB scale properties. You should study existing SDB scales, we recommend Crowne & Marlowe's (1960) 33-item MCSD scale.
2. Study in depth, just 1 variable that you suspect will be over-responded for imaging purposes by the respondent.
3. Develop instruments that cannot be perceived a priori by respondents. Also make sure that respondents are not disturbed by the presence of data collectors.
4. Before you further execute the research, perform PIM on the selected SDB variables. If there is a high potential for SDB to occur then eliminate this variable.
5. If you don't want to eliminate it, then we suggest mitigating it in other ways. SDB should not be ignored.

It would be a good idea for you to study IRT with its supporting software, so that you can eliminate respondents who are infected with SDB

## 5 Results and discussion

SDB is an important issue in ethical/social research because it relates to the validity of research results. There are contradictory results where the most influential MCSD scale

from Crowne & Marlowe (1960), has verifiable intersubjectivity in Brazil and Argentina, but not elsewhere. Various short version scales of the MCSD (Crowne & Marlowe, 1960) have different performances for the same sample, meaning that these scales measure different constructs. The constructs of depression and anxiety were shown to be no different from SDB. There is even a negative correlation between anger and the MC-10 scale. The author views that inconsistencies in research results occur because SDB is a very contextual construct, depending on the values and norms that apply physically as perceived by respondents.

In this research, SDB is measured with a strong conceptual premise, namely the desire to be socially accepted regarding ethical matters that apply in society. By providing a stimulus in the form of a statement related to religiosity, SDB has great potential to occur. A number of respondents are also very likely to have a high level of narcissism based on the results of self-reporting regarding religious activities. The method proposed in this research is mitigation in nature and can improve research results that have already been completed without SDB mitigation.

**Acknowledgments.** The authors would like to thank LPPM Universitas Serang Raya. The author also would like to thank Universitas Gadjah Mada for providing access to Scopus which enabled the author to collect data via PoP.

**Disclosure of Interests.** This study was funded in part by LPPM Universitas Serang Raya and is one of the promised outputs. Both authors are lecturers at Universitas Serang Raya who successfully obtained an internal grant. The authors have taken the necessary steps to avoid problems in the future.

## References

- [1] R. A. Bernardi and J. Nash, "The importance and efficacy of controlling for social desirability response bias," *Ethics Behav*, vol. 33, no. 5, pp. 413–429, 2023, doi: 10.1080/10508422.2022.2093201.
- [2] D. C. Ganster, H. W. Hennessey, and F. Luthans, "Desirability Response Effects," *Academy of Management Journal*, vol. 26, no. 2, pp. 321–331, 1983, [Online]. Available: <http://www.jstor.org/stable/255979>.
- [3] S. A. Desmond, S. E. Soper, and R. Kraus, "Religiosity, Moral Beliefs, and Delinquency: Does the Effect of Religiosity on Delinquency Depend on Moral Beliefs?," *Sociological Spectrum*, vol. 29, no. 1, pp. 51–71, 2008, doi: 10.1080/02732170802480527.
- [4] A. M. Patwardhan, M. E. Keith, and S. J. Vitell, "Religiosity, Attitude Toward Business, and Ethical Beliefs: Hispanic Consumers in the United States," *Journal of Business Ethics*, vol. 110, no. 1, pp. 61–70, Sep. 2012, doi: 10.1007/s10551-011-1147-8.
- [5] B. K. Robinson and M. N. Wicks, "Religiosity, Self-Efficacy for Exercise, and African American Women," *J Relig Health*, vol. 51, no. 3, pp. 854–864, 2012, doi: 10.1007/s10943-010-9397-9.

- [6] H. A. Wibowo and M. R. Masitoh, "Measuring religiosity and its effects on attitude and intention to wear a hijab: Revalidating the scale," in *Increasing Management Relevance and Competitiveness*, 1st ed., Badri Munir Sukoco, Rahmat Heru Setianto, Nidya Ayu Arina, Ade Gafar Abdullah, and R. H. Asep Bayu Dani Nandiyanto, Eds., London: CRC Press, 2018, ch. 44, p. 538. doi: 10.1201/9781351241892-44.
- [7] H. Usman, D. Mulia, C. Chairy, and N. Widowati, "Integrating trust, religiosity and image into technology acceptance model: the case of the Islamic philanthropy in Indonesia," *Journal of Islamic Marketing*, vol. 13, no. 2, pp. 381–409, 2022, doi: 10.1108/JIMA-01-2020-0020.
- [8] R. Moosa and S. Kashiramka, "Objectives of Islamic banking, customer satisfaction and customer loyalty: empirical evidence from South Africa," *Journal of Islamic Marketing*, vol. 14, no. 9, pp. 2188–2206, Aug. 2023, doi: 10.1108/JIMA-01-2022-0007.
- [9] D. P. Crowne and D. Marlowe, "A new scale of social desirability independent of psychopathology.," *J Consult Psychol*, vol. 24, no. 4, pp. 349–354, Aug. 1960, doi: 10.1037/h0047358.
- [10] S. D. Barger, "The Marlowe-Crowne affair: Short forms, psychometric structure, and social desirability," *J Pers Assess*, vol. 79, no. 2, pp. 286–305, 2002, doi: 10.1207/S15327752JPA7902\_11.
- [11] J. W. Buckley, M. H. Buckley, and H.-F. Chiang, *RESEARCH METHODOLOGY & BUSINESS DECISIONS*. New Jersey & Ontario: National Association of Accountants & The Society of Industrial Accountants of Canada, 1976.
- [12] J. W. Buckley, M. H. Buckley, and H.-F. Chiang, *RESEARCH METHODOLOGY & BUSINESS DECISIONS*. New Jersey & Ontario: National Association of Accountants & The Society of Industrial Accountants of Canada, 1976.
- [13] M.-S. Jo, "Controlling social-desirability bias via method factors of direct and indirect questioning in structural equation models," *Psychol Mark*, vol. 17, no. 2, pp. 137–148, Feb. 2000, doi: 10.1002/(SICI)1520-6793(200002)17:2<137::AID-MAR5>3.0.CO;2-V.
- [14] N. M. Bradburn, S. Sudman, E. Blair, and C. Stocking, "Question Threat and Response Bias," *Public Opin Q*, vol. 42, no. 2, p. 221, 1978, doi: 10.1086/268444.
- [15] A. Durmaz, İ. Dursun, and E. T. Kabadayi, "Mitigating the Effects of Social Desirability Bias in Self-Report Surveys," in *Applied Social Science Approaches to Mixed Methods Research*, 2020, pp. 146–185. doi: 10.4018/978-1-7998-1025-4.ch007.
- [16] D. L. Paulhus, "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, Third Revi., Elsevier, 1991, pp. 17–59. doi: 10.1016/B978-0-12-590241-0.50006-X.
- [17] D. L. Paulhus, "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, Third Revi., Elsevier, 1991, pp. 17–59. doi: 10.1016/B978-0-12-590241-0.50006-X.

- [18] M. F. King and G. C. Bruner, "Social desirability bias: A neglected aspect of validity testing," *Psychol Mark*, vol. 17, no. 2, pp. 79–103, Feb. 2000, doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0.
- [19] H. C. Tan, J. A. Ho, R. Kumarusamy, and M. Sambasivan, "Measuring social desirability bias: Do the full and short versions of the Marlowe-Crowne Social Desirability scale matter?," *Journal of Empirical Research on Human Research Ethics*, vol. 17, no. 3, pp. 382–400, 2022, doi: 10.1177/15562646211046091.
- [20] A. L. Edwards, *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden Press, 1957.
- [21] A. L. Edwards, *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden Press, 1957.
- [22] C. G. Davis, J. Thake, and N. Vilhena, "Social desirability biases in self-reported alcohol consumption and harms," *Addictive Behaviors*, vol. 35, no. 4, pp. 302–311, 2010, doi: 10.1016/j.addbeh.2009.11.001.
- [23] J. Adong *et al.*, "Social Desirability Bias Impacts Self-Reported Alcohol Use Among Persons With HIV in Uganda," *Alcohol Clin Exp Res*, vol. 43, no. 12, pp. 2591–2598, Dec. 2019, doi: 10.1111/acer.14218.
- [24] J. Adong *et al.*, "Social Desirability Bias Impacts Self-Reported Alcohol Use Among Persons With HIV in Uganda," *Alcohol Clin Exp Res*, vol. 43, no. 12, pp. 2591–2598, Dec. 2019, doi: 10.1111/acer.14218.
- [25] C. Bir and N. O. Widmar, "Consistently biased: documented consistency in self-reported holiday healthfulness behaviors and associated social desirability bias," *Humanit Soc Sci Commun*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.1057/s41599-020-00665-x.
- [26] C. G. Davis, J. Thake, and N. Vilhena, "Social desirability biases in self-reported alcohol consumption and harms," *Addictive Behaviors*, vol. 35, no. 4, pp. 302–311, 2010, doi: 10.1016/j.addbeh.2009.11.001.
- [27] S. P. Singh and J. Tir, "Threat-Inducing Violent Events Exacerbate Social Desirability Bias in Survey Responses," *Am J Pol Sci*, vol. 67, no. 1, pp. 154–169, 2023, doi: 10.1111/ajps.12615.
- [28] M. F. Arezzo, I. A. Horodnic, C. C. Williams, and G. Guagnano, "Measuring participation in undeclared work in Europe using survey data: A method for resolving social desirability bias," *Socioecon Plann Sci*, vol. 91, no. November 2023, 2024, doi: 10.1016/j.seps.2023.101779.
- [29] B. M. King *et al.*, "Extreme under-reporting of body weight by young adults with obesity: relation to social desirability," *Obes Sci Pract*, vol. 4, no. 2, pp. 129–133, Apr. 2018, doi: 10.1002/osp4.153.
- [30] W. L. Teh *et al.*, "Measuring social desirability bias in a multi-ethnic cohort sample: its relationship with self-reported physical activity, dietary habits, and factor structure," *BMC Public Health*, vol. 23, no. 1, p. 415, Mar. 2023, doi: 10.1186/s12889-023-15309-3.
- [31] M. Fraboni and D. Cooper, "Further Validation of Three Short Forms of the Marlowe-Crowne Scale of Social Desirability," *Psychol Rep*, vol. 65, no. 2, pp. 595–600, Oct. 1989, doi: 10.2466/pr0.1989.65.2.595.

- [32] E. Carmines and R. Nassar, "How Social Desirability Bias Affects Immigration Attitudes in a Hyperpolarized Political Environment," *Soc Sci Q*, vol. 102, no. 4, pp. 1803–1811, 2021, doi: 10.1111/ssqu.12982.
- [33] O. Marcucci, "Implicit Bias in the Era of Social Desirability: Understanding Antiracism in Rehabilitative and Punitive School Discipline," *Urban Review*, vol. 52, no. 1, pp. 47–74, 2020, doi: 10.1007/s11256-019-00512-7.
- [34] H. Kuokkanen and W. Sun, "Social desirability and cynicism biases in CSR surveys: an empirical study of hotels," *Journal of Hospitality and Tourism Insights*, vol. 3, no. 5, pp. 567–588, 2020, doi: 10.1108/JHTI-01-2020-0006.
- [35] H. Kuokkanen and W. Sun, "Social desirability and cynicism biases in CSR surveys: an empirical study of hotels," *Journal of Hospitality and Tourism Insights*, vol. 3, no. 5, pp. 567–588, 2020, doi: 10.1108/JHTI-01-2020-0006.
- [36] S. Ahmad, R. Lensink, and A. Mueller, "Religion, social desirability bias and financial inclusion: Evidence from a list experiment on Islamic (micro-)finance," *J Behav Exp Finance*, vol. 38, p. 100795, 2023, doi: 10.1016/j.jbef.2023.100795.
- [37] S. Ahmad, R. Lensink, and A. Mueller, "Religion, social desirability bias and financial inclusion: Evidence from a list experiment on Islamic (micro-)finance," *J Behav Exp Finance*, vol. 38, p. 100795, 2023, doi: 10.1016/j.jbef.2023.100795.
- [38] A. L. Thorne-Lyman *et al.*, "How does social desirability bias influence survey-based estimates of the use of antenatal care in rural Nepal? A validation study," *BMJ Open*, vol. 13, no. 7, pp. 1–8, 2023, doi: 10.1136/bmjopen-2022-071511.
- [39] I. Krumpal, "Determinants of social desirability bias in sensitive surveys: a literature review," *Qual Quant*, vol. 47, no. 4, pp. 2025–2047, Jun. 2013, doi: 10.1007/s11135-011-9640-9.
- [40] S. Nazari, W. L. Leite, and A. C. Huggins-Manley, "A Comparison of Person-Fit Indices to Detect Social Desirability Bias," *Educ Psychol Meas*, vol. 83, no. 5, pp. 907–928, 2023, doi: 10.1177/00131644221129577.
- [41] W. M. Reynolds, "Development of reliable and valid short forms of the marlowe-crowne social desirability scale," *J Clin Psychol*, vol. 38, no. 1, pp. 119–125, Jan. 1982, doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I.
- [42] A. B. Silverstein, "Validity of random short forms: II. The Marlowe-Crowne social desirability scale," *J Clin Psychol*, vol. 39, no. 4, pp. 582–584, 1983, doi: 10.1002/1097-4679(198307)39:4<582::AID-JCLP2270390420>3.0.CO;2-C.
- [43] D. G. Fischer and C. Fick, "Measuring social desirability: Short forms of the marlowe-crowne social desirability scale," *Educ Psychol Meas*, vol. 53, no. 2, pp. 417–424, 1993, doi: 10.1177/0013164493053002011.
- [44] A. Franzen and S. Mader, "Do phantom questions measure social desirability?," *Methods, Data, Analyses*, vol. 13, no. 1, pp. 37–57, 2019, doi: 10.12758/mda.2019.01.
- [45] R. D. Hays, T. Hayashi, and A. L. Stewart, "A Five-Item Measure of Socially Desirable Response Set," *Educ Psychol Meas*, vol. 49, no. 3, pp. 629–636, Sep. 1989, doi: 10.1177/001316448904900315.

- [46] F. W. Rudmin, "Norwegian short-form of the Marlowe-Crowne Social Desirability Scale Norwegian short-form of the Marlowe-Crowne Social Desirability Scale," *Scand J Psychol*, vol. 40, no. September, pp. 229–233, 1999.
- [47] W. M. Reynolds, "Development of reliable and valid short forms of the marlowe-crowne social desirability scale," *J Clin Psychol*, vol. 38, no. 1, pp. 119–125, Jan. 1982, doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I.
- [48] D. L. Paulhus, "Two-component models of socially desirable responding.," *J Pers Soc Psychol*, vol. 46, no. 3, pp. 598–609, Mar. 1984, doi: 10.1037/0022-3514.46.3.598.
- [49] S. N. Beretvas, J. L. Meyers, and W. L. Leite, "A Reliability Generalization Study of the Marlowe-Crowne Social Desirability Scale," *Educ Psychol Meas*, vol. 62, no. 4, pp. 570–570, Aug. 2002, doi: 10.1177/001316402128775021.
- [50] R. Ballard, "SHORT FORMS OF THE MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE," *Psychol Rep*, vol. 71, pp. 1155–1160, 1992.
- [51] R. Van Dick, G. K. Fisher, M. A. West, and J. F. Dawson, "A Test of Basic Assumptions of Affective Events Theory ( AET ) in Call Centre Work 1," vol. 17, pp. 237–254, 2006, doi: 10.1111/j.1467-8551.2006.00489.x.
- [52] R. Loo and K. Thorpe, "Confirmatory Factor Analyses of the Full and Short Versions of the Marlowe-Crowne Social Desirability Scale," *J Soc Psychol*, vol. 140, no. 5, pp. 628–635, Oct. 2000, doi: 10.1080/00224540009600503.
- [53] H. Seol, "A Psychometric Investigation of the Marlowe-Crowne Social Desirability Scale Using Rasch Measurement," *Measurement and Evaluation in Counseling and Development*, vol. 40, no. 3, pp. 155–168, Oct. 2007, doi: 10.1080/07481756.2007.11909812.
- [54] V. Vésteinsdóttir, U.-D. Reips, A. Joinson, and F. Thorsdottir, "Psychometric properties of measurements obtained with the Marlowe–Crowne Social Desirability Scale in an Icelandic probability based Internet sample," *Comput Human Behav*, vol. 49, pp. 608–614, Aug. 2015, doi: 10.1016/j.chb.2015.03.044.
- [55] R. Ballard, "SHORT FORMS OF THE MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE," *Psychol Rep*, vol. 71, pp. 1155–1160, 1992.
- [56] R. Loo and K. Thorpe, "Confirmatory Factor Analyses of the Full and Short Versions of the Marlowe-Crowne Social Desirability Scale," *J Soc Psychol*, vol. 140, no. 5, pp. 628–635, Oct. 2000, doi: 10.1080/00224540009600503.
- [57] F. B. Scagliusi *et al.*, "Tradução da escala de desejo de aceitação social de Marlowe & Crowne para a língua portuguesa," *Archives of Clinical Psychiatry (São Paulo)*, vol. 31, no. 6, pp. 272–278, 2004, doi: 10.1590/S0101-60832004000600001.
- [58] A. C. COSENTINO and Y. A. C. SOLANO, "Adaptación y validación Argentina de la Marlow-Crowne Social Desirability Scale," *Interdisciplinaria*, vol. 25, no. 2, pp. 197–216, 2008.
- [59] H. Seol, "A Psychometric Investigation of the Marlowe-Crowne Social Desirability Scale Using Rasch Measurement," *Measurement and Evaluation in Counseling and Development*, vol. 40, no. 3, pp. 155–168, Oct. 2007, doi: 10.1080/07481756.2007.11909812.

- [60] V. Vésteinsdóttir, U.-D. Reips, A. Joinson, and F. Thorsdottir, "Psychometric properties of measurements obtained with the Marlowe–Crowne Social Desirability Scale in an Icelandic probability based Internet sample," *Comput Human Behav*, vol. 49, pp. 608–614, Aug. 2015, doi: 10.1016/j.chb.2015.03.044.
- [61] J. W. Creswell and J. D. Creswell., *Research design : qualitative, quantitative, and mixed methods approaches*, 6th ed., vol. 94, no. 4. Thousand Oaks, California 91320: SAGE Publications, Inc, 2023.
- [62] T. Salzberger and M. Koller, "Towards a new paradigm of measurement in marketing," *J Bus Res*, vol. 66, no. 9, pp. 1307–1317, Sep. 2013, doi: 10.1016/j.jbusres.2012.02.030.
- [63] B. Sumintono and W. Widhiarso, *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial (edisi revisi)*, no. November 2014. 2016.
- [64] B. Sumintono and W. Widhiarso, *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*, 2nd ed., no. November. Bandung: Trim Komunikata Publishing House, 2014.
- [65] T. G. Bond and C. M. Fox, *Applying the Rasch Model Fundamental Measurement in the Human Sciences*, 3rd ed. New York: Routledge, 2015.
- [66] S. H. Wijanto, *STRUCTURAL EQUATION MODELING DENGAN LISREL 8.8 KONSEP & TUTORIAL*, 1st ed. Yogyakarta: Graha Ilmu, 2008.
- [67] H. A. Wibowo and N. Indarti, "Blue-Collar Workers Entrepreneurial Intentions and The Extended Theory of Reasoned Action: Incorporating SEM and Person-Item Map Analysis," *Journal of Indonesian Economy and Business*, vol. 35, no. 3, p. 204, Sep. 2020, doi: 10.22146/jieb.52046.
- [68] J. M. Linacre, *A User's Guide to WINSTEP & MINISTEP: Rasch-Model Computer Programs*. Chicago, 2012.
- [69] J. M. Linacre, *A User's Guide to WINSTEP & MINISTEP: Rasch-Model Computer Programs*. Chicago, 2012.
- [70] E. V. Wolfe, E.W., & Smith, "Instrument development tools and activities for measure validation using Rasch models: Validation activities," *Journal of Applied Measuremen*, vol. 8, no. 2, pp. 204–234, 2007.
- [71] Jr. Smith, E. V. and E. W. Wolfe, "Understanding Rasch measurement: Instrument development tools and activities for measure validation using Rasch models: Part 2–Validation activities.,", *J Appl Meas*, vol. 8, no. 2, pp. 204–234, 2007.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

