

# Detection Careles From Responden Within Examination Outlier Data Identifying Respondent's Carelessness Within The Outlier Data

Iwan Gunawan, Amin Kuncoro, Heru Yulianto

**Abstract:** This research examined data of the selected respondents of 300 students majoring in management by processing instrument tests and identifying primary data from the possibility of respondents' carelessness in filling in the questionnaire. This study applied data analysis technique of Mehanolibis test in order to diminish bias results of the questionnaire. The use of questionnaires for this quantitative research is to answer and describe the extent to which the respondent answers with responsibility or only perfunctory. The researcher uses Rasch model in an attempt to acquire a more accurate result compares to any other models. Rasch model assists researcher to obtain a maximum result even to the level of person correlation since it generates a more replicable one. Instrument testing and validation are inevitable and essential elements before stepping into inferential statistics that seeks to acquire an answer to the proposed research question.

**Keywords:** Rasch model and replicable

## Introduction

Scientific work is a scientific-related piece produced by a researcher which gains scientific recognition. Scientific works possess a more meaning and positive value for mankind, therefore, scientific work will be more useful if it is officially registered and has legality in accordance with the government regulations. Each scientific work has its own specialty based on the field of knowledge of its scholars. This means that the more scholars producing patented scientific papers in accordance with the regulations, the more contributions devoted to the state since these scientific works will be intellectual properties of the government. According to Susilo, Eko, (1995) an article is obtained with a particular property, field, method, and appropriate language that can be accounted for. This is also in line with the argument of Yamilah and Samsorizal that the diversity of scientific work is very useful to empower and develop its field of knowledge. Researchers have their own field of knowledge and expertise, yet every researcher will defend the results of his/her research in the national and international forums. Every study provides benefits in accordance with its field of knowledge. Having said that, it will further contribute to the more diverse areas of studies along with their benefits. Quantitative research emphasizes empirical results because it uses primary data which will be processed into a particular data in order to support the research findings and theoretical framework. Researchers who prioritize the process of the primary data demonstrate an ethical conduct in presenting the results in a structured way.

This aims to make the study describable and enable readers to comprehend the results of the empirical research. Primary data has a drawback in its unstructured form, however, there is an easier way to make the data 'more normal', more structured, and continuable for the next data processing. The varied data is often defined as data derived from variables that are inconsistent with the research framework, compared to the results of well-distributed values (Bethlehem, 2009). Primary data in this research are obtained from the respondents' answers on the questionnaires. The shortcoming of this obtained primary data is that the respondents tend to provide perfunctory answers without a strong reasoning. Respondents who answer the questionnaires without reading and comprehending the substantial questions will likely lead the results to be biased; this is because the data is clumped and meaningless. Errors or confusions of the data collected from the respondents are often useless because of its irregular model (Lee, 2008). Data error with an abnormal distance is referred to as Outlier which exceeds error or variability level (Anscombe, 1960). Accordingly, scholars often discover such invalid results which means that some data set may represent the valid results but some others comprise only a few indications of validity. Another way to indicate an outlier (Evan, 1999), Osborn and Overbay, (2004). Before categorizing data through coding, editing, and tabulating; primary data should be processed into valid and reliable data. However, it is important to note that some tools, including free and paid software, have a considerable error rate especially the level of validity and reliability. Most researchers still use the old method by referring to the Likert scale which is basically the result of the respondent's answer serving as the ordinal data. Previous research has made it clear that various data from structured and unstructured data may provide different results (e.g., Leys, Ley, Klein, Bernard, & Licata, 2013; Wang & Serfling, 2015). Research using Rasch model is developed for various approaches especially inferential research. Inferential research is a study that uses complex metadata yet it is presented in a simple method (Kuncoro, Amin, and Sudarman, 2018). Metadata is still a set of complex numbers which is examined to find its validity and reliability. The results of the instrument test (validity and

- Iwan Gunawan, Amin Kuncoro, Heru Yulianto
- Lecture of Indonesia Shipping Academy, Email: [captiwanunawan@gmail.com](mailto:captiwanunawan@gmail.com)
- Lecture and Researcher of STIE AKA Semarang, Indonesia; Email: [kiaisekuler@gmail.com](mailto:kiaisekuler@gmail.com)
- Lecture and Researcher of STIE AKA Semarang, Indonesia;

reliability) are further used as a reference for the metadata test. Rasch model includes a simple approach and generates a better result because this model facilitates to think a logical model. The various numbers of the weighted metadata are expected to respond to the residual data with a certain weight (Karabatsos, 2003).

## Objective

The study aims to explain the results of the test instrument which is often found unclear because of the varied answers of the respondents. Besides, the researcher gives an optional score ranging from 1 to 5. Therefore, respondents are tended to be forced in choosing the available option which sometimes does not suit their conscience. The impact of researcher's 'coercion' on respondent occurs when the respondent does not have the relevant knowledge of the questions. Hence, it is likely that the respondents will answer in a perfunctory way. Castro, (2013) illustrates how factor assessment may cause the possibility of incorrect data. Some researchers often use the instrument test to legalize the metadata which is further processed as a ready-to-use primary data. Nevertheless, (Curtis, 2004) argues that the instrument test has various possibilities which explain the differential functions and some other relevant possibilities in order to avoid confusion. Researchers should be more careful since the results of the test instrument are more about data interpretation which often lies on the researcher's perceptions. Other results explain that outliers in a group of data must be anticipated because it can interfere the metadata of the researchers. This study is expected to answer the error rate of the researchers, especially those who use primary data from respondent's answers using a scale of 1 to 5. Another factor causing bad results of the test instrument lies on the lack of respondent's answer and a tendency level of the error factor of the researchers. The error rate is a positive response when using Rasch model because the resulting total score will be close to the manual data improvement. Most respondents answer that they do not have the knowledge of the meaning of each data instrument. When data is correlated, it becomes ambiguous and fails to provide the researchers a better result. The public anomaly of the researcher shows a particular distribution proving an excellent statistical data that can be accounted for, however, it is without knowing the minimum score which makes the framework unequal. The hypothesis of this study explains: (1) to what extent the majority of respondents give uncertain (perfunctory) answers; and (2) to what extent respondents provide desirable answers. The research framework of this research is rooting to what extent respondents engage a positive action by understanding research instrument so that the results of respondents' answer will represent other respondents in a different context.

## Methodology

### Data Analysis

Descriptive statistics provided more explanation about data which the researchers used. The model of the research examined the primary data, that was, the data of the selected respondents answering all questionnaire questions. Based on the description of respondents, they

were classified into several categories: (1) respondents who answered the questions by understanding the substantial contents of the questions; (2) respondents who randomly answered the questions; and (3) respondents who did not answer at all. Respondents who answered by understanding the contents of the question were given the value of 1.5; respondents who randomly answered the questions (perfunctory) were given the value of 1.5; and respondents who did not answer the questions were given the value of 0.5. The outliers of the study applied Mahalonobis Distance statistics (MD). Data processing used Winsteps 3.02 software (Linacre, 2000). Descriptive statistics used MD with Chi-Square model by referring to Fidel and Tabachnick (2003) that MD answered the congenital deviation.

### Data Collection

The accuracy of the infit-outfit model on Rasch model was applied to identify data discrepancies. Rasch model indicated data irregularities because of the respondent's distorted and unexpected answers. Widhiarso and Sumintono (2016) state that Rasch model is more intense in delivering a maximum result, especially to data irregularities. Furthermore, in some items of the respondent answer pattern, including person correlation, will be shown through Rasch model (Sumintono & Widhiarso, 2013). This research used primary data of the enrolled university students with a clear description of their gender and age. The technique was providing information to the university students by asking them to answer questionnaires consisting of eight instruments. These instruments contained cognitive aspect, personality aspect, and interest of the students in taking Management majors.

### Result

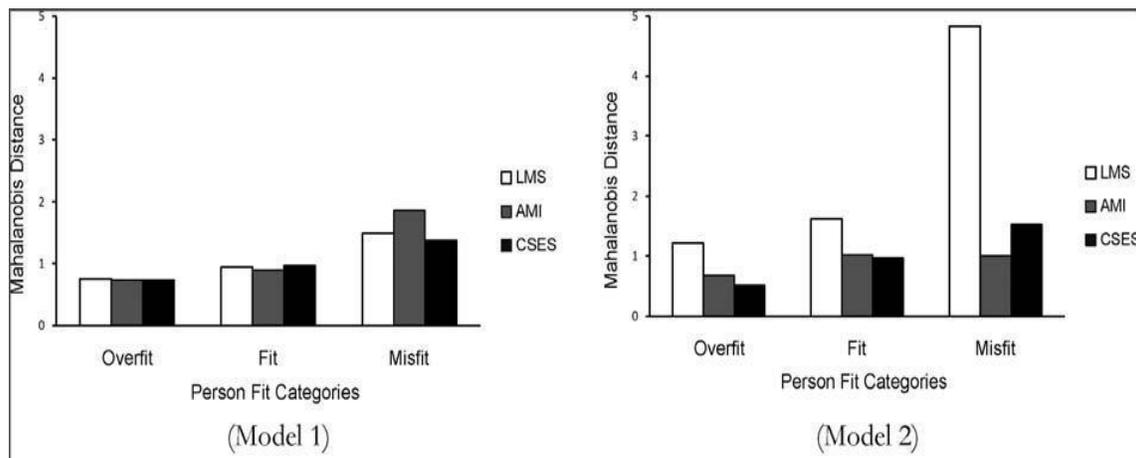
The results of validity and reliability tests using software other than Rasch model identified an uncertainty of each instrument value. This was because the respondents answered with a lack of efforts. Another factor was the lengthy questions, unmatched indicators with the respondent's perception, and the measurement scales of 1 to 5. Another finding discovered a deviation of the results and the discrepancy of the model, therefore, better and more reliable tests were needed. Referring to the fit model, some items were not favorable with the analysis results of below 0.5 and above 1.5. Sumitro and Widhiarso (2013) explain that the fit model is between 0.5 and 1.5. Another result demonstrated two items of the research instruments which emphasized in MB, PA; indicated no unused IP items; showed consistency of the results ( $\alpha$ ); and identified an index that provided a separated track (G) of some recent items of size 83 and 2.34 (MB), 70 and 1.52 (PA), as well as 0.74 and 1.68 (IP). This study applied the latest methods on each research instrument in an effort to generate maximum results. The results of frequency distribution explained the respondent's interest as measured by mean (M) result that was close to 1.0 and the standard deviation (SD) that was close to 50. Several respondents with different results of answers were the deviation of PA (infit = 22.3%, usability = 22.5%) or MB (infit = 29.3% and usability = 29.0%). The correlation between them was 0.37-0.46, meaning that the respondents gave their best efforts to respond the research instruments. Another result using MB

in PA (model 1) and MB in IP (model 2) showed that this analysis revealed a consistency. This was done through outlier test on any descriptive statistics analysis.

**The Correlation between Outlier and Fit Model**

The correlation between incompatibility of each outlier (infinitive statistics) and application of the smallest outlier scores is Model 1 ( $r = .16$  and  $r = .22$ , Prob. 001). Model 2 describes that  $r = .11$  and  $r = .11$  with  $p < .000$ ). Each model in the first assessment has a significant relationship between the x variable and the y level. This proves that the model suitability using the Rasch method is very significant. These results considerably support the hypothesis because there is a good relationship of variable x to variable y without a deviation of the model. In Figure 1, it is explained that the comparison value of Mahalanobis Distance has

three criteria of descriptive statistics model. Each number in the Mahalanobis Distance acquires different result, that is, each picture has a different picture. This means the respondents respond with a variation of their perception and condition at the time of the interview. The value of MD (Mahalanobis Distance) shows a significant difference, especially in the value of PA. Another factor of the variation of the respondent's answer is that they feel saturated to read the instruments with long questions. Other descriptive statistics are presented in Table 1, that is, a particular item (Mahalanobis Distance) with various levels explain one best value and three of which cannot be explained by descriptive statistics. The standard deviation among the three possesses a deviating result hence the outlier value is varied. The result of ANOVA test reveals three categories which suit the value of Mahalanobis Distance (MD).



**Figure 1.** Histogram, Mahalanobis Distance, Fit dan Underfit. PA, MB, IP.

**Table 1**  
Mahalanobis Statistical Descriptif with All Categories

Instruments	Person-Fit Categories	Model 1		F-test	Model 2		F-test
		S	KK		S	KK	
MB	Overfit	0.65	0.77	F = 21.72 (p b 0.001)	1.21	0.06	F = 5.58 (p b 0.01)
	Fit	0.90	0.92		1.61	0.04	
	Underfit	1.40	1.59		4.80	0.23	
PA	Overfit	0.71	1.27	F = 51.10 (p b 0.001)	0.69	1.13	F = 1.37 (p N 0.05)
	Fit	0.82	1.58		1.07	2.53	
	Underfit	1.84	2.77		1.00	1.56	
IP	Overfit	0.70	1.01	F = 11.90 (p b 0.001)	0.55	0.81	F = 15.07 (p b 0.001)
	Fit	0.98	1.78		0.99	2.48	
	Underfit	1.39	2.58		1.58	2.38	

**Note** = MB : Motivation to learn, PA = Academic Achievement, IP = Achievement Index

The table demonstrates that the regression in model 1 matches the MB of 1.37 with a p-value of N.0.05. This means when model 1 uses post hoc, there a difference between people and strong academic achievement (S.diff = .56: p < .001). There is good value of academic achievement (M.diff = .78: p < .000). Nevertheless, there is

still an unfavorable value of the achievement index (M.diff = .21: p N. 05). The second model explains the same result, that is, academic achievement has a higher significance compared to interest to learn. The second figure illustrates the non-clustered results between MB and PA. This means that the results are normally distributed. This also clarifies

that the respondent's answer has a particular intention which is qualified to get such attention. Some respondents explain that there are differences in academic achievement. This proves the achievement of the value of both variables. The deviation of the MB result with a higher PA score emerges because it lies on the straight line dimension which approaches regression. Accordingly, it can be predicted that MB and PA are not significantly different when the achievement index is included with the value of  $b = 7.86$ :  $p < .001$  or excluded with the value of  $b: 745$ ;  $p < 0.000$  with the achievement index as the predictor. The correlations of the two variables in model 1 and model 2 show an increase when the achievement index obtains a clearer instrument. The intense correlations indicate that the infit statistics are ranging from  $b1.6$  to  $b1.1$ . This examines a particular support stating that the variable of interest to learn and the variable of academic achievement are correlated; and the achievement index gives more picture of the student assessment on their final achievement as university students.

## Discussion

This study is conducted with an assumption that data of respondents, collected through questionnaire distribution, shows errors or mistakes in filling in the provided Likert scale. The error often lies in the less effective answers hence the researchers need to apply Rasch model. Rasch test provides a significant result because it identifies and sorts the less precise answers of the respondents. Some researchers use SPSS Software as their instrument test yet it does not give a maximum result. This is because respondent's answer is viewed carelessly or the research questions as described in the questionnaire are not well-comprehended. If the researchers apply a particular tool to explain the extent of the questionnaire's validity, it is necessary to answer with such certainty because the effect will be further analyzed when regression test is conducted. The results of this study answer the confusion of the researchers because of the not optimum results of the regression test. For that reason, implications of the research must have a strong narrative. Regression test has a minimum error rate, to be exact, it is at the level of error. However, it has not provided an exact answer on how big the ability of the independent variables toward the dependent ones. These results are also in line with the previous studies of Zijlstra, van der Ark, and Sijtsma (2010) as well as Woods, Oltmanns, and Turkheimer (2009). They explain deviations or different results may be found when Rasch model is not applied, this is due to carelessness of the respondents in answering questionnaires (Meijer & Sijtsma, 2001). Insertion of the inferential results causes a different scope of the respondent's answers. This is because the respondents tend to give perfunctory answers and they do not show such commitment to complete the provided optional answers. Another factor is the insufficient time and the lack of understanding to read in detail. This analysis also illustrates that respondents are relatively stable in giving their response, their stability is due to the high outliers. Other different results are also presented by (Panayides & Tymms, 2013). They found respondent's inconsistency with a high deviation of response.

## Reference

- [1]. Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, 2(2), 123–146. <http://dx.doi.org/10.1080/00401706.1960.10489888>.
- [2]. Bethlehem, J. (2009). *Applied survey methods*. Hoboken, NJ: John Wiley & Sons, Inc.
- [3]. Castro, R. (2013). Inconsistent respondents and sensitive questions. *Field Methods*, 25(3), 283298. <http://dx.doi.org/10.1177/1525822x12466988>.
- [4]. Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal*, 5(2), 125–143.
- [5]. Evans, V. P. (1999). Strategies for detecting outliers in regression analysis: An introductory primer. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.
- [6]. Fidell, L. S., & Tabachnick, B. G. (2003). Handbook of psychology: Research methods in psychology. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), Hoboken, NJ: John Wiley & Sons.
- [7]. Kuncoro, Amin & Sudarman, 2018 *Metode penelitian manajemen*, Andy Offset, Yogyakarta.
- [8]. Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirtysix person-fit statistics. *Applied Measurement in Education*, 16(4), 277–298. [http://dx.doi.org/10.1207/S15324818AME1604\\_2](http://dx.doi.org/10.1207/S15324818AME1604_2).
- [9]. Linacre, J. M. (2000). *WINSTEPS*, version 3.02. Chicago: Winstep.com.
- [10]. Lee, H. (2008). Outliers. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*. Thousand Oaks, CA: SAGE Publications, Inc.
- [11]. Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education*, 9(1), 3–8.
- [12]. Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107–135. <http://dx.doi.org/10.1177/01466210122031957>.
- [13]. Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, 9(6).
- [14]. Panayides, P., & Tymms, P. (2013). Investigating whether aberrant response behaviour in classroom maths tests is a stable characteristic of students. *Assessment in Education: Principles, Policy & Practice*, 20(3), 349–368. <http://dx.doi.org/10.1080/0969594X.2012.723610>.

- [15].Widhiarso dan Sumintono, (2016)Aplikasi Model Rasch untuk Penelitian Ilmu- Ilmu Sosial. See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/256498376>
- [16].Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2010). Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics*. <http://dx.doi.org/10.3102/1076998610366263>.
- [17].Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2011). Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics*, 36(2), 186–212 <http://dx.doi.org/10.3102/1076998610366263>.