

# Controversy in Determining Criteria and Categories in Summarizing and Exploring The Research Data; Analysis of Assessment Procedures in The Social Science Research

Zadrian Ardi<sup>1</sup>, Daharnis<sup>2</sup>, Verlanda Yuca<sup>3</sup>, Ifdil Ifdil<sup>4\*</sup>

<sup>1234</sup>Guidance and Counseling, Faculty of Education, Universitas Negeri Padang, Indonesia

## ABSTRACT

Determination of criteria and categorization of data in measurement is a crucial moment when processing research data. Overview, exploration and presentation of data will be effective and valid if the criteria set meets the requirements, both from a theoretical and statistical perspective. However, the fact that often arises, especially in beginner researchers made a mistaken in determining the criteria and categories. These errors are often indicated in the form of incomprehension and negligence of researchers in placing data types with the type of methodology. This paper describes various errors in determining criteria and categories of research data, determining criteria and categories of research data, and suggesting their meaning, especially in social science research. This research involved respondents from four different types of assessments, which is 428 respondents from Daharnis-Zadrian Majoring Choices Inventory (DZ-MCI) and 174 respondents from the Acceptability of Mental-Health Mobile-App Survey (AMMS), as well as 1190 respondents from the development of DLAS and DLESS inventory. The results of the study show that the determination of categories and criteria in data processing greatly impacts the truth of data exposure and advanced data analysis. Implications of these studies for determining categories and criteria of research on social science research are discussed.

## Keywords

Category and Criteria Data Research, Assessment Procedure, Social Science Research, Criterion-Referenced Measurement, Norm-Referenced Measurement

*Article Received: 20 September 2020, Revised: 30 November 2020, Accepted: 18 December 2020*

## Introduction

Many assessors and researchers, especially the amateur ones have difficulty in determining the criteria or categorization of research results that have been obtained (Meyer, Doromal, Wei, & Zhu, 2017). Sometimes they use criteria from the results of previous studies that are considered "true" even though the categorization is actually "wrong". There are also those who adopt raw assessment criteria for an educational institution that are not in accordance with the objectives of the research conducted (de Klerk, Veldkamp, & Eggen, 2015; Rose, Shevlin, Winter, & O'Raw, 2015).

Indeed there is controversy between researchers in setting criteria which are sometimes not in accordance with the objectives of the study (Heinrich et al., 2018; Sekaran & Bougie, 2016) (Coolican, 2017). There are those who use criterion-referenced measurement and norm-referenced measurement. Both can be justified in accordance with the objectives of the research and

the categorization follows scientific rules and can be accounted for logically and statistically (Coolican, 2017).

The problem is the researchers tend to use criteria, besides not being suitable with the purpose of the study (Duckworth & Yeager, 2015; Walliman, 2015) (Busk & Marascuilo, 2015), nor do they follow scientific procedures and logic (D Daharnis et al., 2018).

This article aims to provide an overview of how criteria are determined from research based on criterion-referenced measurement and norm-referenced measurement that meets scientific and logical requirements, also based on statistical analysis.

## Methods

### Participants

The sample in this study was based on the distribution of respondents according to the type

of measuring instrument used. The instruments and samples were adapted from several previous studies. The data came from 428 respondents from the Daharnis-Zadrian Majoring Choices Inventory (DZ-MCI) (Daharnis Daharnis & Ardi, 2016), 174 respondents from Acceptability of Mental-Health Mobile-App Survey (AMMS) (Ardi, Sukmawati, Afdal, Rangka, & Suranata, 2018), and 1190 respondents from Daharnis Learning Activities Scale (DLAS) (D Daharnis, Erlamsyah, Ifdil, Ardi, & Hariko, 2014). The instrument had previously met good measurement requirements, so that in this manuscript it was discussed about how the instrument categorized the position of the respondent on the measured variable.

### Data Analysis

Research data analysis was carried out in accordance with various alternative categorization and grouping of data based on comparison with ideal (criterion-referenced measurement) and inter-respondent scores (norm-referenced measurement). Data grouping based on score distribution is done by utilizing Microsoft Excel software (Triola, 2013) and Statistical Package for Social Science (SPSS). Data analysis was also carried out using Rasch analysis (Sumintono & Widhiarso, 2014) through Winstep Version 3.72 software (Linacre, 2009).

## Result and Discussion

### Result

Data grouping and categorization can be done in two times, which are before the process of collecting research data using ideal score to establish criterion-referenced measurement and after data is collected by using real data through norm-referenced measurement (Škoda, Doulik, Bílek, & Šimonová, 2016; Thompson & McDonald, 2016). The grouping and categorization is based on rationality, literature review and the purpose of conducting research (Idris, Nurhayati, & Satriani, 2018; Nahhas, Bamasag, Khemakhem, & Bajnaid, 2018).

In various sources of data analysis instructions that are popular with academics and researchers, categorization and grouping of data can be done by searching for class intervals and class limits (Lipsey et al., 2017; Shapiro et al., 2017). These results will be made justification about the position of the respondent on the variable being measured, for example in a study, researchers will measure the condition of student achievement motivation in a school, the researcher can categorize the position of respondents which to be known as "high", "moderate", " low" (Buckley, Moore, Boardman, Arya, & Maul, 2017; Dahlke, Kostal, Sackett, & Kuncel, 2018) (Xiao, Bowen, & Lindsey, 2018).

Determination of data categories based on criterion-referenced measurement can be done by calculating the range of scores (highest score - lowest score) plus one point, and then divided by the number of categories needed (Kim, Ad, McAna, & Dicker, 2016; Schumacher et al., 2018) (Estebarez, 2017; Santelli, 2016). By applying this procedure will produce a decimal number. To overcome the absence of a decimal number (rounding), the interval for one category can be raised by one point on the class set by the researcher.

For example, the application of criteria for the procedure for 5 scaling instruments (1-5 scale), and the establishment of five categories can be seen in the results of research on the conditions of student learning activities in West Sumatra-Indonesia whose the data was obtained from the Daharnis Learning Activities Scale (DLAS) with 30 items as shown in Table 1. In the grouping process the lowest and highest score (ideal) becomes the benchmark in determining the data class, where 30 is the lowest score and 150 is the highest score on a 1-5 scale. For data sourced from instruments other than 1 to 5 and categorization other than five the category can adjust according to the designation procedure. Based on the steps to determine the category, the group of data can be presented as follows:

Table 1 Data categories based on interval score and percentage

Interval Score	Percentage	Categories
126 - 150	≥ 84%	Very High
102 - 125	68% – 83%	High
78 - 101	52% – 67%	Moderate
54 - 77	36% - 51%	Low
30 - 53	20% - 35%	Very Low

\*Note: for 1 – 5 scale; 30 items; 5 categories

However, the determination of the steps in Table 1 still reaps quite different controversies and interpretations among academics and researchers in interpreting respondents' achievements, especially in respondents who achieved a threshold score between two categories. This condition is often discussed in the term "statistical casualties" in the data.

The illustration of the case is if a respondent gets a score of 101 on a DLAS measuring instrument, then he will be included in the "moderate" category, whereas if there are other respondents who get a score of 102 then they will get a "high" category. In this case, the score achievement is only 1 point adrift between the two respondents, so that this condition tends to be detrimental and wrong in interpreting a data. Therefore, related to processing and interpretation using these categories, there is a need for in-depth analysis supported by other data to see the condition of the respondents in a holistic and detailed manner. Alternative of data grouping and categorization can also be done by using real / empirical data for

establishing norm-referenced measurement. One application that can be used is the Rasch Model (Hagquist, Bruce, & Gustavsson, 2009; Seo, Taherbhai, & Paek, 2013). In the Rasch Model, real/empirical data is assumed to be ordinal data that is transformed into equal interval data. The transformation was carried out through calculating opportunities in each choice of answers in the form of odd ratio data in the form of opportunities that can be obtained in each answer option calculated based on logarithms. The result of the transformation is in the form of an index of separation amount that will determine how many ideal categories.

The illustrations of determining data categories based on Rasch analysis are presented in Figure 1, whose data is obtained from the Acceptability of Mental-Health Mobile App Survey (AMMS) administration. In Figure 1 it can be seen that the data separation index is 3.20. The numbers of groups in the AMMS data category are 3 categories, namely "high", "medium", and "low" groups.

SUMMARY OF 172 MEASURED Person

	TOTAL SCORE	COUNT	MODEL MEASURE	INFINIT S.E.	OUTFIT MNSQ	ZSTD	MNSQ	ZSTD
MEAN	28.1	11.0	.09	.53	1.00	-.2	1.01	-.2
P.SD	5.9	.0	1.61	.07	.65	1.5	.66	1.5
S.SD	5.9	.0	1.61	.07	.65	1.5	.66	1.5
MAX.	43.0	11.0	5.36	1.06	3.79	4.2	3.72	4.2
MIN.	13.0	11.0	-4.27	.48	.14	-3.3	.14	-3.2
-----								
REAL RMSE	.60	TRUE SD	1.49	SEPARATION	3.20	Person RELIABILITY	.89	
MODEL RMSE	.53	TRUE SD	1.52	SEPARATION	3.40	Person RELIABILITY	.91	
S.E. OF Person MEAN = .12								

Figure 1: Rasch summary statistics from amms survey

Through the index separation, the researcher can group data using the mean and standard deviation.

The categorization using this approach can be seen in Table 2.

Table 2 Data categories are based on the Logit Interval and Score Interval on AMMS (n=172)

Interval Logit	Interval Score	Category
> +1.61	>34	High
-1.61 - +1.61	22,2 – 34	Moderate
< -1.61	<22,2	Low

\*Note: for 1 – 5 scale; 11 items; 5 categories

The explanation in Table 2 can be interpreted that the grouping of data carried out by referring to the norm standard is not too much different than the benchmark on the ideal score. The illustration of the grouping category is presented in Figure 2.

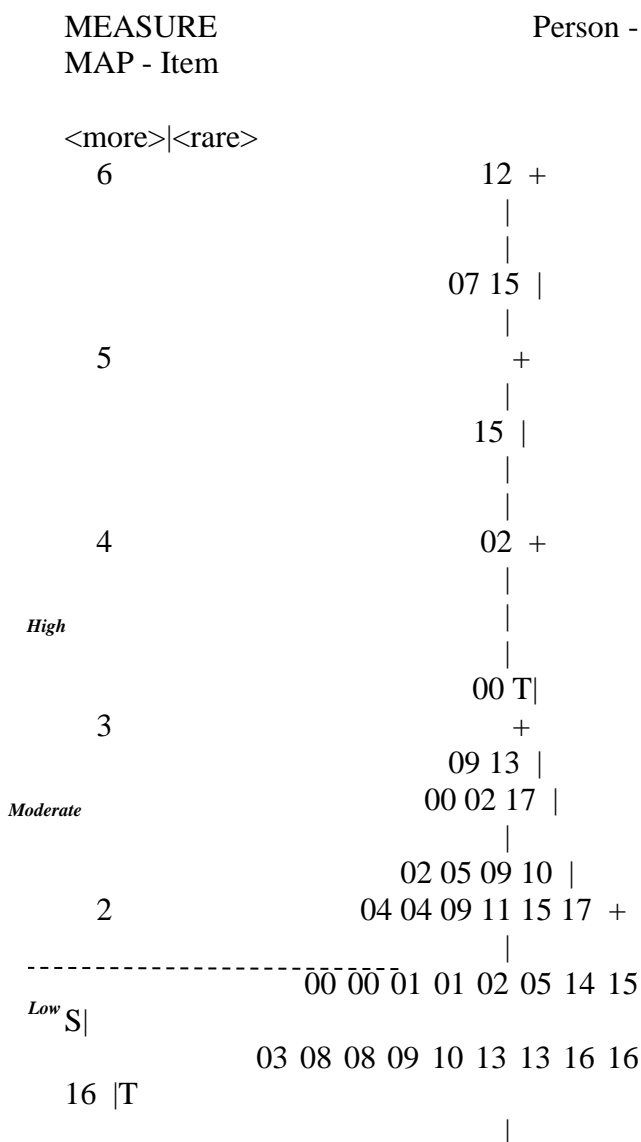
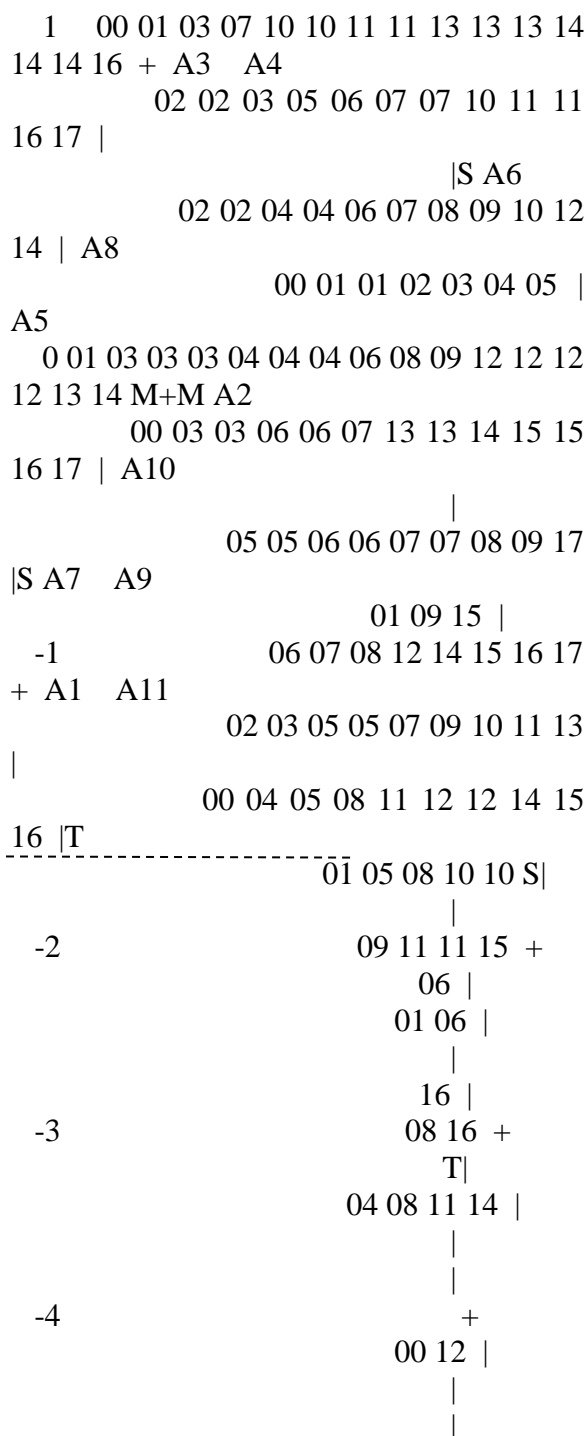




Figure 2 The example of categories from Rasch perspective

The exposure in Figure 2 shows the division of categories based on the standard deviation value. The data distribution is displayed in the form of a normal curve in the person graph. This curve is then divided at point +1SD (+1.61) to -1SD (-1.61) so that respondents whose logit values are in this range are categorized as respondents with "moderate" achievement. Transforming this value to raw score can also be seen in the table, that in the category "moderate" is the respondent with raw score 22.2 to 34. The condition of the data with this norm reference will also change according to the conditions of empirical data and the quality of the measuring instrument itself. Therefore, it is important for researchers to determine what approach will be used in determining the process of grouping and categorizing research data (Ardi, 2017; D Daharnis, et al., 2018) (Ardi, et al., 2018).

### Conclusion

In determining the category of data distribution can be used two standards, namely criterion-referenced measurement and norm-referenced measurement. The selection of these types of benchmarks is based on the theoretical basis and research objectives. If the researcher wants to compare the position of the respondent with ideal conditions, then criterion-referenced measurement is used. Meanwhile, if researchers want to compare positions between respondents, norm-referenced measurement is used. For the purposes of evaluation, intervention and treatment, it will be better if criterion-referenced measurement is used. However, if for the sake of selection it is used as a standard norm-referenced measurement.

### References

[1] Ardi, Z. (2017). *Unsuitable Majoring: Does the Reorientation Would Help the Student for Revitalize Learning Activities?* Paper

presented at the 9th International Conference for Science Educators and Teachers (ICSET 2017).

- [2] Ardi, Z., Sukmawati, I., Afdal, A., Rangka, I. B., & Suranata, K. (2018). Exploring the acceptability of internet-based mental health mobile app services using network psychometrics analysis.
- [3] Buckley, P., Moore, B., Boardman, A. G., Arya, D. J., & Maul, A. (2017). Validating a fidelity scale to understand intervention effects in classroom-based studies. *American Educational Research Journal*, 54(6), 1378-1413.
- [4] Busk, P. L., & Marascuilo, L. A. (2015). Statistical analysis in single-case research. *Single-Case Research Design and Analysis (Psychology Revivals): New Directions for Psychology and Education*, 159.
- [5] Coolican, H. (2017). *Research methods and statistics in psychology*: Psychology Press.
- [6] Daharnis, D., & Ardi, Z. (2016). The compatibility student choice of university majoring; A preliminary studies. *GUIDENA: Jurnal Ilmu Pendidikan, Psikologi, Bimbingan Dan Konseling*, 6(1), 101-109.
- [7] Daharnis, D., Ardi, Z., Alizamar, A., Ifdil, I., Rangka, I., & Suranata, K. (2018). *Adaptation and validation of mathematics anxiety: Rasch and network psychometrics analysis*. Paper presented at the Journal of Physics: Conference Series.
- [8] Daharnis, D., Erlamsyah, E., Ifdil, I., Ardi, Z., & Hariko, R. (2014). *Gambaran kegiatan belajar siswa Sumatera Barat*. Paper presented at the Seminar Internasional Bimbingan dan Konseling.
- [9] Dahlke, J. A., Kostal, J. W., Sackett, P. R., & Kuncel, N. R. (2018). Changing abilities vs. changing tasks: Examining validity degradation with test scores and college performance criteria both assessed longitudinally. *Journal of Applied Psychology*, 103(9), 980.
- [10] de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a

- Bayesian network example. *Computers & education*, 85, 23-34.
- [11] Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- [12] Estebaranz, J. M. A. (2017). La evaluación de los aprendizajes. Problemas y soluciones. *Profesorado. Revista de Currículum y Formación de Profesorado*, 21(4), 381-404.
- [13] Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: an introduction and illustrative example. *International journal of nursing studies*, 46(3), 380-393.
- [14] Heinrich, M., Lardos, A., Leonti, M., Weckerle, C., Willcox, M., Applequist, W., et al. (2018). Best practice in research: consensus statement on ethnopharmacological field studies—ConSEFS. *Journal of ethnopharmacology*, 211, 329-339.
- [15] Idris, H., Nurhayati, N., & Satriani, S. (2018). Developing Computer-Assisted Instruction Multimedia For Educational Technology Course of Coastal Area Students. *E&ES*, 156(1), 012049.
- [16] Kim, H., Ad, V. B., McAna, J., & Dicker, A. P. (2016). Evaluating the quality, clinical relevance, and resident perception of the radiation oncology in-training examination: A national survey. *Practical radiation oncology*, 6(1), 44-49.
- [17] Linacre, J. (2009). A User's guide to Winsteps-ministep: Rasch-model computer programs. Program Manual 3.68. 0. Chicago, IL.
- [18] Lipsey, M. W., Nesbitt, K. T., Farran, D. C., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Learning-related cognitive self-regulation measures for prekindergarten children: A comparative evaluation of the educational relevance of selected measures. *Journal of Educational Psychology*, 109(8), 1084.
- [19] Meyer, J. P., Doromal, J. B., Wei, X., & Zhu, S. (2017). A criterion-Referenced approach to student ratings of instruction. *Research in Higher Education*, 58(5), 545-567.
- [20] Nahhas, S., Bamasag, O., Khemakhem, M., & Bajnaid, N. (2018). Added values of linked data in education: A survey and roadmap. *Computers*, 7(3), 45.
- [21] Rose, R., Shevlin, M., Winter, E., & O'Raw, P. (2015). Project IRIS-Inclusive Research in Irish Schools: A Longitudinal Study of the Experiences of and Outcomes for Children with Special Educational Needs (SEN) in Irish Schools.
- [22] Santelli, J. S. (2016). Development and Psychometric Testing of the Criterion-Referenced Measurement Tool for Genetics. *Journal of Nursing Measurement*, 24(1), 69-91.
- [23] Schumacher, D. J., Michelson, C., Poynter, S., Barnes, M. M., Li, S.-T. T., Burman, N., et al. (2018). Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Medical Teacher*, 40(1), 70-79.
- [24] Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*: John Wiley & Sons.
- [25] Seo, D., Taherbhai, H., & Paek, I. (2013). A multidimensional diagnostic perspective on academic achievement goal orientation structure, using the Rasch measurement models. *Journal of applied measurement*, 14(2), 129-148.
- [26] Shapiro, E. S., Gebhardt, S., Flatley, K., Guard, K. B., Fu, Q., Leichman, E. S., et al. (2017). Development and validity of the Rating Scales of Academic Skills for Reading Comprehension. *School Psychology Quarterly*, 32(4), 509.
- [27] Škoda, J., Doulík, P., Bílek, M., & Šimonová, I. (2016). Learning style as a factor influencing the effectiveness of the inquiry-based science education at lower secondary schools. *Journal of Baltic Science Education*, 15(5), 588.
- [28] Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*: Trim Komunika Publishing House.

- [29] Thompson, M. S., & McDonald, S. (2016). Race, skin tone, and educational achievement. *Sociological Perspectives*, 59(1), 91-111.
- [30] Triola, M. F. (2013). *Elementary statistics using Excel*: Pearson.
- [31] Walliman, N. (2015). *Social research methods: The essentials*: Sage.
- [32] Xiao, Y., Bowen, N. K., & Lindsey, M. A. (2018). Racial/ethnic measurement invariance of the School Success Profile (SSP)'s future orientation scale. *Journal of school psychology*, 71, 85-107.