A Rasch model analysis on the ability to analyze descriptive statistics: A case Study of FMIPA Universitas Syiah Kuala new students

R Ferdhiana¹, I Santigana¹, F Abdurrahman¹ and Nurhasanah¹

¹ Universitas Syiah Kuala, Mathematics and Natural Sciences Faculty, Department of Statistics, Kopelma Darussalam, Banda Aceh, Indonesia

E-mail: ridha.ferdhiana@unsyiah.ac.id.

Abstract. Data literacy skills are essential in the current industrial 4.0 era. Data literacy skills include the ability to read and analyze data in a descriptive form. This study aims to measure the ability to analyze descriptive data and analyze the statistical analysis results of responses from new students of the Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala. The test instrument consisted of 29 questions divided into three parts: understanding, application, and analysis. The tool was tested with a time limit of 30 minutes to 436 respondents from 7 study programs at the Faculty of Mathematics and Natural Sciences. The average score obtained by students was 67.66, with a standard deviation of 12.48. The implementation of Rasch model provides a difficulty index ranging from -3.80 to 2.45, with a difficulty average approaching zero and a standard deviation of 1.4. The result concluded that the respondent's ability to answer the questions is average.

Keywords: Rasch model, difficulty index, data literacy, descriptive data test

1. Introduction
Industry 4.0 is here and is believed to have a significant influence on human life today. Industry 4.0 has excellent potential to increase productivity levels, reduce costs, and significantly improve the quality of products produced [1]. Indonesia is a country that is in the process of planning to implement Industry 4.0 with the launch of a program called "Making Indonesia 4.0" [2]. The application of industry 4.0 requires Indonesia to prepare human resources, which, in this case, will affect the world of education. Reference [3] states that hard skills alone are not enough for industry 4.0; graduates must be able to adapt to good soft skills.

The necessary capabilities that must be possessed to be able to compete in industry 4.0 are understanding IT infrastructure (cloud computing), reaching a consensus for decision making (central data and analytics), changing processes (business process management), connecting from production to management (MES integration), analyzing performance, as well as working together (in the work ecosystem) [4]. Meanwhile, according to [5], companies in the industrial era 4.0 need competent experts in network issues (IT and IoT), design and data experts (scientists), software developers and...
programmers, and cyber security. Data literacy skills are always expected in the industrial 4.0 era. Data literacy can be seen in the competencies of Big Data, Problem Solving, Data/Network Security, and Decision Making.

In the Industry 4.0 era, an understanding of how decisions are informed by data is an essential ability required in the global knowledge ([4], [6]). Data literacy skill can be interpret as the ability to collect, manage, evaluate, and apply data, in a critical manner to support of evidence based decision making ([6]-[7]). Data literacy follows the general concept of literacy but focuses on the competencies involved in working with data. The main difference from other literacy, for example, text literacy, which only emphasizes the ability to read a text, data literacy requires specific skills that involve reading and understanding data, which is also known as information.

According to Rumsey [7], students who pass the Introductory Statistics course do not necessarily have good data literacy skills. Furthermore, Rumsey [7] states that the Introductory Statistics course's success is when someone can question, trace problems, compare with other results, explain, and evaluate these results at a higher level. Almost all undergraduate (S1) study programs at Universitas Syiah Kuala get Basic Statistics or Introductory Statistics or Statistical Methods. Even though it uses many names, this course's essence is to provide tools and methods for hypothesis testing that are indispensable in any research activity. Hypothesis testing is closely related to the data so that all Universitas Syiah Kuala students should have reasonably good data literacy skills.

Since the 2016/2017 school year, the Faculty of Mathematics and Natural Sciences Unsyiah has implemented the same competency standards for the Introductory Statistics course that is taught in the first year for the new student. The purpose is to get all FMIPA students the same lesson plans, materials, learning methods, teaching materials and practicum, assessments, and exam questions. The average final score for the Introductory Statistics course for FMIPA students in the 2017/2018 school year is 61.96, with a variance of 22.65 and a median of 67.42, while for the 2018/2019 school year, the statistics are 67.05, 18.41, and 67.11, respectively. The mean and median show that the students' final letter grades are around BC (a little bit above C). The standard deviation indicates a relatively broad range of values, with the scores ranging from 35 to 84. Similar results will be found in study programs outside FMIPA. These results indicate that new students' understanding of data and data manipulation at the lower tertiary education level is insufficient.

This study aims to see the new students' data literacy skills where the students have not taken the introductory statistics course yet. The study intends to measure their ability to read, understand, analyze, and interpret data in a descriptive form, which is considered a necessary foundation of data literacy skills. This literacy ability will be measured using a questionnaire in the form of questions related to data literacy at the basic concepts, knowledge of collecting data, data management, and data evaluation, which are the Introduction to Statistics course's main objectives. The questionnaire results were analyzed using the RAasch method to get the questions' difficulty level, person fit, and item fit. The research results will be used to improve the content and level of competence of the course.

2. Research Methodology

This research begins by designing a questionnaire that contains questions in the form of statistical descriptions. A pilot test was then conducted on the intended questionnaire, and after passing the validity and reliability test, applied to a larger respondent. In this article, those testing processes are not shown because we emphasize the Rasch method's results. The Rasch method outputs are difficulty index, item fit, plot of the person item map, and item characteristic curves. Based on those outputs, we will be able to detect the weakness of the respondent's ability.
2.1. The Assessment Tool and respondents

In contrast to statistical literacy, data literacy emphasizes understanding of data and its meaning, including the ability of reading graphs and charts and conclude them [8]. As mention above, the purpose of this study is to measure their ability to read, understand, analyze, and interpret data in a descriptive form; thus, the questions were developed based on three common topics and three skills. The questionnaire consisted of 29 questions about statistical descriptions and divided among those topics and skills, as shown in table 1.

| Topic                          | Understanding | Application | Analysis |
|-------------------------------|---------------|-------------|----------|
| One Variable Diagram          | 1, 2, 27, 28  | 17, 26, 29  | 14, 15, 16, 18 |
| Two Variable Diagram          | 4, 7, 10, 19, 23 | 5, 6, 9, 11, 21, 22 | 3, 8, 12, 13, 20, 24, 25 |

This study's respondents are new students in the Faculty of Mathematics and Natural Sciences Universitas Syiah Kuala. All respondents are in their first term in University with an assumption that they do not have a college-level course in Statistics. With this assumption, we expect that their ability to read data is considered primary. The questionnaire is distributed to 436 respondents from 7 study programs, Mathematics, Physics, Chemistry, Biology, Computer Science, Statistics, and Pharmacy.

2.2. Difficulty Index

The Rasch model is the Item Response Theory (IRT) model with one parameter: the level of difficulty. The Rasch model equation is given by equation (1) where \( \theta \) is the respondent's ability, and \( \beta \) is the item's level of difficulty.

\[
P_i(X_i = 1) = \frac{e^{(\theta - \beta_i)}}{1 + e^{(\theta - \beta_i)}},
\]

The difficulty level is a measure of the difficulty of the question in the eyes of the respondent. In general, the difficulty level ranges from -3 to 3, where a negative score indicates that the problem is an easy one ([9]-[10]). The questions with a difficulty level below -3 or above 3 are questions that can be eliminated because they are very easy or challenging questions.

2.3. Item Fit

The Rasch model uses the fit statistics to measure or show the magnitude of the data's prediction accuracy ([11]-[12]). Fit statistics also indicate whether an item should be deleted, reconstructed, or rewritten [13]. In general, fit statistic is given by two values, namely the mean squared mean (MNSQ) and the z-standard (ZSTD) [14]. MNSQ is the mean of the squared residuals for an item, while ZSTD is transforming the mean square value with a sample size correction, commonly referred to as the standardized statistic or z-score. The expected value of the MNSQ is a value close to 1.0 (one) because it is a chi-square statistic divided by the degrees of freedom. An MNSQ value of more than 1.0 indicates that the item is under fitted and indicates some noise that is not apply in the model, or there are other
sources of variance in the data. Overfit is when the MNSQ value is less than 1.0, indicating that model predicts the data very well. The t-statistic is often used as the ZSTD when the degrees of freedom are limited. Student's t-statistic values were adjusted to standard normal values, and using 95% confidence intervals would be in the range from -1.96 to 1.96 [9].

Inlie-patter-sensitive fit statistic (infit) and outlier-sensitive fit statistic (outfit) are based on the chi-square statistic. Infit is sensitive to unexpected patterns because it is based on each observation weighted by the model's variance. Outfits are more sensitive to unforeseen observations, for example, items that are too easy or too difficult. Outfit is based on conventional chi-square statistic. Reference [15] guides the interpretation of the infit, clothing of the MNSQ values, as shown in table 2.

Table 2. Interpretation of parameter-level mean-square fit statistics.

| Mean-Square | Implication for Measurement                                           |
|------------|-----------------------------------------------------------------------|
| > 2.0      | Distorts or degrades the measurement system                           |
| 1.5 – 2.0  | Unproductive for construction of measurement, but not degrading       |
| 0.5 – 1.5  | Productive for measurement                                            |
| < 0.5      | Less productive for measurement, but not degrading. May produce       |
|            | misleadingly good reliabilities and separations.                      |

2.4. Item Characteristic Curves (ICC)

In the Item Response Theory, the item characteristic curve illustrates the relationship between latent ability and test item performance. The formed curve is a description of the possibility of the respondent's ability to answer the item correctly. The latent ability is plotted on the x (horizontal) axis and the value on the right of the threshold (zero) requires a higher ability to answer the item correctly. The probability of answering correctly (X = 1) for each item is plotted on the y (vertical) axis. Items with a high difficulty level will be below the threshold (0.5), while the easy items will be above the threshold.

3. Results and Discussion

Before we discuss the analysis results with the Rasch model, we presented the questionnaire results in figure 1a and figure 1b, which consisted of a histogram and a boxplot. Based on the histogram image, it can be seen that the score obtained by the respondent is skewed to the left, meaning that there are more respondents with a score above the mean than respondents who get a value below it. From the boxplot, it can be seen that there are two bottom outliers, meaning two respondents have a meagre ability in understanding descriptive data.

![Figure 1a. Histogram of the questionnaire result.](image1.png)

![Figure 1b. Boxplot of the questionnaire result.](image2.png)
The summary of the questionnaire results, the minimum, maximum, mean, median, and standard deviation scores were 24.1379, 93.1034, 67.6575, 68.9655, and 12.4796. Although more scores are above the mean, but because the mean's value is not sufficient, just about the letter grade C, overall, the result is not satisfying. The big standard deviation, 12.4796, indicates that the scores are quite widespread.

Table 3. Item Difficulty Parameters (Beta) with 0.95 CI.

| Beta | Estimate | Std. Error | Lower CI | Upper CI |
|------|----------|------------|----------|----------|
| Q1   | -2.976   | 0.328      | -2.332   | -3.619   |
| Q2   | -3.804   | 0.487      | -2.850   | -4.759   |
| Q3   | -0.673   | 0.132      | -0.415   | -0.930   |
| Q4   | -2.312   | 0.243      | -1.836   | -2.788   |
| Q5   | 0.633    | 0.103      | 0.834    | 0.432    |
| Q6   | -1.215   | 0.157      | -0.908   | -1.522   |
| Q7   | 0.291    | 0.107      | 0.500    | 0.082    |
| Q8   | -0.397   | 0.122      | -0.158   | -0.637   |
| Q9   | -0.281   | 0.119      | -0.048   | -0.514   |
| Q10  | 0.984    | 0.101      | 1.182    | 0.787    |
| Q11  | 1.351    | 0.101      | 1.549    | 1.152    |
| Q12  | -0.041   | 0.113      | -0.262   | 0.180    |
| Q13  | 2.448    | 0.120      | 2.683    | 2.213    |
| Q14  | 0.704    | 0.102      | 0.904    | 0.505    |
| Q15  | -0.800   | 0.137      | -0.533   | -1.068   |
| Q16  | 2.434    | 0.119      | 2.668    | 2.200    |
| Q17  | 0.423    | 0.105      | 0.628    | 0.217    |
| Q18  | -1.094   | 0.150      | -0.799   | -1.388   |
| Q19  | -0.474   | 0.125      | -0.230   | -0.718   |
| Q20  | -0.003   | 0.112      | -0.223   | 0.216    |
| Q21  | -0.041   | 0.113      | -0.262   | 0.180    |
| Q22  | 1.605    | 0.103      | 1.402    | 1.808    |
| Q23  | 0.223    | 0.108      | 0.012    | 0.434    |
| Q24  | 0.925    | 0.101      | 0.727    | 1.123    |
| Q25  | -0.041   | 0.113      | -0.262   | 0.180    |
| Q26  | 0.745    | 0.102      | 0.546    | 0.944    |
| Q27  | 1.291    | 0.101      | 1.092    | 1.489    |
| Q28  | -0.412   | 0.123      | -0.653   | -0.172   |
| Q29  | 0.508    | 0.104      | 0.305    | 0.711    |

The item difficulty index is an estimate of the Beta value in the Rasch model equation. Index with negative value shows that the questions are seen as easy by the respondent. Meanwhile, the positive index means that the problem is difficult. A good range of index values is between -3 to 3. Questions with an index value below -3 or more than 3 are questions that can be eliminated by considering other factors. Table 3 provides the difficulty index's magnitude along with the standard error, lower limit, and upper limit at the 95% interval. The mean of Betas is 0.10, with a 1.30 standard deviation. There is one question outside the boundaries of -3 to 3, the second question (Q2). The number of questions with a
negative index is 15, thus can be concluded that the questions are fair enough in term of its difficulty. Other information that can be obtained from table 3 is that question 2 (Q2) tends to be eliminated. Easy questions (index value close to -3) are Q1 and Q4, while difficult questions (close to 3) are Q13 and Q16. The other questions are in the range -2 to 2, which is a fair zone for the questions' difficulty level.

Table 4. Item Fit Based on MNSQ and t.

| Item | Chisq   | df | p-value | Outfit MNSQ | Infit MNSQ | Outfit t | Infit t | Discrim |
|------|---------|----|---------|-------------|------------|----------|--------|---------|
| Q1   | 167.955 | 435| 1       | 0.385       | 0.886      | -1.828   | -0.292 | 0.283   |
| Q2   | 586.644 | 435| 0       | 1.346       | 0.937      | 0.704    | 0.017  | -0.021  |
| Q3   | 363.29  | 435| 0.995   | 0.833       | 0.954      | -1.476   | -0.542 | 0.287   |
| Q4   | 354.681 | 435| 0.998   | 0.813       | 0.943      | -0.579   | -0.205 | 0.137   |
| Q5   | 460.235 | 435| 0.194   | 1.056       | 1.049      | 1.185    | 1.378  | 0.194   |
| Q6   | 375.37  | 435| 0.982   | 0.861       | 0.946      | -0.849   | -0.445 | 0.248   |
| Q7   | 372.278 | 435| 0.987   | 0.854       | 0.901      | -2.558   | -2.339 | 0.436   |
| Q8   | 389.036 | 435| 0.945   | 0.892       | 0.975      | -1.111   | -0.341 | 0.27    |
| Q9   | 423.866 | 435| 0.64    | 0.972       | 1.004      | -0.281   | 0.09   | 0.204   |
| Q10  | 479.417 | 435| 0.069   | 1.1         | 1.083      | 2.434    | 2.613  | 0.127   |
| Q11  | 451.072 | 435| 0.287   | 1.035       | 1.012      | 0.822    | 0.397  | 0.234   |
| Q12  | 475.987 | 435| 0.085   | 1.092       | 1.036      | 1.192    | 0.669  | 0.168   |
| Q13  | 528.299 | 435| 0.001   | 1.212       | 1.037      | 2.209    | 0.61   | 0.005   |
| Q14  | 430.436 | 435| 0.553   | 0.987       | 0.992      | -0.272   | -0.212 | 0.297   |
| Q15  | 414.652 | 435| 0.751   | 0.951       | 0.923      | -0.351   | -0.864 | 0.283   |
| Q16  | 453.65  | 435| 0.259   | 1.04        | 1.01       | 0.473    | 0.179  | 0.112   |
| Q17  | 420.477 | 435| 0.683   | 0.964       | 0.984      | -0.648   | -0.399 | 0.291   |
| Q18  | 368.395 | 435| 0.991   | 0.845       | 0.902      | -1.036   | -0.924 | 0.315   |
| Q19  | 554.434 | 435| 0       | 1.272       | 1.077      | 2.467    | 1.066  | 0.034   |
| Q20  | 471.124 | 435| 0.112   | 1.081       | 1.079      | 1.081    | 1.472  | 0.102   |
| Q21  | 359.847 | 435| 0.996   | 0.825       | 0.892      | -2.41    | -2.038 | 0.457   |
| Q22  | 444.28  | 435| 0.369   | 1.019       | 0.985      | 0.399    | -0.418 | 0.219   |
| Q23  | 365.105 | 435| 0.994   | 0.837       | 0.893      | -2.721   | -2.414 | 0.459   |
| Q24  | 370.84  | 435| 0.988   | 0.851       | 0.872      | -3.886   | -4.222 | 0.487   |
| Q25  | 356.181 | 435| 0.998   | 0.817       | 0.902      | -2.536   | -1.841 | 0.424   |
| Q26  | 421.638 | 435| 0.668   | 0.967       | 0.98       | -0.749   | -0.592 | 0.301   |
| Q27  | 490.542 | 435| 0.034   | 1.125       | 1.102      | 2.93     | 3.135  | 0.073   |
| Q28  | 425.595 | 435| 0.617   | 0.976       | 0.932      | -0.211   | -0.983 | 0.314   |
| Q29  | 479.321 | 435| 0.07    | 1.099       | 1.052      | 1.909    | 1.359  | 0.191   |

Misfit is a situation where an item does not fit. Measurements that are seen to see the misfit are infit and outfit. If there are an MNSQ infit value and an MNSQ outfit that is more than 2, then the item is declared unfit. Meanwhile, the value of infit t and outfit t must be in the range -1.96 and 1.96. Table 4 is the result of calculating the infit and output values for both MNSQ and t. Underlined values are those that are outside the fit limit. It can be seen that items Q7, Q10, Q21, Q23, Q24, and Q27, based on the
t-statistic, are in the misfit area. However, these six items are still within the reasonable limits of MNSQ. Item Q2 based on MNSQ and t are within fit limits, so they are not excluded from the questionnaire.

Person Item Map in figure 2 shows that the questions on the left of the latent dimension -2 are Q2, Q1, and Q4. Item Q2, based on the difficulty index, is an item that can be eliminated, but the MNSQ and t infit and outfit do not suggest this. Item Q1 and Q4 are not items that are outside the fit range. We concluded that the three questions are not problematic. The questions that are in the right of latent dimension 2 are Q13 and Q16; both questions have no problem with the difficulty index or fit item. Fig. 3 is a picture of the Item Characteristic Curve of all the Questions. We can see that Q2, Q1, and Q4 are the fastest to P (X = 1), while Q16 is the slowest item. The rest of the items appear normal with an almost similar curve. Based on all the curves, we can say that the respondent's response to the questions is average.

![Person-Item Map](image1)

**Figure 2.** Person-Item Map.

![Item Characteristic Curve](image2)

**Figure 3.** Item Characteristic Curve of all the Questions.
4. Conclusion
The questionnaire results showed that more students scored more than the mean, even though two respondents scored very low. Based on the Rasch model analysis, one question is obtained outside the difficulty index range, six are outside the range of the t-statistic fit, and none is misfit based on the MNSQ. Items with a negative difficulty index were 15 out of 29 questions. It can be concluded that the questionnaire built tends to be easy, but the average respondent's ability is still in the range of C values.

Acknowledgment
The author would like to thank Syiah Kuala University for providing research funding through the 2020 Lector Research scheme. This research would not have been possible without discussion from the Computation and Applied Statistics research Department of Statistics Universitas Syiah Kuala.

References
[1] Khan A and Turowski K 2016 A Perspective on Industry 4.0: From Challenges to Opportunities in Production Systems In The Proceedings of the International Conference on Internet of Things and Big Data (IoTBD 2016) 441–448.
[2] Hidayatno A, Rahmawan A and Hulu C 2019 Industry 4.0 Technology Implementation Impact to Industrial Sustainable Energy in Indonesia: A Model Conceptualization Energy Procedia 156 227–233
[3] Baygin M, Yetis H, Karakose M and Akin E 2016 An effect analysis of industry 4.0 to higher education In The Proceedings of 5th International Conference on Information Technology Based Higher Education and Training (ITHET) 1–4
[4] Six skills needed to successfully roll-out Industry 4.0 pilot projects | ITProPortal [Online] Accessed: 17-Oct-2019 https://www.itproportal.com/features/six-skills-needed-to-successfully-roll-out-industry-40-pilot-projects/
[5] Must-Have Skills for Industry 4.0 [Online] Accessed: 17-Oct-2019 https://www.controlglobal.com/blogs/guest-blogs/must-have-skills-for-industry-4-0/
[6] Risdale C, Rothwell J, Smit M, Bliemel M, Irvin D, Kelley D E, Matwin S S, Wuetherick B and Ali-Hassan A 2015 Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report Key Message (Dalhause University)
[7] Rumsey D J 2002 Statistical literacy as a goal for introductory statistics courses J. Stat. Educ., 10 3
[8] Carlson J and Johnston L 2015 Data Information Literacy : Librarians, Data and the Education of a New Generation of Researchers (West Lafayette, IN: Purdue University Press)
[9] Boone W J 2016 Rasch analysis for instrument development: Why,when,and how? CBE Life Sci. Educ. 15 4
[10] Demars C E 2017 Classical test theory and item response theory In The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development 1–2 pp. 49-73 (John Wiley & Sons, Ltd.)
[11] What do Infit and Outfit, Mean-square and Standardized mean? [Online] Accessed: 06-Nov-2020 https://www.Rasch.org/rmt/rmt162f.htm.
[12] Rasch Analysis for Instrument Development: Why, When, and How? | Enhanced Reader [Online] Accessed: 04-Nov-2020 www.lifescied.org
[13] Christensen K B and Kreiner S 2013 Item Fit Statistics Rasch Model. Heal. 83–104
[14] HumBond T G and Fox C M 2007 Applying the Rasch Model : Fundamental Measurement in the Human Sciences Second Edition (University of Toledo)
[15] Linacre J M 2002 What do infit and outfit, mean-square and standardized mean Rasch Meas. Trans. 16 2 878