Exploration of Students' Statistical Reasoning Ability in the Context of Ethnomathematics: A Study of the Rasch Model

Rahmi Ramadhani[1,2], Sahat Saragih[1], E. Elvis Napitupulu[1]

[1]Postgraduate Program, Universitas Negeri Medan, Medan, Indonesia, [2]Universitas Potensi Utama, Medan, Indonesia

rahmiramadhani3@gmail.com, saragihpps@gmail.com, elvisnapit@gmail.com

Abstract: *Statistical reasoning ability is one of the essential skills in developing competence, which is one of the Sustainable Development Goals (SDGs). This study aims to explore the statistical reasoning ability of junior high school students in descriptive statistics learning. The investigation directs students to determine their level of statistical reasoning ability. The Rasch model analysis analyzes whether there is a bias in the item questions used in exploring these abilities based on demographic factors such as gender, initial mathematical ability, and ethnic background. This study uses a mix-method (a combination of quantitative and qualitative methods). Based on the Rasch Model analysis, quantitative and qualitative data were obtained from 22 students utilizing a statistical reasoning ability test instrument based on the ethnomathematics context, which was valid and reliable. The results of data analysis show that students have different levels of statistical reasoning abilities and are related to students' initial mathematical ability. Students with low initial mathematical skills are at level 1 idiosyncratic, while students with high initial mathematical skills are at procedural level 4. The results of the DIF (Differential Item Functioning) analysis showed no significant difference between the use of the ethnomathematics context-based test instrument and the demographic factors of the research subjects. This study concludes that students can explore statistical reasoning ability through ethnomathematics-based problem solving but still pay attention to other demographic factors.*

Keywords: *Statistical Reasoning Ability, Ethnomathematics, Descriptive Statistics, Rasch Model*

INTRODUCTION

Currently, the development of society in a country is supported by the development of advances in Science and Technology. This is in line with the emergence of the Industrial Revolution 4.0

which brings innovation and technology as part of the development of life. The factor that gives the biggest role in the advancement of science and technology is the factor of educational development. The education factor is also included in one of the Sustainable Development Goals (SDGs) which are used as indicators for countries in their efforts to develop society and sustainable economic development. Education provides space for people to develop innovation and technology in higher-order thinking processes. This can be seen with the increasing importance of understanding data which will later be applied in the development of technology-based on the Internet of Things (IoT) and Artificial Intelligence (Yaqoob et al., 2019). Understanding the data ultimately leads to the development of analytical skills on data that is closely related to statistics. Data analysis using statistics aims to filter information that appears massively so that later it can be concluded that the right and significant information is used as further recommendations regarding the development of a system (Setiawan & Sukoco, 2021).

Based on this situation, the education factor clearly has an important role in creating a society that is critical of information, quick to collect valid data, so that later it can test and analyze valid data for conclusions and applications in a broad scope. This important role requires good statistical reasoning ability. In this context, statistical reasoning ability is defined as a general conclusion expressed with uncertainty but can be proven using available data sets (Ben-Zvi et al., 2015).

Statistical reasoning emphasizes how students use their reasoning in gathering information taken from statistical random data (Pfannkuch, 2006). Statistical reasoning also helps students understand statistical data and can describe the data obtained in various forms of data presentation, interpret data to draw conclusions from the collection of statistical data obtained (Pfannkuch, 2005; Rubin et al., 2006). Statistical reasoning ability also focuses on how to find the meaning of the statistical data set that has been identified, whether there are data deviations, and then decide whether the data analysis result can be concluded or not (Makar, 2013). The importance of statistical reasoning skills based on the key concepts of NCTM did not yield positive outcomes in field studies. Facts on the ground show that high school students still have difficulty describing data in various forms of data presentation, as well as reasoning and statistical thinking processes in the context of statistical problems (Ramadhani & Fitri, 2020; Ramadhani & Narpila, 2018).

The results of this study are also supported by other findings, where students still make mistakes in answering questions related to the size of the central value, variability, and standard deviation (Chan et al., 2016; Rufiana et al., 2018) which have an impact on students' low reasoning abilities related to literacy and mathematics (especially in the context of statistics). The level of statistical reasoning ability of students on average is still at the level of idiosyncratic reasoning (Level 1) (Chan et al., 2016; delMas, 2002; Garfield, 2002). Students at Level 1 can employ a variety of statistical terms and symbols, but they cannot fully interpret and apply them to the correct data. As a result, their responses are frequently incorrect. For example, such students may be familiar with the term' standard deviation,' but cannot correctly apply it (Chan et al., 2016). In general, idiosyncratic, level 1 replies are not helpful in classifying thinking. For example, a student's argument would be classed as idiosyncratic if it was entirely focused on the context of the work and did not refer to the data. Based on the confirmation, we found the results of the world assessment where Indonesia is ranked in the bottom 6 (72 out of 79 countries) for reading, mathematics, and science assessments (PISA 2018), while for the 2015 TIMSS assessment, Indonesian students did not achieve the average score. determined by TIMSS (TIMSS Scale Centrepoint) which is 397 out of 500 points on average (Mullis et al., 2016; OECD, 2019).

Another fact is that most students in Indonesia, namely 75.7% of students do not reach Level 2 which is set as the basic level in understanding mathematical literacy. These students can only solve math problems using familiar contexts, clear questions, and present all relevant information. On the other hand, students are actually able to identify relevant information and carry out routine mathematical procedures only if explicit instructions are given (Zulkardi et al., 2020). Based on this, students at Level 2 (Statistical Reasoning Level) are better able to employ a variety of statistical phrases and symbols since they are more familiar with the meanings of various statistical ideas. They still don't know how to use them properly. Students may, for example, select the proper quartile explanation yet fail to answer conceptual questions. These findings and facts are reinforced by the statement of Maryati & Priatna (2018) where students show difficulties in reading the statistical data presented, understanding statistical problems, and solving problems related to the average value of a data. Likewise, the results obtained by Mahdayani (2016) where as many as 54.6% of high school students have difficulty in reading and understanding statistical data; 83.5% of students have difficulty in transforming data; and as many as 91.7% of students have difficulty in processing data and concluding the data presented. Based on the result, we can find of one possible explanation is that students tend to memorize so it has an impact on the process of understanding and statistical reasoning (Nuralam & Gadeng, 2018). Statistical reasoning for high school students is an important skill, because it affects the understanding process in other scientific fields (English, 2014; Sharma, 2017). Students' difficulties in reasoning statistical problems will affect students' abilities in spatial and numerical terms. Students will feel anxious and confused when faced with simple numerical problems, describe spatial problems to count objects on a graphic display which is the basis of statistical reasoning abilities (Ching et al., 2020).

Based on the facts and findings obtained, it is necessary to diagnose what makes students tend to have low statistical reasoning abilities. In response to this, we can explore students' statistical reasoning abilities using five levels of statistical reasoning ability in descriptive statistics based on the statistical reasoning model initiated by Garfield (2002), namely idiosyncratic (level 1), verbal (level 2), transitional (level 3), procedural (level 4), and investigative process reasoning (level 5). Students with level 1 and 2 statistical reasoning skills continue to struggle with symbols, terminology, statistical concepts, and their application in real-life situations. They can understand one or two components of the statistical process, but they cannot use these principles to get answers, unlike students at level 3. Such students, for example, can recognize the shape, central tendency measures, and variability of graphical representations but are unable to integrate them into their solutions. Students at Level 4 can identify statistical processes with accuracy, but they are unable to completely comprehend or integrate them. These students, for example, may understand the concept of averages but are unable to interpret it fully. Level 5 students have a thorough understanding of statistical processes and are capable of synchronizing rules and behavior as well as explaining the function in their own words. Students at Level 1 (idiosyncratic reasoning) would display pre-structural traits in the iconic style. Meanwhile, students at Level 2 (verbal reasoning) would exhibit characteristics of the uni-structural level using the concrete symbolic mode. Students at Level 3 would likewise use the concrete figurative approach to depict components of the multi-structural level (transitional reasoning). Students at Level 4 (procedural reasoning) would show tangible symbolic representations of relationship level traits. Finally, Level 5 students formally reveal extended abstract level features. (Reasoning based on integrated processes).

In addition to Garfield's five levels of statistical reasoning ability; Biggs & Collins (1982) also introduced the taxonomic model Structure of Observed Learning Outcomes (SOLO) which was used as a cognitive model for the development of statistical reasoning, where junior high school students should have been in a formal function mode. Garfield & Ben-Zvi (2008) further identified that there are 9 important things in the development of statistical reasoning abilities, including data (including the nature of data and types and various data sources), statistical models (regression and normal distribution), distribution (involving ideas about, form, centering, and dispersion of data), center (includes the difference between the median and mean), variability (including measurement and spread of distribution), comparing groups based on concentration and dispersion, sampling (such as the effect of sample size), inferential statistics (testing hypothesis or confidence interval), and covariation with scatterplots, correlation, and linear regression.

Statistical reasoning abilities can also be investigated by taking into account the characteristics of the learning provided, including (1) an inquiry-based learning environment that builds collaborative norms among students; (2) statistical concepts and tools to support and expand students' informal statistical constructions in concluding the data presented; and (3) providing data-rich tasks that trigger conflicts with beliefs (contextual and/or statistical) to encourage students to seek deeper insights and explanations (Makar, 2013; Makar & Fielding-Wells, 2011). Wild & Pfannkuch (1999) also stated that statistical learning which emphasizes the development of statistical reasoning abilities is carried out through real situations. Students who obtain data from real situations will gain new knowledge to provide information and make decisions and actions on evidence-based situations that will later have an impact on similar problems that exist in people's lives. The ability to philosophize on data, argue about patterns, and provide suggestions regarding solutions to problems encountered depends on a good foundation of statistical and contextual knowledge. This is because statistics demand data-based and social argumentation skills (Pfannkuch, 2011).

Several types of tests used to measure students' statistical reasoning abilities have been developed, such as The Reasoning about P-values and Statistical Significance (RPASS) which was developed in the form of an application and focused on measuring advanced inferential statistics (Lane-Getaz, 2013), the Comprehensive Assessment of Outcome in Statistics (CAOS) (DelMas et al., 2007; Zieffler et al., 2010), Statistical Reasoning Assessment (SRA) (Garfield, 2002), Statistical Concept Inventory (SCI) (Stone et al., 2003) which was developed focusing on measuring various introductory statistical concepts, and The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) which focused on the topic of test significance. Zieffler et al. (2008) also developed tests that focus on types of tests such as: estimating and graphing a population, comparing two data samples, and making an assessment of the two data samples being compared to draw conclusions and final decisions on the two data samples owned. The same thing was also done by Huey & Jackson (2015) using a test instrument that focuses on inferential, unstructured, open-ended, context-based dimensions, and represents visual displays (such as tables, graphs, plots, diagrams, and others).

Based on the test instruments developed as a whole using data derived from contextual problems. The real problems used in developing students' statistical reasoning abilities can be integrated with the cultural context and traditions inherent in students' daily lives. Real experiences that are contextual and come from cultures and traditions that are close to students will increase students' motivation and interest. The use of culture in the context of mathematics is not new, because mathematics is a product of culture (Marsigit et al., 2018) and mathematics is part of local wisdom (Madusise, 2015). Culture-based mathematics known as ethnomathematics makes it easy for students to do mathematical modeling based on ideas, methods, and techniques from what has been developed by the surrounding community and can be used as an alternative to introduce students to be closer to the phenomena that occur in their lives (Prahmana et al., 2021). Investigation of data originating from phenomena that occur in students' lives will help develop students' mathematical reasoning abilities, to students' creativity (D'Ambrosio & Rosa, 2017; Rosa & Orey, 2017), and this can also be integrated into the test instruments used in measuring students' statistical reasoning ability. Incorporating the ethnomathematics framework in testing statistical reasoning abilities further supports the SOLO model's definition of cognitive thinking. Students who respond to their physical environment are at the sensorimotor stage in the SOLO model. Students that can internalize actions through images are at the iconic moment. Students in the concrete symbolic stage are interested in symbolic systems such as number schemes, maps, and the written word. Students can also translate abstract concepts (informal mode) into formal thoughts (formal mode). Post-formal students are those who comprehend the fundamental structure of theories and disciplines. Based on this, junior high school students according to the SOLO thinking model, are in a formal model (Chan et al., 2016), so that measurements using the ethnomathematics context can help students optimize their statistical reasoning abilities.

This study aims to analyze the statistical reasoning ability possessed by junior high school students which focuses on descriptive statistics measured using the level of statistical reasoning ability developed by Garfield. This study also wants to see how the pattern of answers given by students, is classified hierarchically based on student demographics (gender, ethnicity, and initial mathematical ability). We will also compare the answers given by students to see bias or differences based on student demographics, as well as consistent student answer patterns. This study uses a test instrument that was developed based on the ethnomathematics context. The ethnomathematics context used is Malay culture which will later be integrated into simple descriptive statistical problems. This study has several research questions to be explored, including:

1. How do students demonstrate statistical reasoning ability using ethnomathematics-based problems?
2. What is the level of students' statistical reasoning ability using ethnomathematics-based problems based on Rasch Model analysis?
3. Are there biases of gender, ethnicity, and students' initial mathematical abilities on students' statistical reasoning abilities using ethnomathematics-based problems based on Rasch Model analysis?

RESEARCH METHODOLOGY

This study uses a combination of qualitative and quantitative approaches (mix-method) to gain a clearer understanding of the exploration of students' statistical reasoning abilities in descriptive statistics learning as well as to triangulate findings (Creswell & Clark, 2018). The subjects of this study were junior high school students who were in the final level, totaling 22 people. Descriptions of research subjects can be seen in Table 1 below:

Table 1.

Demographic Background of Research Sample

Demographic Background		Number of Participants	Percentage
Gender	Male	15	68.18%
	Female	7	31.82%
Age	13 years old	3	13.64%

	14 years old	17	77.28%
	> 15 years old	2	9.08%
Initial Mathematics Ability	Low	4	18.18%
	Average	14	63.64%
	High	4	18.18%
Cultural Tribal Background	Pure (Malay ethnic)	17	77.28%
	Mix (Malay ethnic and Javanese ethnic)	5	22.72%

Based on Table 1 above, the demographic indicators on the research subjects used refer to the gender of the students, the students' initial mathematics ability, and the cultural tribal background of each student. Information related to ethnic and cultural backgrounds possessed by students is important to ensure that students understand the cultural context that will be integrated into descriptive statistics learning. This is also done to ensure that the findings of this study do not have bias related to differences in cultural backgrounds, even though the research was conducted in areas that have a Malay cultural background. The three demographic indicators are used on research subjects with the aim that the analysis carried out on students' statistical reasoning abilities can be carried out at the individual level of students and the level of the items used.

Multi-Tier Test Instrument of Statistical Reasoning Ability

The test instrument used in this study was a test instrument in the form of a multi-tiered essay test, which consisted of 3-tiered essay questions (the total number of tests was 10 questions). The research test instrument used to analyze students' statistical reasoning abilities based on the answer process carried out by students referred to the results of statistical learning tests (limited to the concept of data collection and calculation of central values on date data). The assessment of the test instrument is based on a rating scale that is adjusted to the indicators of statistical reasoning ability on each item (ie idiosyncratic, verbal, transitional, procedural, and investigative process reasoning). The highest score for each test is 5, and the lowest score for each test is 1. If the student does not answer the given test, then no score is given (no assessment).

Item Response Test (IRT) using Rasch Model Measurement

The test instrument in analyzing students' statistical reasoning abilities was developed based on constructs in descriptive statistics learning and the level of statistical reasoning abilities. The description of the test instrument used is based on the descriptive statistics learning construct and the level of statistical reasoning ability which can be seen in Table 2 and Table 3 below:

Table 2.

Description of Developed Test Instruments Based on the Descriptive Statistics Learning Construct

Constructs	Code	Sub-Process	Items
Describing data	A1	Extraction and generation of information from data or graphs depending on a contextual situation in Malay culture	How much cotton, silk, gold, and silver thread is needed in the weaving of the traditional Malay Songket based on the graph?
	A2	Demonstrating sensitivity to the graphical representation's exhibited qualities	Do you agree, based on the graph above, that silk thread has the fewest uses in the Malay Songket production process compared to other types of thread?
	A3	Recognizing the graphical representation's general features	What is the difference in the usage of silk and cotton, silk and gold-silver, cotton and gold-silver threads in the production of a typical Malay Songket, according to the graph above?
Organizing and reducing data	B1	Making a table out of the data	Arrange the tabular data acquired and adjust it to the weavers' production and number of Okik used (based on the information presented related to the weaving culture)!
	B2	Using the measure of center to reduce the data, either by calculation or analysis	In which month is the highest number of Okik used
	B3		Determine the number of Okik in the middle of the sequence if you order the number of Okik from least to most!
Representing data	C1	Recognizing multiple representations of the same data set	Analyze the outcomes of these sales to determine which month saw the

			highest profit from Putu Mayam sales based on the graph!
	C2		Re-examine, which month saw the greatest growth in profit?
Analyzing and interpreting data	C3	Comparing data from the same data set	What is the total profit earned over the course of a year based on the findings of your analysis?
	C4	Using data or graphs to make predictions, inferences, or draw conclusions	Determine the average profit obtained from the sale of Putu Mayam!

Table 3.

Description of Developed Test Instruments Based on Statistical Reasoning Ability Level

Construct Level	Level 1 Idiosyncratic	Level 2 Verbal	Level 3 Transitional	Level 4 Procedural	Level 5 Integrated Process
Describing data	A1L1	A1L2	A1L3	A1L4	A1L5
	Do not extract or generate idiosyncratic or meaningful information from graphs.	Some information is extracted and generated from the graph, but the interpretation is imprecise or uncertain.	Extracts and creates information from a graph in one or two dimensions.	Correctly extracts and creates information from the graph	Completely extracts and creates information from the graph
	A2L1	A2L2	A2L3	A2L4	A2L5
	Does not demonstrate awareness of the graphical representation's displayed properties.	Shows understanding of the displayed properties of pictorial representation in writing, but is only partially right.	Displays minimal knowledge of the graphical representation's displayed features.	Demonstrates some understanding of the graphical representation's displayed attributes.	Demonstrates comprehensive understanding of the graphical representation's displayed attributes

	A3L1	A3L2	A3L3	A3L4	A3L5
	Does not identify the graphical representation's general properties.	Recognize general characteristics of the pictorial representation in words, however, it is only partially accurate	Recognizes one or two broad characteristics of the graphical depiction	Accurately recognizes general elements of the graphical representation	Completely recognizes general elements of the graphical depiction
Organizing and reducing data	B1L1	B1L2	B1L3	B1L4	B1L5
	The data could not be organized into a table.	Provides oral assertions when putting data into a table, but is only partially correct.	Organizes the data into a table with notable errors	Organizes the data into a table with minimal errors	Correctly organizes data into a table
	B2L1	B2L2	B2L3	B2L4	B2L5
	Incapable of reducing the data using central tendency measures through calculation or analysis	Reduces data by employing central tendency measures in words, either by computation or analysis, but is only partially correct.	Data is reduced using measures of central tendency with considerable errors, either by calculation or analysis.	Reduces the data by calculating or analyzing metrics of central tendency with minor errors.	Completely reduces the data utilizing central tendency metrics, either by calculation or analysis
	B3L1	B3L2	B3L3	B3L4	B3L5
	Unable to reduce the data using central tendency metrics via math or analysis	Reduces data by employing central tendency measures in words, either by computation or analysis, but is only partially correct.	Reduces data using measures of central tendency with significant errors, either by computation or analysis	Reduces the data by calculating or analyzing metrics of central tendency with minor errors.	Completely reduces the data utilizing central tendency metrics, either by computation or analysis

Representing data	C1L1	C1L2	C1L3	C1L4	C1L5
	Does not distinguish between alternative representations of the same data set.	Identifies many word representations for the same data set, but is only partially right.	Identifies one or two features shared by multiple representations of the same data set.	Identifies different representations of the same data collection correctly.	Determines the effectiveness of two distinct representations for the same data collection.
	C2L1	C2L2	C2L3	C2L4	C2L5
	Does not distinguish between alternative representations of the same data set.	Identifies many word representations for the same data set, but is only partially right.	Identifies one or two aspects of multiple representations of the same data set.	Recognizes different representations of the same data set.	Determines the effectiveness of two distinct representations for the same data gathering.
Analyzing and interpreting data	C3L1	C3L2	C3L3	C3L4	C3L5
	Do not compare data sets from the same source.	Compares multiple data sets verbally, however, the comparisons are imperfect.	Performs one or two comparisons within the same data sets	Accurately compares data sets from the same source.	Completely compare data sets inside the same data sets.
	C4L1	C4L2	C4L3	C4L4	C4L5
	Make no predictions, judgments, or conclusions based on the graphics.	Predictions, judgments, and conclusions derived from graphs are conveyed in words, yet they are insufficient.	One or two predictions, inferences, or conclusions are drawn from the graphs.	Predicts, infers, or draws suitable conclusions from graphs	Completely and comprehensively predicts, infers, or draws conclusions from graphs

Modified from: Chan et al. (2016)

The test instrument was developed in accordance with Table 2 above, then tested for validation and reliability using the Rasch Model Measurement analysis assisted by the Winstep application. Rasch Model Measurement is a measurement analysis based on Item Response Theory (IRT) developed by George Rasch (a Danish Mathematician) using the Joint Maximum Likelihood Estimation (JMLE) equation which converts raw data into the interval (logit) data (Soeharto, 2021). Rasch's estimation is based on the interaction of item-person and Probability estimations.

The interaction between items and persons (in this case, students) may be expressed mathematically using logit values (log odd units). The measurement's probability is governed by the difficulty of the item and person at the same time, yielding the logit item value (using the odd probability of each item) and the logit person value (using the odd probability of each respondent). In other words, the probability is closely proportional to the gap between item difficulty and student abilities (Boone, 2016; Boone et al., 2014). This means that the student's ability to measure remains the same regardless of the difficulty level of the item, but the item's difficulty level remains unchanged regardless of the student's ability. In this work, Rasch analysis was utilized to overcome some of the limitations of Classical Test Theory (CTT). In describing the measuring model, the CTT has four limitations: (a) the measurements are constructed using ordinal data results rather than an interval (logit) scale; (b) the items and people in the dependent measurement are highly dependent on the sample; (c) the nature of the instrument's measurement in terms of reliability and validity is highly dependent on the sample; and (d) data is centered on group-centered statistics but is insufficient for explaining individual respondent measurements (Barbic & Cano, 2016).

The logit scale in Rasch Model Measurement can indicate a person's skill and item difficulty from positive infinity to negative infinity. The results of the validation show that the multi-tiered description questions which the total number of questions used are 10 questions (statistical reasoning ability test) developed are valid (according to the valid item criteria regarding item response theory), namely the MNSQ OUTFIT value is in the range .5 to 1.5, the OUTFIT Z-STANDARD (ZSTD) value is in the range of -2 to +2 and the Point Measurement Correlation (Pt. Measure Corr) value is in the range of .4 to .85. The mean item size (logit) is .00, and the standard deviation (SD) is 1.36, indicating that the variance of the item measurements in item difficulty is wide across the logit scale. For the person (students), the average size is -.45 logit, which indicates that if the average value is less than the logit value of .00. Based on this logit value, we concluded the student's ability tends to be smaller than the difficulty level of the question. However, the SD on the person (students) is 1.75 which indicates that the variation of person is suitable for data analysis. The overall test instrument developed is also declared reliable by referring to the results of the summary statistical test (Rasch Model Measurement) on the Reliability value (alpha = .95) in the special category and the Reliability value (alpha = .88) in the good category for persons or students (Sumintono & Widhiarso, 2015). The summary of test items and person statistics can be seen in Table 4 below.

Table 4.

The Summary of the Statistics Based on Pearson and Items

Statistic Test	Test Group	
	Person	Item Test
N	22	10
Measure	-.45	.00
Mean	27.6	60.8
SD	1.75	1.36
SE	.98	.45
Mean Outfit MNSQ	.92	.92
Mean Outfit ZSTD	-.02	.08
Separation	2.76	4.28
Reliability	.88	.95
Cronbach's Alpha	.90	

After the test instrument for statistical reasoning ability is valid and reliable, then the test instrument can be given to the research subject to explore the abilities possessed by the research subject (junior high school students). The results of the answers given by students were analyzed based on the process of answers given in accordance with the indicators of statistical reasoning ability and descriptive statistics learning constructs (see Table 3). The answering process was then analyzed to see if there is a gender bias, initial mathematical ability, and cultural background that students have on the answer process given by students on each item of the question. The bias analysis was carried out by measuring Rasch using Differential Item Functioning (DIF). This test is carried out to answer the research problem formulation which allows that an instrument or item has a bias if it is found that one individual with certain characteristics is more advantageous than individuals with other characteristics. Demographic indicators (demographic variables) used to analyze the existence of bias in the results of answers made by students are gender factors, students' initial mathematical abilities, and cultural background. Furthermore, the DIF analysis can also provide information on whether there are students who have a tendency to answer questions in an unusual way (such as students with low initial abilities being able to answer difficult questions, but still unable to answer easy questions) (Sumintono & Widhiarso, 2015).

RESULT AND DISCUSSION

Students' Statistical Reasoning Ability Level and Difficulty Level of Test Instruments Based on Rasch Measurement Model Analysis

Based on the results of the analysis of the answers given by 22 students on the multi-tier statistical reasoning ability test, the results obtained as shown in the Wright Map below:

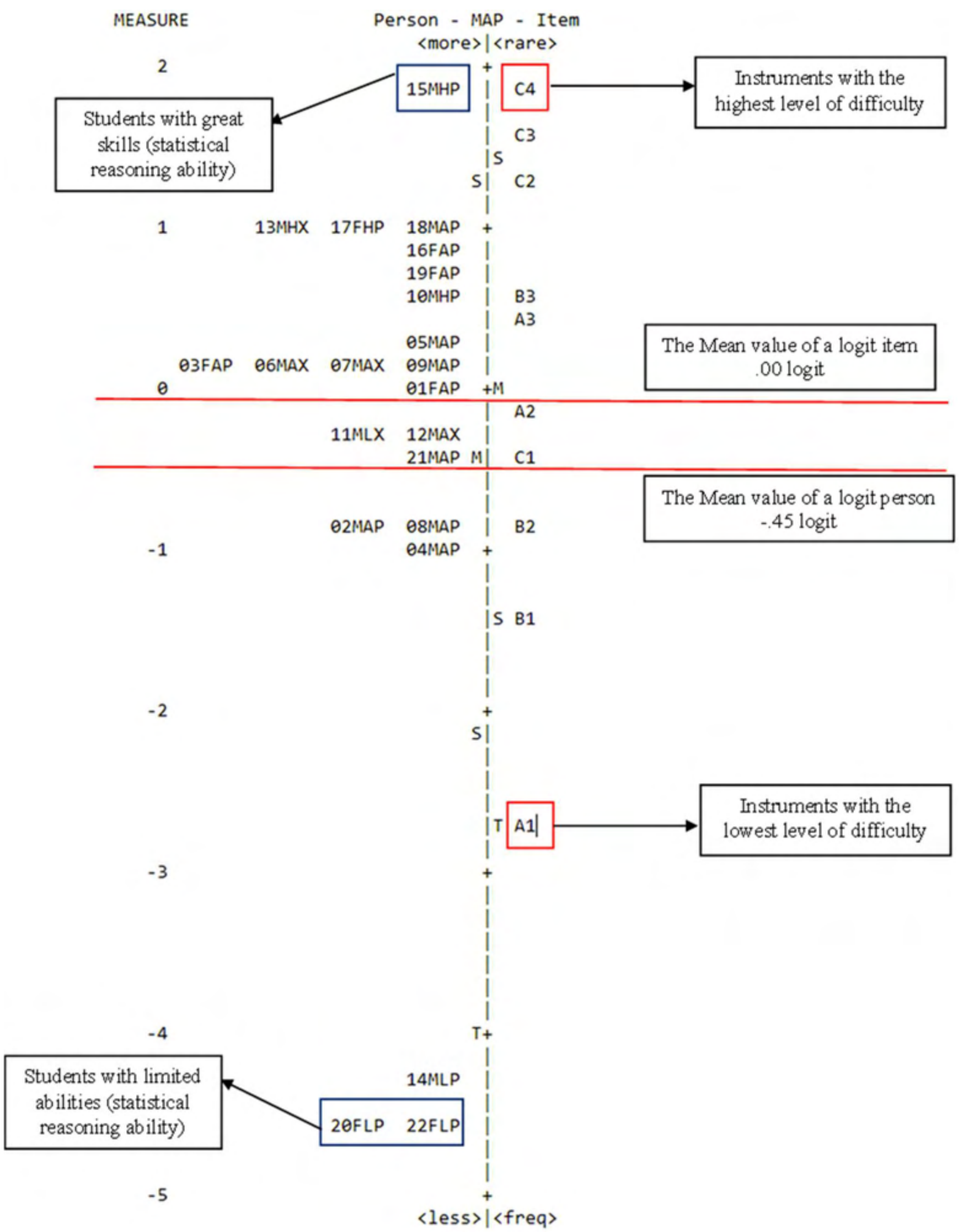


Figure 1.

Wright's Map Showing Students' Statistical Reasoning Ability Level and Test Instrument Difficulty Level Based on Logit Score

Based on Figure 1 above, it could be seen that student with code 15MHP (Male, High Initial Mathematics Ability, Pure Ethnic) are students with demonstrates high statistical reasoning abilities. This can be seen from the position of students on Map Wright who are in the top position with logit values approaching +2.00 which is +1.82, while students with low statistical reasoning abilities are students with codes 20FLP and 22FLP (Female, Low Initial Mathematics Ability, Pure Ethnic) with a logit value of more than -4.00 i.e., -4.64. Figure 1 above also shows that more than 50% of students are above the average logit person value (.00 logit), which is 13 students. In addition to the level of students' statistical reasoning abilities, the Wright Map also shows that the test instruments fall into the categories of difficult tests and easy tests. The test instrument with code C4 which has the construct of Analyzing and interpreting data is in the category of difficult test, while the test instrument with code A1 which has the construct of Describing data is included in the category of easy test.

Analysis of Student Answer Patterns on Statistical Reasoning Ability Tests in Cases of Descriptive Statistics in the Context of Malay Ethnomathematics

Further analysis was conducted to see how the pattern of answers of students with high statistical reasoning abilities, namely students with 15MHP code and students with low statistical reasoning abilities, namely students with 20FLP and 22FLP codes based on the level of statistical reasoning ability.

Construct: Describing Data (A1, A2, and A3)

The pattern of student answers with code 15MHP (high statistical reasoning ability) can be seen in Table 5 below:

Table 5.

Student Answer Patterns Code 15MHP (high ability) on A1, A2, A3

a. Berapakah jumlah benang katun, sutera, dan emas-perak yang digunakan dalam penenunan kain songket khas Melayu? katun = 40 sutera = 70 emas-perak = 90	a. How much cotton, silk, gold, and silver thread is needed in the weaving of the traditional Malay Songket based on the graph? Cotton: 40: Silk: 70: Gold-Silver: 90

b. Berdasarkan grafik di atas, setujukah kamu jika benang sutera merupakan benang paling sedikit digunakan dibandingkan dengan jenis benang yang lain? Jika tidak setuju, jabarkan alasan kamu! tidak setuju, karena di bawah sutra masih ada benang katun yg jumlahnya sedikit	b. Do you agree, based on the graph above, that silk thread has the fewest uses in the Malay Songket production process compared to other types of thread? I can't entirely agree because the amount of silk threads remains more than the number of cotton threads.
c. Berdasarkan grafik di atas, bandingkan berapa selisih penggunaan benang sutera dan katun, benang sutera dan benang emas-perak serta benang katun dan benang emas-perak dalam proses produksi songket Batubara khas Melayu? sutra dan katun : 70 - 40 = 30 katun dan emas perak : 90 - 40 = 50 sutra dan emas-perak : 70 - 90 = 20	c. What is the difference in the usage of silk and cotton, silk and gold-silver, and cotton and gold-silver threads in the production of a typical Malay Songket, according to the graph above? Silk and Cotton: $70 - 40 = 30$ Silk and Gold-Silver: $90 - 70 = 20$ Cotton and Gold-Silver: $90 - 40 = 50$

Based on students' answers to the 15MHP code on tests A1, A2, and A3 it could be seen that students extract information about the number of cotton, silk, and gold-silver threads well from the graphs presented; is aware of the appropriate information on the graphs presented; can compare the information presented on the graph with the appropriate calculations. The pattern of answers given by students with the 15MHP code is in accordance with level 4 Procedural, namely students can well display, realize, and compare the information presented on the graph. The pattern of the answers was as expected, although the students did not provide a more detailed explanation regarding the answers given. This can be seen in the answers to the A2 test, where 15MHP did not describe the least number of threads according to the graph presented with the aim of supporting disagreement on the statement. 15MHP students have high initial mathematical abilities, so they are able to give correct answers on tests with easy (A1 and A2), and moderate (A3) categories. The student with code 15MHP also has a native Malay ethnic background (not mixed), so that the integrated Malay cultural context (Songket Malay) on the A1, A2, and A3 tests can be easily understood. Different things can be seen from the pattern of students' answers to codes 20FLP and 22FLP which do not present the right information based on the graphs presented. The following is a description of the answer patterns for students with codes 20FLP and 22FLP which are presented in Table 6.

Table 6.

Student Answer Patterns Code 20FLP and 22FLP (low ability) on A1, A2, A3 Tests

Student Answer Patterns Code 20FLP	Student Answer Patterns Code 22FLP
a. Berapakah jumlah benang katun, sutera, dan emas-perak yang digunakan dalam penenunan kain songket khas Melayu? 20 pengguna benang 50 kain pada a. How much cotton, silk, gold, and silver thread is needed in the weaving of the traditional Malay Songket based on the graph? 20 threads dan 50 fabrics	a. Berapakah jumlah benang katun, sutera, dan emas-perak yang digunakan dalam penenunan kain songket khas Melayu? jumla adalah 90 a. How much cotton, silk, gold, and silver thread is needed in the weaving of the traditional Malay Songket based on the graph? The total number of threads is 90.
b. Berdasarkan grafik di atas, setujukah kamu jika benang sutera merupakan benang paling sedikit digunakan dibandingkan dengan jenis benang yang lain? Jika tidak setuju, jabarkan alasan kamu! dibenun dengan menggunakan jenis yang dipakai sol songket. b. Do you agree, based on the graph above, that silk thread has the fewest uses in the Malay Songket production process compared to other types of thread? The threads are woven using the type used in Songket fabrics.	b. Berdasarkan grafik di atas, setujukah kamu jika benang sutera merupakan benang paling sedikit digunakan dibandingkan dengan jenis benang yang lain? Jika tidak setuju, jabarkan alasan kamu! yang di gunaka untuk kartun b. Do you agree, based on the graph above, that silk thread has the fewest uses in the Malay Songket production process compared to other types of thread? Cotton is the least commonly used fabric.
c. Berdasarkan grafik di atas, bandingkan berapa selisih penggunaan benang sutera dan katun, benang sutera dan benang emas-perak serta benang katun dan benang emas-perak dalam proses produksi songket Batubara khas Melayu? 40 dengan menggunaan dengan 70 atas bandingka 40-70: 60 c. What is the difference in the usage of silk and cotton, silk and gold-silver, and cotton and gold silver threads in the production of a typical Malay Songket, according to the graph above? 40 threads compared to 70 threads by making a comparison between 70 - 40 = 60	c. Berdasarkan grafik di atas, bandingkan berapa selisih penggunaan benang sutera dan katun, benang sutera dan benang emas-perak serta benang katun dan benang emas-perak dalam proses produksi songket Batubara khas Melayu? Berdasarka di atas bandingka 40 c. What is the difference in the usage of silk and cotton, silk and gold-silver, and cotton and gold silver threads in the production of a typical Malay Songket, according to the graph above?? According to this comparison, the difference is 40.

Based on Table 5 above, it can be seen that the pattern of answers given by the two students (20FLP and 22FLP) has the same pattern. The pattern of answers given does not provide information that matches the graph shown. Students coded 20FLP and 22FLP gave answers that were very far from what was expected on the A1, A2, and A3 tests. The tendency of the answer pattern was due to the fact that the two students did not understand what was being asked on the test. Both students were perplexed and found it difficult to discern the correct response. This finding is in accordance with the results of interviews related to the statistical reasoning of the two students.

R (Researcher)	*: Do you understand the information presented on the graph?*
20FLP	*: Honestly, I understand the question (pause).*
22FLP	*: I understand the numbers that appear on the graph.*
R (Researcher)	*: Why did you answer 20 threads and 50 clothes on the A1 test (To 20FLP students)?*
20FLP	*: I was thinking if the A1 test meant subtracting the numbers 70, 90, and 40 on the graph (the face looks confused) and I didn't know that what the A1 test really meant was rewriting the number of each yarn according to the graph.*
R (Researcher)	*: How did you get an answer of 90 on the A1 test?*
22FLP	*: I was wrong with the test given. I thought that what was being asked was only the number of gold-silver threads, so that's why I wrote 90. I only realized after I got my answers. I panicked because I couldn't see clearly which numbers were on the silk thread, the cotton thread, and the gold-silver thread.*

Based on the results of the interviews above, it is evident that the 20FLP and 22FLP students did not understand well the questions given on the A1, A2, and A3 tests even though the A1 and A2 tests were easy questions. Perplexed and panic are some of the factors that underlie the inappropriateness of the answers given. In addition, the factor of the low initial mathematical ability of the two students could also be one of the factors causing the students' inability to understand and reason about the meaning of the tests given. Initial abilities are needed to support subsequent learning performance (Nogues & Dorneles, 2021). The two students also had inadequate mathematical literacy; thus, they did not interpret the information on the test correctly. Based on this, students with codes 20FLP and 22FLP can only be at level 1 idiosyncratic-statistical reasoning ability.

Construct: Organizing and reducing data (B1, B2, and B3)

The pattern of student answers with code 18MAP (medium ability) on tests B1, B2, and B3 can be seen in Table 7 below.

Table 7.

Student Answer Pattern Code 18MAP (medium ability) on Tests B1, B2, and B3

<table>
<tr>
<td>
Berdasarkan pada informasi di atas,
a. Kelompokkan data yang diperoleh dalam bentuk tabel dan sesuaikan dengan bulan produksi serta jumlah Okik yang digunakan oleh penenun!
<table>
<tr><th>Bulan Produksi</th><th>Jumlah Okik</th></tr>
<tr><td>Januari 2020</td><td>10 Buah</td></tr>
<tr><td>Febuari 2020</td><td>10 menjadi 12</td></tr>
<tr><td>Maret 20-20</td><td>9 Buah</td></tr>
<tr><td>April 2020</td><td>9 Buah</td></tr>
<tr><td>Mei 2020</td><td>9 Buah</td></tr>
<tr><td>Juni 20-20</td><td>9 Buah</td></tr>
<tr><td>Juli 20-20</td><td>14 Buah</td></tr>
<tr><td>November 2020</td><td>14 Buah</td></tr>
<tr><td>Agustus 20-20</td><td>14 Buah</td></tr>
<tr><td>September 20-20</td><td>14 Buah</td></tr>
<tr><td>OktoBer 20-20</td><td>14 Buah</td></tr>
<tr><td>Desember 20-20</td><td>17 Buah</td></tr>
</table>
2
4
5
</td>
<td>
a. Arrange the tabular data acquired and adjust it to the weavers' production and number of Okik used (based on the information presented related to the weaving culture)!
<table>
<tr><th>Production Month</th><th>Total of Okik</th></tr>
<tr><td>January 2020</td><td>10 Okik</td></tr>
<tr><td>February 2020</td><td>10 becomes 12 Okik</td></tr>
<tr><td>March 2020</td><td>9 Okik</td></tr>
<tr><td>April 2020</td><td>9 Okik</td></tr>
<tr><td>May 2020</td><td>9 Okik</td></tr>
<tr><td>June 2020</td><td>9 Okik</td></tr>
<tr><td>July 2020</td><td>14 Okik</td></tr>
<tr><td>August 2020</td><td>14 Okik</td></tr>
<tr><td>September 2020</td><td>14 Okik</td></tr>
<tr><td>October 2020</td><td>14 Okik</td></tr>
<tr><td>November 2020</td><td>14 Okik</td></tr>
<tr><td>December 2020</td><td>17 Okik</td></tr>
</table>
</td>
</tr>
<tr>
<td>
b. Analisislah pada bulan berapa jumlah Okik yang paling banyak digunakan?
Bulan DesemBer mencapai 17 Buah
</td>
<td>
b. In which month is the highest number of Okik used?
In December, the number of Okik reached 17 pieces.
</td>
</tr>
<tr>
<td>
c. Jika kamu mengurutkan jumlah Okik dari yang paling sedikit hingga paling banyak, tentukan jumlah Okik yang berada pada posisi tengah urutan tersebut!
Jumlah = 10 Buah menjadi 2
Buah ~~10~~ Jumlah = 9 Buah menjadi 4
Jumlah = 14 Buah menjadi 5
Jumlah = 17 Buah menjadi 1
</td>
<td>
c. Determine the number of Okik in the middle of the sequence if you order the number of Okik from least to most!
The total of Okik = 10 pieces in 2 months
The total of Okik = 9 pieces in 4 months
The total of Okik = 14 pieces in 5 months
The total of Okik = 17 pieces in 1 months
</td>
</tr>
</table>

Based on Table 7 above, it can be seen that 18MAP code students were able to give the correct answer, but still had difficulty communicating the analysis and reasoning process obtained in written form. Seen in the pattern of answers to the B1 test, 18MAP students were able to reduce data from the information presented in tabular form. However, it appears that there is a typo in the February 2020 production month line. 18MAP students have understood that in the February production month, the Okik used by the weavers increased to 12 pieces from the previous 10 pieces. 18MAP students did not reason well with the information they received, but instead wrote down the information they received in a table, even though the answers they gave were not wrong, only inaccurate and unclear. The same thing is also seen in the pattern of answers to the B3 test, where students are still wrong in understanding what is asked in the test. 18MAP students did not sort the number of Okik used from the smallest number to the largest number so the purpose of the test given was not achieved. The B3 test is intended so that students are able to determine the median value on date data that has been sorted from the smallest to the largest data. This is obtained from interviews with student code 18MAP as follows:

R (Researcher) : *What do you mean by the answer that in February 2020, the number of Okik used was 10 to 12?*
18MAP : *Oh yes, ma'am, what I mean is that in February 2020, the Okik used by weavers increased, ma'am. In the question, it was written that in February 2020 there were additional weavers, so the Okik used also increased. So, in the February 2020 table, I wrote down Okik which was used from 10 to 12 pieces. That means increased to 12 pieces ma'am.*
R (Researcher) : *Okay, how about the answers on the B3 test? What do you mean by the number of 10 pieces in 2 months, 9 pieces in 4 months, and so on? What do you mean that there are 10 Okik in 2 months of production? Please explain again!*
18MAP : *Oh yes ma'am (pause for a moment… continue again). What I understand from the problem is to write down the number of Okik used starting from the least to the most. So, I wrote that there were 10 Okik used in January and February 2020, and soon, ma'am.*
R (Researcher) : *Are you sure that's what the B3 test meant?*
18MAP : *I think so ma'am, but it still seems wrong, ma'am (students are not sure about the answer that has been given)*

Based on the pattern of answers and confirmation of the answer process given by students with code 18MAP, it was concluded that students had errors in understanding what was meant by the test given. Understanding the questions or problems given is an important factor in carrying out the data analysis process so that later it can be continued in determining the appropriate statistical calculations. The understanding of the questions given is included in the context of mathematical literacy. Students with low mathematical literacy ability will have an impact on improving other mathematical abilities such as problem-solving, reasoning, and other cognitive abilities. Mathematical literacy ability is proven to provide support for reasoning processes and other cognitive abilities, so that it will help students in improving mathematics learning outcomes (Holenstein et al., 2021). Based on these findings, students with code 18MAP on tests B1, B2 are at level 3 Transitional, where students have been able to reduce data in tabular form even though there are still errors in one of the data provided; able to reduce the data and use it in determining the largest data. However, on the B3 test, students with code 18 MAP were at level 2 verbal. This is because students are still wrong in determining the most frequently used data according to the information in the ethnomathematics context presented.

Construct: Analyzing and interpreting data (C3 and C4)

The pattern of students' answers with code 15MHP (high statistical reasoning ability) on the C3 and C4 tests can be seen in Table 8 below.

Table 8.

The Pattern of Student Answer Code 15MHP (high ability) on C3 and C4 Tests

c. Berapakah total keuntungan yang diperoleh selama 12 bulan menurut hasil analisis kamu? 45,7 Juta selama 12 bulan	c. What is the total profit earned over the course of a year based on the findings of your analysis?? 45.7 million for 12 months.
d. Coba kamu putuskan, berapa rata-rata keuntungan yang diperoleh dalam penjualan Putu Mayam tersebut? ~~45~~ 3,833	d. Determine the average profit obtained from the sale of Putu Mayam! 3.833

Based on Table 8 above, it can be seen that the 15MHP code students did not write down their answers completely and clearly. In the answers to the C3 and C4 tests, students with 15MHP codes were able to analyze the answers correctly but did not provide procedural explanations. The results of the verification and confirmation regarding the answers given by students with code 15MHP through interviews are as follows.

R (Researcher) : *How do you get the figure of 45.7 million as a total profit earned for 12 months? Please explain the steps.*

15MHP : *I got the figure of 45.7 million by adding up all the sales profits from month 1 to month 12.*

R (Researcher) : *Are you sure 3833 is the correct answer on the C4 test? How did you get that result?*

15MHP : *At first, I was sure, but after I looked again at the graph, I became unsure of the answer. I have a hard time dividing numbers in decimal form (pause for a moment, then think). So, I think my answer is still not correct, right? (Shows confused face).*

R (Researcher) : *How did you get that result?*

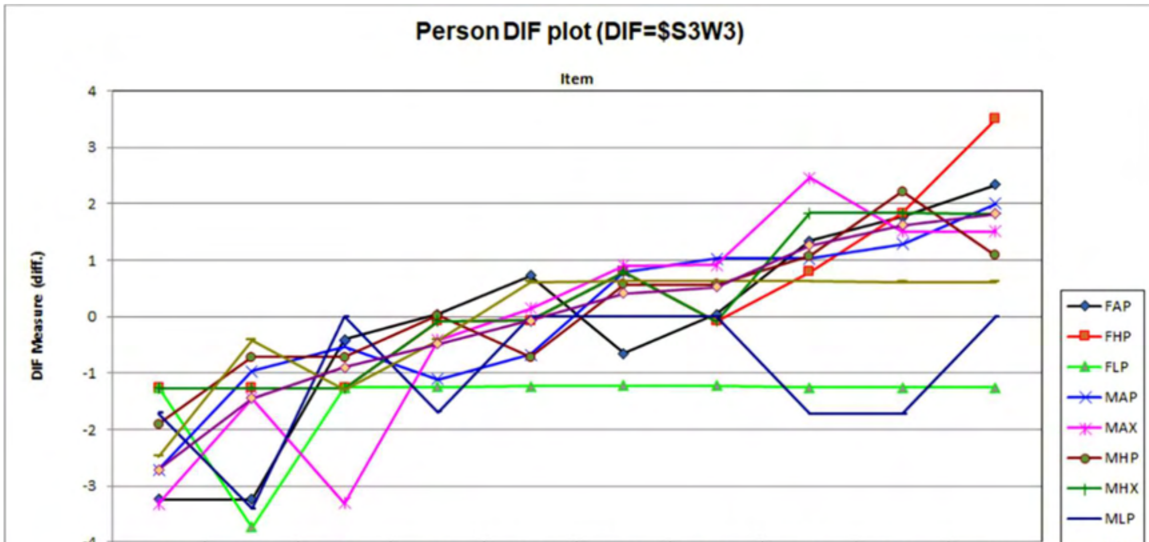
15MHP : *Initially, I added up all the sales profits from month 1 to month 12. Then, I divided the previous amount by 12 because the question asked was the average profit for 12 months. This is what I doubt, whether the answer is correct or still wrong.*

Based on the confirmation of the answer process obtained from students with code 15MHP, students have difficulty when faced with calculating data by hand in the form of decimal numbers. This is in accordance with the findings of Roell (2017) where rational numbers and operations involving fractions and decimals are one of the biggest obstacles for students in learning mathematics, especially statistics. Students will find it difficult to complete decimal number operations without adequate knowledge of the concepts involved. The difficulties experienced by the 15MHP code student ultimately provided further obstacles in the analysis and reasoning process of the results that had been obtained. The results of the 15MHP students' answers on the C4 test were still not correct, because there was an error in placing the position of the decimal number in the division operation (in order to determine the average value). However, 15MHP students can understand the concept of determining the average score on descriptive statistics referring to the results of the interviews obtained. Therefore, students who code 15MHP on the C3 and C4 tests are at level 2 verbally, because students are still not able to carry out the reasoning process correctly and completely. But verbally students are able to understand how to get the right answer (understand the concept of average value). Overall, based on the results of the answers to 22 students, it was found that almost all students had difficulties in completing the C3 and C4 tests, this was also because the two tests were included in the difficult test category (can be seen in Figure 1 Map Wright).

The pattern of students' answers which were analyzed based on the level of initial mathematical ability they had, the applied descriptive statistics learning construct, and the level of statistical reasoning ability, it was found that students with low initial mathematical abilities were still at level 1 idiosyncratic. This is clearly seen from the pattern of answers given and the confirmation of the answers given. Anxiety and students' lack of understanding related to mathematical literacy are the biggest challenges for students with low initial abilities. Meanwhile, students with moderate initial abilities (average) are at level 2 verbal and Level 3 Transitional. Both levels of statistical reasoning ability are dominantly seen in students with moderate initial abilities. This is because students have been able to carry out the process of reduction, analysis, and reasoning on the given problem, but still verbally, and have difficulty communicating it in written form. Students who are able to reach level 4 Procedural statistical reasoning abilities in descriptive statistics learning are students with high initial abilities. This can be seen from the pattern of answers and confirmation of the answers given to indicate that students with high initial abilities are able to perform reduction, analysis, and reasoning on the given problem. Although there are still incomplete and unclear solutions to problems, they have given the right results. The unique findings are seen specifically in questions C3 and C4 which are difficult questions, where students with high initial abilities also have difficulty in analyzing and reasoning the problems given. Students are still wrong in doing calculations and are not sure about the results and conclusions obtained. In the questions, C3 and C4, students with high initial abilities were still at level 2 verbal. This is because students are able to verbally understand and analyze the problem, but are still wrong in executing the answer process. The results of the analysis of the pattern of answers obtained provide the findings that the level of students' statistical reasoning abilities varies according to the learning construct used. However, the integration of the ethnomathematics context plays an important role for students, both students who have a native Malay ethnic background and a mixed Malay ethnic background, greatly assisting students in understanding the context of the problem presented. This is because students feel close to the problem presented. The ethnomathematics context used also provides a new experience for students in solving math problems more meaningfully (Fouze & Amit, 2018; Jurdak, 2016).

Differential Item Functioning (DIF) based on Gender, Initial Mathematics Ability, and Ethnics Background

The analysis of the pattern of student answers that have been described has answered the first and second research questions. The findings suggest that students' statistical reasoning abilities are closely related to their initial mathematical abilities. The results of the exploration of statistical reasoning abilities show how the level of students' statistical reasoning abilities in each descriptive statistical learning construct is presented in the form of ethnomathematics-based problems (Malay culture). Then, do the results of the exploration of statistical reasoning ability have a bias towards the gender, early mathematics ability, and ethnic background of the students? To answer this research question, we can perform an analysis using Differential Item Functioning (DIF) on the Rasch Model measurement. DIF shows the bias of differences in the answers given by students based on demographic factors, which can be seen by paying attention to the probability value of each item of the question. If the probability value of each item is below 5%, then the item questions that have been completed by students provide a bias and will tend to favor one of the demographic factors (Soeharto & Csapó, 2021). Based on the results of the DIF analysis, it was found that the probability value of all items used in analyzing the pattern of student answers related to the exploration of students' statistical reasoning abilities was not below 5%, which means that the items used did not have a bias towards the pattern of student answers to favor one or the other one demographic factor (gender, early mathematical ability, and ethnic background). The results of a further analysis can be seen in the graphic presented below.



	1 A1	4 B1	5 B2	7 C1	2 A2	3 A3	6 B3	8 C2	9 C3	10 C4
FAP	-3.238	-3.2398	-0.4022	0.0464	0.7283	-0.6566	0.0462	1.3399	1.7592	2.339
FHP	-1.2725	-1.2698	-1.2658	-0.0768	-0.074	0.7935	-0.0782	0.7896	1.837	3.5024
FLP	-1.2581	-3.7265	-1.2545	-1.2474	-1.2379	-1.2266	-1.2241	-1.2551	-1.2548	-1.2546
MAP	-2.7103	-0.9678	-0.5355	-1.1152	-0.6778	0.7924	1.0208	1.0221	1.2863	1.9953
MAX	-3.3129	-1.4491	-3.3113	-0.4218	0.145	0.9072	0.9097	2.4539	1.5046	1.5098
MHP	-1.9033	-0.7195	-0.725	0.0189	-0.7251	0.5693	0.5674	1.0724	2.2131	1.0763
MHX	-1.2725	-1.2698	-1.2658	-0.0768	-0.074	0.7935	-0.0782	1.8304	1.837	1.8217
MLP	-1.7152	-3.4017	0	-1.698	0	0	0	-1.728	-1.727	0
MLX	-2.4795	-0.4176	-1.2886	-0.4225	0.6153	0.6267	0.6279	0.6176	0.6135	0.6165
*	-2.7103	-1.4491	-0.8984	-0.4904	-0.074	0.4064	0.5201	1.2596	1.6144	1.8217

Figure 2.

Person DIF on Student Answer Patterns Regarding Gender, Early Mathematics Ability, and Ethnic Background

Based on Figure 2 above, the DIF analysis confirmed that students with female gender, low early math abilities, and ethnic Malay backgrounds had a fixed pattern of answers and did not experience significant changes. This can be seen in the DIF Measure that students have with the FLP code where 9 out of 10 items get scores that are not much different. The results of the DIF analysis concluded that although students have supportive demographic factors, such as gender, high early math abilities, and ethnic backgrounds that are in accordance with the given ethnomathematics-based problems, they do not provide benefits for students in improving learning outcomes, especially those closely related to improving students' statistical reasoning ability. However, it cannot be omitted that students' initial mathematical abilities also have their own role to support the development of other mathematical abilities. This condition is possible, and it can be seen that students with code 15MHP who have high initial mathematical abilities have good answer patterns and are at level 4 procedural on item questions with the construct of analyzing and interpreting data. On the other hand, students with codes 20FLP and 22FLP who have low initial mathematical abilities and low scorers have a poor answer pattern and are at level 1 idiosyncratic on all items about statistical reasoning abilities. Wyse & Mapuranga (2009) also reported that bias on test items may occur based on demographic factors of research respondents. Based on this, the findings obtained in this study can be re-analyzed to see how big the implications are between other demographic factors by giving different treatments.

CONCLUSIONS

Students' statistical reasoning abilities can be explored well if students and teachers know the level of statistical reasoning abilities they have. The level of statistical reasoning ability can be analyzed by providing a test instrument that is developed based on contextual problems such as culture and traditions that are close to students' daily lives. The integration of the ethnomathematics context in the statistical reasoning ability test instrument helps students understand the statistical problems presented and makes it easier for students to carry out the process of reducing, interpreting, analyzing, and reasoning data for further conclusions. Exploration of the statistical reasoning ability of high school students by giving problems in the ethnomathematics context of Malay culture resulted in the findings that students have different levels of statistical reasoning ability. However, the findings of this study also concluded that although there is no bias in the pattern of answers given by students to demographic factors (such as gender, early mathematics ability, and ethnic background), the early mathematics ability factor still provides its own support in influencing students' statistical reasoning abilities. This shows that students' statistical reasoning abilities can be further improved by paying attention to students' initial mathematical abilities, as well as using the context of the problems presented.

The Rasch Model analysis provides an important role in checking for possible biases in student response patterns based on demographic factors. Rasch's analysis made it possible to further explore biases on demographic factors other than students' initial mathematical ability, gender, and student background. Other internal factors such as the level of learning motivation, the location of the student's residence area, as well as external factors such as giving different treatment to learning can be explored further using the Rasch Model analysis by providing optimal learning treatment.

In this study, giving problems with ethnomathematics contexts was proven to help students understand the problems presented to assist students in conducting the analysis and reasoning process of the data obtained further. These findings provide recommendations for researchers to integrate cultural contexts and traditions close to students in presenting mathematical problems. This study produces limited findings, where the ethnomathematics context presented uses the cultural context and traditions of the Malay tribe in general, and the number of research subjects used is also small. Therefore, other researchers can continue similar research by paying attention to the demographic factors used, research subjects, and the existence of learning treatments so that the findings obtained can provide significant results.

References

[1] Barbic, S. P., & Cano, S. J. (2016). The application of Rasch measurement theory to psychiatric clinical outcomes research: Commentary on … Screening for depression in primary care†. *BJPsych Bulletin*, *40*(5), 243. https://doi.org/10.1192/PB.BP.115.052290

[2] Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, *88*(3), 291–303. https://doi.org/10.1007/S10649-015-9593-3/FIGURES/1

[3] Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. Academic.

[4] Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

[5] Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.

[6] Chan, S. W., Ismail, Z., & Sumintono, B. (2016). A Framework for Assessing High School Students' Statistical Reasoning. *PLoS ONE*, *11*(11), 1–32. https://doi.org/10.1371/journal.pone.0163846

[7] Ching, B. H. H., Kong, K. H. C., Wu, H. X., & Chen, T. T. (2020). Examining the reciprocal relations of mathematics anxiety to quantitative reasoning and number knowledge in Chinese children. *Contemporary Educational Psychology*, *63*, 101919. https://doi.org/10.1016/j.cedpsych.2020.101919

[8] Creswell, J. W., & Clark, V. L. P. (2018). *Designing and Conducting Mixed Methods Research* (Third Edit). SAGE Publications, Inc.

[9] D'Ambrosio, U., & Rosa, M. (2017). Ethnomathematics and Its Pedagogical Action in Mathematics Education. In *Ethnomathematics and its Diverse Approaches for Mathematics Education* (pp. 285–305). https://doi.org/10.1007/978-3-319-59220-6_12

[10] delMas, R. C. (2002, January). *Statistical Literacy, Reasoning, and Thinking: A*

Commentary. Journal of Statistics Education. https://doi.org/10.1080/10691898.2002.11910674

[11] DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28–58. http://www.stat.auckland.ac.nz/serj

[12] English, L. D. (2014). Promoting statistical literacy through data modelling in the early school years. In Chernoff E. & Sriraman B. (Eds.), *Probabilistic Thinking* (pp. 441–457). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7155-0_23

[13] Fouze, A. Q., & Amit, M. (2018). Development of mathematical thinking through integration of ethnomathematic folklore game in math instruction. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(2), 617–630. https://doi.org/10.12973/ejmste/80626

[14] Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910676

[15] Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.

[16] Holenstein, M., Bruckmaier, G., & Grob, A. (2021). Transfer effects of mathematical literacy: an integrative longitudinal study. *European Journal of Psychology of Education*, *36*(3), 799–825. https://doi.org/10.1007/S10212-020-00491-4/TABLES/2

[17] Huey, M. E., & Jackson, C. D. (2015). A framework for analyzing informal inferential reasoning tasks in middle school textbooks . In T. G. Bartell, K. N. Bieda, R. T. Putnam, K. Bradfield, & H. Dominguez (Eds.), *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 431–438). Michigan State University.

[18] Jurdak, M. (2016). Learning and teaching real world problem solving in school mathematics: A multiple-perspective framework for crossing the boundary. In *Learning and Teaching Real World Problem Solving in School Mathematics: A Multiple-Perspective Framework for Crossing the Boundary*. Springer International Publishing. https://doi.org/10.1007/978-3-319-08204-2

[19] Lane-Getaz, S. J. (2013). Development of a reliable measure of students' inferential reasoning ability. *Statistics Education Research Journal*, *12*(1), 20–48. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=15701824&v=2.1&it=r&id=GALE%7CA450999592&sid=googleScholar&linkaccess=fulltext

[20] Madusise, S. (2015). Cultural villages as contexts for mediating culture and mathematics education in the South African curriculum. *Revista Latinoamericana de Etnomatemática*, *8*(2), 11–31. http://www.redalyc.org/articulo.oa?id=274041586002

[21] Mahdayani, R. (2016). Analisis kesulitan siswa dalam pemecahan masalah matematika pada materi aritmetika, aljabar, statistika. dan geometri. *PENDAS MAHAKAM: Jurnal Pendidikan Dan Pembelajaran Sekolah Dasar*, *1*(1), 86–98. https://jurnal.fkip-uwgm.ac.id/index.php/pendasmahakam/article/view/39

[22] Makar, K. (2013). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics* , *86*(1), 61–78. https://doi.org/10.1007/S10649-013-9526-Y

[23] Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education-New ICMI Study SeriesNew ICMI Study Series* (Vol. 14, pp. 347–358). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-1131-0_33

[24] Marsigit, M., Setiana, D. S., & Hardiarti, S. (2018). Pengembangan Pembelajaran Matematika Berbasis Etnomatematika. *Prosiding Seminar Nasional Pendidikan Matematika Etnomatnesia*, *0*(0), 20–38. https://jurnal.ustjogja.ac.id/index.php/etnomatnesia/article/view/2291

[25] Maryati, I., & Priatna, N. (2018). Analisis Kemampuan Literasi Statistis Siswa Madrasah Tsanawiyah dalam Materi Statistika. *Journal of Medives : Journal of Mathematics Education*, *2*(2), 205–212. https://doi.org/10.31331/medives.v2i2.640

[26] Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS Advanced 2015 International Results in Advanced Mathematics and Physics*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2015/international-results/advanced/

[27] Nogues, C. P., & Dorneles, B. V. (2021). Systematic review on the precursors of initial mathematical performance. *International Journal of Educational Research Open*, *2–2*, 100035. https://doi.org/10.1016/J.IJEDRO.2021.100035

[28] Nuralam, & Gadeng, T. (2018). Kualitas Hasil Belajar Statistika Siswa MTs yang dibelajarkan dengan Pendekatan Problem Posing. *Suska Journal of Mathematics Education*,

4(2), 88–99.

[29] OECD. (2019). *PISA 2018 Results: Combined Executive Summaries, Volume I, II & III*. OECD Publishing. https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf

[30] Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267–294). Kluwer Academic Publishers.

[31] Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1–2), 27–46. https://doi.org/10.1080/10986065.2011.538302

[32] Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*. International Statistical Institute. http://iaseweb.org/documents/papers/icots7/6A2_%0APFAN.pdf

[33] Prahmana, R. C. I., Yunianto, W., Rosa, M., & Orey, D. C. (2021). Ethnomathematics: pranatamangsa system and the birth-death ceremonial in Yogyakarta. *Journal on Mathematics Education*, *12*(1), 93–112. https://doi.org/10.22342/jme.12.1.11745.93-112

[34] Ramadhani, R., & Fitri, Y. (2020). A Project-based learning into flipped classroom for ePUB3 electronic mathematics learning module (eMLM)-based on course design and implementation. *Universal Journal of Educational Research*, *8*(7), 3119–3135. https://doi.org/10.13189/ujer.2020.080740

[35] Ramadhani, R., & Narpila, S. D. (2018). Peningkatan Kemampuan Berpikir Statistik Melalui Problem-Based Learning Berbudaya Lokal pada Siswa Kelas XI SMA YPK Medan. *Jurnal PARADIKMA*, *11*(1), 74–84.

[36] Roell, M., Viarouge, A., Houdé, O., & Borst, G. (2017). Inhibitory control and decimal number comparison in school-aged children. *PLOS ONE*, *12*(11), e0188276. https://doi.org/10.1371/JOURNAL.PONE.0188276

[37] Rosa, M., & Orey, D. C. (2017). STEM education in the Brazilian context: An ethnomathematical perspective. In *STEM Education in the Junior Secondary: The State of Play* (pp. 221–247). Springer, Singapore. https://doi.org/10.1007/978-981-10-5448-8_11

[38] Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualisation software. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*. International Statistical Institute. http://iaseweb.org/documents/papers/icots7/2D3_RUBI.pdf

[39] Rufiana, I. S., Sa'dijah, C., Subanji, Susanto, H., & As'ari, A. R. (2018). Informal Statistical Reasoning of Students Taken Formal Statistics Learning Related to Distribution. *International Journal of Insight for Mathematics Teaching*, *01*(2), 130–140.

[40] Setiawan, E. P., & Sukoco, H. (2021). Exploring first year university students' statistical literacy: A case on describing and visualizing data. *Journal on Mathematics Education*, *12*(3), 427–448. https://doi.org/10.22342/jme.12.3.13202.427-448

[41] Sharma, S. (2017, January 26). *Definitions and models of statistical literacy: a literature review*. Open Review of Educational Research; Routledge. https://doi.org/10.1080/23265507.2017.1354313

[42] Soeharto, S. (2021). Development of A Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach. *Journal of Turkish Science Education*, *18*(3), 351–370. https://doi.org/10.36681/tused.2021.78

[43] Soeharto, S., & Csapó, B. (2021). Evaluating Item Difficulty Patterns for Assessing Student Misconceptions in Science Across Physics, Chemistry, and Biology Concepts. *Heliyon*, e08352. https://doi.org/10.1016/J.HELIYON.2021.E08352

[44] Stone, A., Allen, K., Rhoads, T. R., Murphy, T. J., Shehab, R. L., & Saha, C. (2003). The statistics concept inventory: A pilot study. *Proceedings - Frontiers in Education Conference, FIE*, *1*, T3D1-T3D6. https://doi.org/10.1109/FIE.2003.1263336

[45] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.

[46] Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/J.1751-5823.1999.TB00442.X

[47] Wyse, A. E., & Mapuranga, R. (2009). Differential Item Functioning Analysis Using Rasch Item Information Functions. *International Journal of Testing*, *9*(4), 333–357. https://doi.org/10.1080/15305050903352040

[48] Yaqoob, I., Hashem, I. A. T., Ahmed, A., Kazmi, S. M. A., & Hong, C. S. (2019). Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. *Future Generation Computer Systems*, *92*, 265–275. https://doi.org/10.1016/J.FUTURE.2018.09.058

[49] Zieffler, A., Garfield, J., DelMas, R., & Björnsdóttir, A. björg. (2010). Development of an instrument to assess statistical thinking. *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*.

[50] Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A Framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40–58. http://www.stat.auckland.ac.nz/serj

[51] Zulkardi, Z., Putri, R. I. I., & Wijaya, A. (2020). Two Decades of Realistic Mathematics Education in Indonesia. In *International Reflections on the Netherlands Didactics of Mathematics* (pp. 325–340). Springer, Cham. https://doi.org/10.1007/978-3-030-20223-1_18