

## Measuring changes in students' informal statistical reasoning skills through the ethno-flipped classroom model: Stacking and racking analysis

**Rahmi Ramadhani<sup>1</sup>, Soeharto<sup>2</sup>, Fitria Arifiyanti<sup>3</sup>, Zsolt Lavicza<sup>4\*</sup>, Edi Syahputra<sup>5</sup>,  
Elmanani Simamora<sup>5</sup>**

<sup>1</sup>Department of Informatics, Universitas Potensi Utama, North Sumatra, Indonesia

<sup>2</sup>Research Center for Education, National Research and Innovation Agency (BRIN), Indonesia

<sup>3</sup>Department of Physics Education, Universitas Pendidikan Indonesia, West Java, Indonesia

<sup>4</sup>Department of STEM Education, Linz School of Education, Johannes Kepler University Linz, Austria

<sup>5</sup>Department of Mathematics Education, Universitas Negeri Medan, North Sumatra, Indonesia

\*Correspondence: [zsolt.lavicza@jku.at](mailto:zsolt.lavicza@jku.at)

**Received:** Apr 14, 2025 | **Revised:** Jul 22, 2025 | **Accepted:** Jul 29, 2025 | **Published Online:** Oct 10, 2025

### Abstract

Statistical reasoning is widely recognized as a fundamental skill for preparing students to navigate a data-driven global society. However, research on the development of informal statistical reasoning, particularly when facilitated through an ethno-flipped classroom model, remains limited. Furthermore, few studies employ longitudinal approaches, such as stacking and racking, to assess continuous changes in students' reasoning development over time. Most existing research relies on pre–post comparisons and lacks comprehensive analysis at both the person and item levels. This study investigates changes in students' informal statistical reasoning using the Partial Credit Rasch Model, incorporating stacking and racking analysis. A total of 152 twelfth-grade students participated in a 12-week ethno-flipped classroom intervention. The informal statistical reasoning test, consisting of five open-ended items, demonstrated sufficient validity for use in both the instructional intervention and the Rasch-based analysis. Item validity was assessed as a key parameter within the Rasch measurement framework. The findings revealed significant improvements in student proficiency and a decrease in item difficulty. Notably, 66 students reached Level 5, reflecting integrated process reasoning. These results support the effectiveness of the ethno-flipped classroom model in enhancing informal statistical reasoning. This study contributes to the design of contextual, adaptive instruction and offers a robust Rasch-based framework for monitoring reasoning development longitudinally.

### Keywords:

Ethno-flipped classroom, Informal statistical reasoning, Racking analysis, Stacking analysis

### How to Cite:

Ramadhani, R., Soeharto, S., Arifiyanti, F., Lavicza, Z., Syahputra, E., & Simamora, E. (2025). Measuring changes in students' informal statistical reasoning skills through the ethno-flipped classroom model: Stacking and racking analysis. *Infinity Journal*, 14(4), 949-972.  
<https://doi.org/10.22460/infinity.v14i4.p949-972>

*This is an open access article under the CC BY-SA license.*



## 1. INTRODUCTION

In global mathematics education, there is widespread recognition that mathematical reasoning is a critical skill, closely tied to the execution of mathematical procedures (Schleicher, 2023; Smit et al., 2023) and essential for understanding concepts across other scientific disciplines (Kollosche, 2021; Valenta et al., 2024). In recent years, the field has placed increasing emphasis on statistical literacy, particularly statistical reasoning, as a vital competency for navigating real world and data rich environments. Statistical reasoning equips individuals with the ability to make sense of data, draw inferences, and solve practical problems using statistical information (Fielding et al., 2025; Hasanah et al., 2024). Within this broader domain, informal statistical reasoning plays a foundational role, particularly in early learning context. It refers to the ability to interpret informal statistical information in informal settings and to formulate questions, model data, and draw reasonable conclusions, all without relying on formal inferential procedures (Hjelte et al., 2020; Makar, 2013; Pfannkuch, 2011). As a subdomain of statistical reasoning, informal statistical reasoning is essential not only for academic progression but also for developing critical data literacy in everyday decision-making (Setianingsih & Rahmah, 2020; Sukoriyanto & Khurin'in, 2023).

The growing importance of informal statistical reasoning in mathematics education is reflected in various international curriculum frameworks. For instance, the GAISE report by the American Statistical Association (Franklin et al., 2007) outlines four key components of the statistical reasoning process, posing questions, collecting data, analyzing data, and interpreting results, that align with NCTM's principles (National Council of Teachers of Mathematics, 2000). Several countries, including the UK (Department for Education, 2014), China (Ministry of Education of People's Republic of China, 2003), Singapore (Ministry of Education Singapore, 2023), and Indonesia (Ministry of Education Culture Research and Technology, 2023) have integrated informal statistical reasoning into their national curricula to enhance students' ability to analyze data and solve real-life problems.

Informal statistical reasoning refers to a skill that emphasizes how students use their logic to gather mathematical information, which involves obtaining information from random statistical data (Pfannkuch, 2006). It is a type of reasoning that is closely related to probabilities and helps to draw conclusions and make decisions in situations that aren't expressed in terms of recognized probabilities (Sariningsih & Herdiman, 2017). This form of reasoning allows students to understand statistical data informally and to form conclusions and judgments based on informally acquired data from specific observational studies, experiments, or sample surveys (Hjelte et al., 2020; Pfannkuch, 2011). Developing informal statistical reasoning is essential for building foundational skills in statistical inference, enabling students to observe, compare, and draw conclusions from data distributions. These skills support generalizing data findings and are assessed through five reasoning levels: level 1 (idiosyncratic reasoning), level 2 (verbal reasoning), level 3 (transitional reasoning), level 4 (procedural reasoning), and level 5 (integrated reasoning). The informal statistical reasoning have four key indicators: describing, organizing, representing, and interpreting data (Jones et al., 2000; Makar & Fielding-Wells, 2011; Makar & Rubin, 2009; Mooney, 2002).

Informal statistical reasoning skills are inherently connected to data interpretation and can be meaningfully contextualized within the framework of ethnomathematics. This approach was demonstrated by Ramadhani, Saragih and Napitupulu (2022), who investigated students' informal statistical reasoning through assessments embedded in ethnomathematical contexts. Their findings revealed that students exhibited varying levels of informal statistical reasoning, which were associated with their initial mathematical competencies. These outcomes suggest that incorporating ethnomathematics-based contexts into assessment tools holds potential for effectively mapping students' reasoning abilities. However, the study did not involve the implementation of specific instructional interventions, focusing solely on identifying existing ability levels. In contrast, other studies assessing informal statistical reasoning have not integrated ethnomathematical contexts into their instruments. The limited use of ethnomathematics contexts in assessing informal statistical reasoning represents a key novelty of this study. Specifically, this research incorporates ethnomathematics-based contexts into statistical problems designed to evaluate students' informal statistical reasoning skills. While previous studies have explored students' reasoning levels, they did not involve specific instructional interventions using cultural contexts, thereby highlighting a gap that this study seeks to address.

Several researchers have demonstrated that integrating ethnomathematics into mathematics education enhances student engagement and provides meaningful learning experiences (Furuto, 2016, 2018; Hidetoshi & Rothman, 2021; Kim & Chae, 2016; Lipka et al., 2005). Beyond cultural contextualization, instructional design models such as the flipped classroom have gained popularity for enabling flexible, interactive, and student-centered learning (Attard & Holmes, 2020). Encouraging peer collaboration within this model further increases students' confidence, interest, motivation, and adaptability when using technology to learn mathematics (Abeysekera & Dawson, 2015).

The ethnomathematics context in this study was integrated with the flipped classroom model. Such an integration was previously implemented by Ramadhani et al. (2023a). This integration resulted in a versatile learning environment that operates in two phases: out-of-class and in-class, designed to foster meaningful learning by incorporating contexts, activities, cultural elements, values, and characteristics embedded in students' sociocultural experiences. Engaging with problems that reflect students' cultural backgrounds and daily activities enhances their ability to recognize, simplify, visualize, analyze, and draw predictions from data in contexts relevant to their lives. While previous research demonstrated the potential of this integration, the current study introduces a further innovation by applying the ethno-flipped classroom model specifically to assess students' informal statistical reasoning, a focus that has not been sufficiently addressed in earlier studies.

To further elaborate, the ethno-flipped classroom model is an innovative pedagogical framework that combines the flexibility and technology integration of the flipped classroom (Attard & Holmes, 2020; Ramadhani et al., 2023b) with the cultural relevance of ethnomathematics (D'Ambrosio & de Campos Almeida, 2017; Prahmana et al., 2021; Rosa & Orey, 2016). By embedding learning in students' cultural contexts, it supports adaptive, student-centered learning experiences. Collaborative and investigative learning through

tiered group work enhances character development, self-regulated learning, and socio-cultural values such as cooperation and perseverance (Ramadhani, Syahputra, et al., 2022; Ramadhani et al., 2023b). Studies show that implementing this model using contexts like Lemang Batok or Nias traditions improves numeracy and informal statistical reasoning (Ramadhani et al., 2024; Ramadhani et al., 2025), confirming the model's value in promoting critical thinking and statistical literacy in real-world cultural contexts.

While the pedagogical model offers promise, rigorous evaluation of its impact on students' reasoning skills remains limited in the literature. This study therefore draws on previous intervention-based research that seeks to assess learning gains meaningfully. Numerous studies have examined improvements in statistical reasoning skills before and after interventions (Conway et al., 2019; Ganesan & Leong, 2020). However, only a few have explored these changes in depth to identify meaningful differences between pre and post-test changes in depth to identify meaningful differences between pre and post-test phases. Ramadhani, Saragih, Maulida, et al. (2022) examined changes in statistical reasoning skills using a TinkerPlots-assisted ethnomathematics approach. Laliyo et al. (2022) investigated shifts in students' understanding of hydrological concepts through the implementation of an inquiry-based model. Ling et al. (2018) analyzed changes in students' initial mathematical abilities using educational games as a learning medium. Hu et al. (2022) assessed variations in self-efficacy in science learning through both cross-sectional and longitudinal analyses. Meanwhile, Amin et al. (2013) explored developments in students' readiness for self-directed learning. Most studies merely confirm the presence of an effect without explaining the magnitude or nature of the improvement in student ability (Laliyo et al., 2022). Many focus solely on identifying student performance at a fixed point rather than investigating how specific instructional interventions lead to progressive development over time (Ilie et al., 2024). As a result, they overlook the dynamic nature of learning and fail to capture longitudinal shifts in student understanding.

These limitations highlight a core concern in the psychometric community about the accuracy of measuring changes over time. Central to this issue is the use of raw scores in Classical Test Theory (CTT), which has been widely criticized for introducing measurement bias (Sumintono, 2018). Raw scores lack the precision required for valid interpretation. Specific concern include: (1) gain scores tend to correlate negatively with pre-test performance (Pentecost & Barbera, 2013); (2) they often show low reliability (Willoughby & Metz, 2009); and (3) scale inconsistency reduces measurement accuracy (Linn & Slinde, 1977).

To address these limitations, more advanced psychometric approaches are needed. Stacking and racking analysis is an analytic techniques within the Rasch Measurement framework, offer a more rigorous and accurate way to evaluate learning gains. Both are grounded in Item Response Theory (IRT), which focuses on modeling the probability of a specific response based on both item characteristics and person ability. These techniques allow researchers to simultaneously examine changes in individual or person ability and item difficulty across time points, making them particularly suitable for longitudinal educational studies in educational settings (Amin et al., 2013; Sunjaya et al., 2021). Stacking arranges each student's data vertically, one row for the pre-test and one for the post-test, to track

individual changes across identical items (Combrinck et al., 2016). Racking, in contrast, places the data horizontally, allowing each test item to appear twice while each student appears once, enabling analysis of changes in item difficulty over time (Wright, 2003). These techniques help evaluate shifts in student abilities—such as informal statistical reasoning and self-regulated learning, by comparing item positions and difficulty levels on the Wright map. Building on this framework, stacking and racking techniques significantly enhance the reliability and interpretability of assessments by: (1) providing detailed insights into changes in both student ability and item difficulty; (2) improving the precision of skill measurement beyond what conventional methods allow; (3) accurately monitoring learning progress over time; and (4) ensuring the equitability and observability of ability shifts within individuals and groups (Uzun & Öğretmen, 2021).

This study aims to examine changes in students' informal statistical reasoning skills by comparing their performance on pre-tests and post-tests following the implementation of the ethno-flipped classroom model. The analysis considers both students' proficiency levels and the difficulty of test items to identify trends and significant differences in their reasoning abilities when solving statistical problems within ethnomathematical contexts. To guide this investigation, the following research questions were formulated.

- RQ1: Does the data collected from a sample of high school students fit the Rasch Model parameters?
- RQ2: Are there any changes in students' informal statistical reasoning skills after the intervention using an ethno-flipped classroom model based on person ability?
- RQ3: Are there any changes in students' informal statistical reasoning skills after the intervention using an ethno-flipped classroom model, based on item test difficulty?
- RQ4: Are there any differences between the students' informal statistical reasoning skills at Time 1 (pre-test) and Time 2 (post-test)?

## 2. METHOD

### 2.1. Research Design

This study applied quantitative research with a one-group pre-test/post-test design. The research data were analyzed using the stacking analysis technique and racking analysis technique on the Rasch Model Measurement. The stacking analysis technique was used to measure changes in students' informal statistical reasoning ability after the intervention of the ethno-flipped classroom model in terms of the level of ability per student. In contrast to the stacking analysis technique, the racking analysis technique was used to measure changes in the students' informal statistical reasoning skills in terms of difficulty levels per test item.

### 2.2. Participants

Participants in this study were final-year students at senior high school level on Nias Island, Indonesia. The study sample was selected using stratified random sampling. The total sample of this study was 235 students, but only 152 of the 12th-grade students were included in the stacking and racking analysis (56% female and 44% male) with an age range of 17-18 years old from six high schools in Nias Island, Indonesia. The research sample participated

in a 14-week ethno-flipped classroom learning experience divided into two learning phases (six out-of-class learning phases and six in-class learning phases), one initial assessment phase (Time 1-Pretest), and one final assessment phase (Time 2-Posttest). The topic taught was statistics, consisting of three sub-topics: measures of data presentation, measures of central tendency, and measures of dispersion.

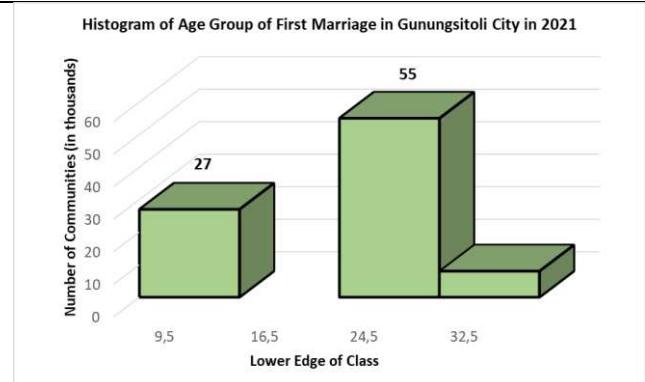
The decision to include only 152 students was based on data completeness. A total of 82 students did not fully participate in the 12 week intervention and failed to complete either the pre-test or the post-test. Consequently, students who were absent for one of the assessments or did not participate consistently throughout the intervention were excluded from the analysis. This exclusion was necessary to avoid bias and distortion in estimating person ability and item difficulty due to incomplete data. Data completeness is a critical requirement in longitudinal studies utilizing stacking and racking analysis, as accurate and valid measurement at both time points (Time 1-pre-test and Time 2-post-test) is essential for meaningful comparisons.

### 2.3. Research Instrument

The instrument used in this study was an informal statistical reasoning test. The test that we designed and developed refers to the statistical learning indicators adopted from Jones et al. (2000) and Mooney (2002), which describe data, organize and reduce data, represent data, and analyze and interpret data. Describing data involves the direct reading and reporting of information presented in charts, tables, or other graphical displays. Organizing and reducing data refers to the processes of classifying, structuring, or summarizing data, often through the use of measures of central tendency and variability. Data representation entails the graphical display of data to facilitate understanding and comparison. Analyzing and interpreting data involves identifying patterns or trends and making predictions or inferences based on the information presented in graphical formats. The test was developed based on the level of informal statistical reasoning ability using a five-item test. Table 1 presents one of the informal statistical reasoning ability test items developed.

**Table 1.** Informal statistical reasoning ability test item

Informal Statistical Reasoning Ability Test Item	Solution
<p>One of the kinship systems of the Nias community is derived from the marriage procession. In Nias culture, marriage is an obligation and responsibility of the parents ("No ibu'a wo'omö nia khö ndraono nia"). The implementation of the marriage process is also carried out in cooperation, following the <i>Amaedola</i> held by the Nias community, namely "<i>Aoha noro nilului wahea, aoha noro nilului waoso, alisi tafadaya-daya, hulu tafewolo-wolo</i>". The following data is given in a histogram related to data on the age of first marriage of Nias people in Gunungsitoli city in 2021 (data source: <a href="https://gunungsitilikota.bps.go.id">https://gunungsitilikota.bps.go.id</a>). The data displayed is in the form of rounding units of thousands.</p>	<p>It is known that the number of Nias people who have their first marriage from the age range of 10-40 years is 114 thousand. The next information obtained is the ratio of Nias people who have their first marriage in the age range of 33-40 years and 17 - 24 years is 1:3 (<i>Students extract the ideoesyncratic or relevant information from the graph presented</i>).</p> <p>Suppose the number of Nias people who had their first marriage in the age range of 33-40 years is <math>x</math>, then:</p> $114 = 27 + 3x + 55 + x$

Informal Statistical Reasoning Ability Test Item	Solution								
<p>Histogram of Age Group of First Marriage in Gunungsitoli City in 2021</p>  <table border="1"> <caption>Data from Histogram</caption> <thead> <tr> <th>Lower Edge of Class</th> <th>Number of Communities (in thousands)</th> </tr> </thead> <tbody> <tr> <td>9,5 - 16,5</td> <td>27</td> </tr> <tr> <td>16,5 - 24,5</td> <td>55</td> </tr> <tr> <td>24,5 - 32,5</td> <td>8</td> </tr> </tbody> </table> <p>If there are 114 thousand Nias people who have their first marriage from the age range of 10-40 years, and the ratio of Nias people who have their first marriage in the age range of 33-40 years and 17 - 24 is 1:3. Analyze how many people in Gunungsitoli city had their first marriage between the ages of 17 and 24!</p>	Lower Edge of Class	Number of Communities (in thousands)	9,5 - 16,5	27	16,5 - 24,5	55	24,5 - 32,5	8	$114 = 82 + 4x$ $4x = 114 - 82$ $4x = 32$ $x = 8$ <p><i>(Students produce ideosyncratic information that complement the graph presented)</i></p> <p>Thus, Nias people who have their first marriage in the age range of 17 years to 24 years are <math>3x = 3(8) = 24</math> thousand</p> <p><i>(Students complete the ideosyncratic information that has been generated to complete the graph presented correctly and completely)</i></p>
Lower Edge of Class	Number of Communities (in thousands)								
9,5 - 16,5	27								
16,5 - 24,5	55								
24,5 - 32,5	8								

**Table 1** illustrates that the informal statistical reasoning ability test incorporates problems framed within the cultural context of Nias. This integration aligns with the implementation of the ethno-flipped classroom model, which emphasizes presenting tasks rooted in local ethnomathematical contexts. Specifically, the use of ethnomathematical elements corresponds to the second component of the ethno-flipped classroom model, namely Cultural Experience (see [Figure 1](#)). Developing informal statistical reasoning assessments that integrate ethnomathematics contexts presents several key challenges. First, it requires balancing cultural authenticity with the psychometric validity of test items; the cultural context must support rather than hinder students' reasoning processes. Second, there is a potential for construct-irrelevant variance, where differences in students' familiarity with cultural content may distort the measurement of their true informal statistical reasoning abilities. Third, it is necessary to ensure that the ethnomathematics context aligns consistently with established indicators of informal statistical reasoning abilities, maintaining both cultural and educational coherence.

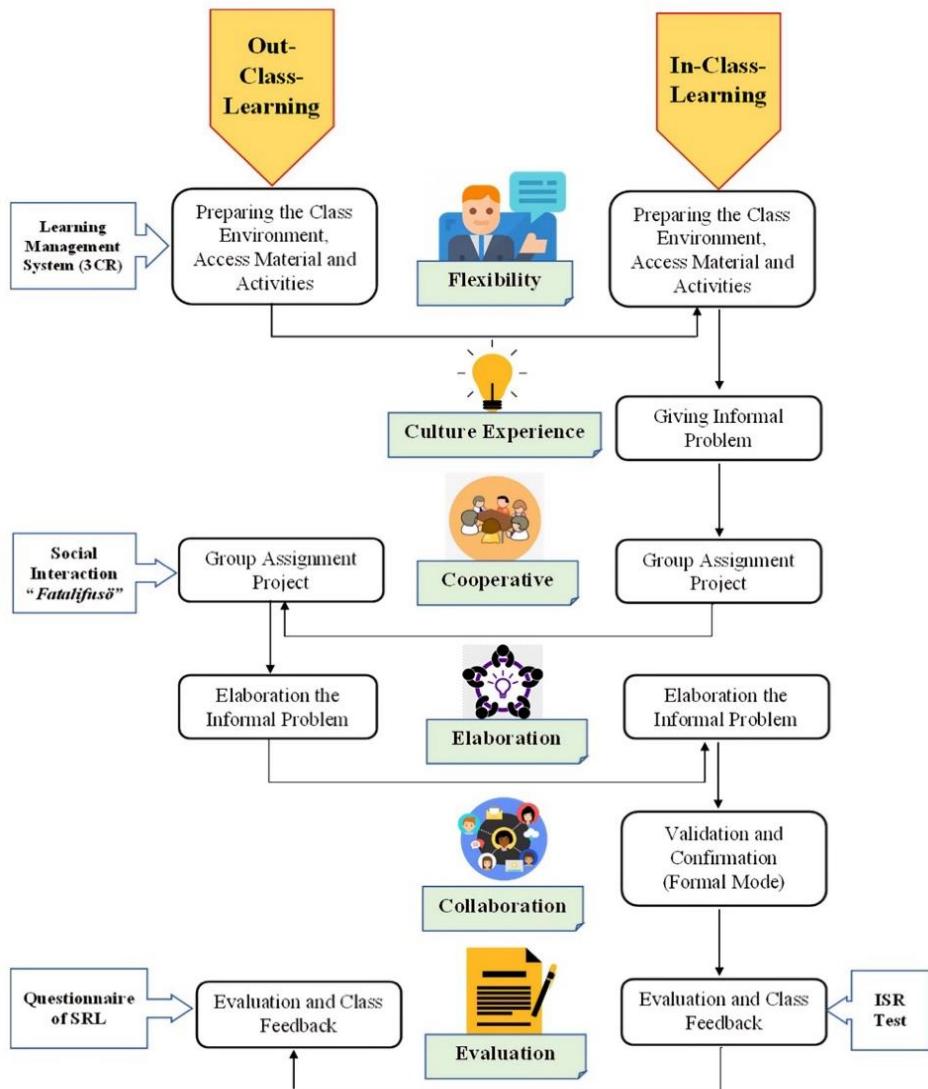
#### 2.4. Procedure

The research was conducted from September to November 2022 and began by administering a pretest (Time-1) to all research samples, and ethical approval letter was obtained from Universitas Potensi Utama. The model intervention was conducted by referring to the syntax of the ethno-flipped classroom model for all research samples for 12 weeks and was conducted using two learning phases (out-of-class learning and in-class learning for each learning cycle). The intervention was conducted in three learning cycles, where one cycle studied one subtopic. The posttest (Time-2) was administered at the end of the third cycle. The first cycle of learning began with the out-of-class learning phase by focusing on the preparation of the learning environment, technology adaptation, and the initial introduction of statistical problems in the context of ethnomathematics; in this case,

using the context of ethnomathematics artifacts of the Nias tribe. After learning with the ethno-flipped classroom model, the final phase was carried out by providing the final assessment (Posttest Time 2). The test given in the post-test session (Time 2) was the same as the test given in the pre-test session (Time 1). The initial test data (pretest) were collected during the first week of the activity, while the final data (posttest) were gathered at the end of the final week.

The lesson began by setting up the learning environment through a pre-arranged learning management system, where all instructional modules are easily accessible. This process is part of the initial phase outlined in the flipped classroom model of out-of-class learning. During this initial out-of-class phase, students explored the material available on the LMS and engaged in discussions with their peers (Syntax 1). The insights gained from these discussions are then used for further exploration in the subsequent in-class learning phase, while the initial phase of the environment, activities, and material preparation continues. In the in-class learning phase, students were presented with informal problems rooted in the sociocultural context of the Nias tribe for discussion (Syntax 2). Prior to these discussions, the teacher divided the students into learning groups based on their proficiency levels (Syntax 3). The day before the learning intervention, the teacher conducted an initial assessment of the students' informal statistical reasoning skills through a pretest. The results of this initial assessment form the basis for grouping students.

In addition, students engaged in collaborative study groups to address informal statistical problems, beginning in the in-class learning phase and continuing in the out-of-class learning phase through the discussion tool built into the LMS. During the out-of-class phase, students extended the collaborative session by expanding the initial discussion results (Syntax 4). This expansion process continued into the in-class phase through additional elaboration activities, with the resulting elaborations serving as validation and confirmation material. The validation and confirmation tasks (Syntax 5) were performed through graded collaborative activities. Representatives of low-ability groups confirmed the earlier elaboration results for the medium- and high-ability groups. Similarly, middle-ability student groups verified and validated their elaborations with those of high-ability student groups. At this stage, representatives from these student groups reached consensus on the results. In addition, representatives from the high-ability groups presented agreed-upon results to the teacher for validation purposes. The learning activity culminated in a final assessment designed to evaluate students' proficiency after the learning intervention (Syntax 6). [Figure 1](#) illustrates the sequence of implementation of the learning process using the ethno-flipped classroom model.



**Figure 1.** Illustrates the sequence of implementation of the learning process using the ethno-flipped classroom model

## 2.5. Data Analysis

Rasch Measurement with stacking and racking analysis techniques assisted by WINSTEPS application. The analysis test began by checking the validity of the informal statistical reasoning test instrument used in this study. Fit validity analysis was conducted using the Rasch Model Measurement concerning certain criteria. The criteria in the fit validity analysis on the informal statistical reasoning ability test are using three values, namely the Outfit Mean Square (MNSQ) value between 0.5 and 1.5, the Outfit Z-Standard (ZSTD) value between -2.0 and 2.0, and the Point Measure Correlation (Pt. Mean Corr) value between 0.4 and 0.85 (Sumintono & Widhiarso, 2015). Further analysis was performed to strengthen the results of the fit validity analysis by using a dimensionality test. The dimensionality test criterion is whether the test instrument developed can measure the range of variables or the ability of the research subject to answer question items if the Raw Variance Explained by Measures is above 40% (Ekinci et al., 2023; Soeharto et al., 2024; Sukarelawan et al., 2021). The test instrument can be used in research if Cronbach's alpha

reliability value is in the interval  $0.70 \leq KR20 \leq 0.80$  or has a reliability measurement interpretation value of "Good" (Bond & Fox, 2015). The next stage of analysis is stacking, and racking analysis is used to check for changes based on persons and items. Finally, the Wilcoxon signed-rank test using SPSS version 25 was used to measure the difference in students' informal statistical reasoning skills.

### 3. RESULTS AND DISCUSSION

#### 3.1. Results

##### *RQ1: Fit Validity Based on Rasch Model Parameters*

**Table 2.** Construct validation test results on the informal statistical reasoning (ISR) ability test instrument using the Rasch analysis

No.	Item Code	Total Score	JMLE Measure	MNSQ Outfit	ZSTD Outfit	PT Measure Corr.
1.	ISR1	802	-1.12	0.81	-2.18	0.84
2.	ISR2	783	-0.92	0.98	-0.18	0.80
3.	ISR3	683	-0.23	1.15	1.62	0.81
4.	ISR4	743	-0.68	1.06	0.66	0.79
5.	ISR5	481	2.49	0.86	-1.20	0.80

The results of the fit validity analysis of the informal statistical reasoning test based on Table 2 show that all item codes meet the fit validity criteria. Although in item ISR1 the Outfit ZSTD value is greater than, this value is still acceptable and is concluded to still meet the valid category (Bond & Fox, 2015). The Cronbach's alpha reliability was categorized as excellent (see Table 3). Further analysis showed that the person separation index for the construct validation of the informal statistical reasoning instrument obtained a value of more than 2. Planinic et al. (2019), and Soeharto and Csápo (2022) stated that the separation value in person separation must be more than 2 logits, where the greater the separation index, the higher the quality of the test. The results of the construct validation analysis of the informal statistical reasoning test indicate that the instrument is capable of effectively measuring the range of abilities among the research participants in responding to the test items. This is supported by the raw variance explained by the measures of item dimensionality, which exceeded 40% (71.1%).

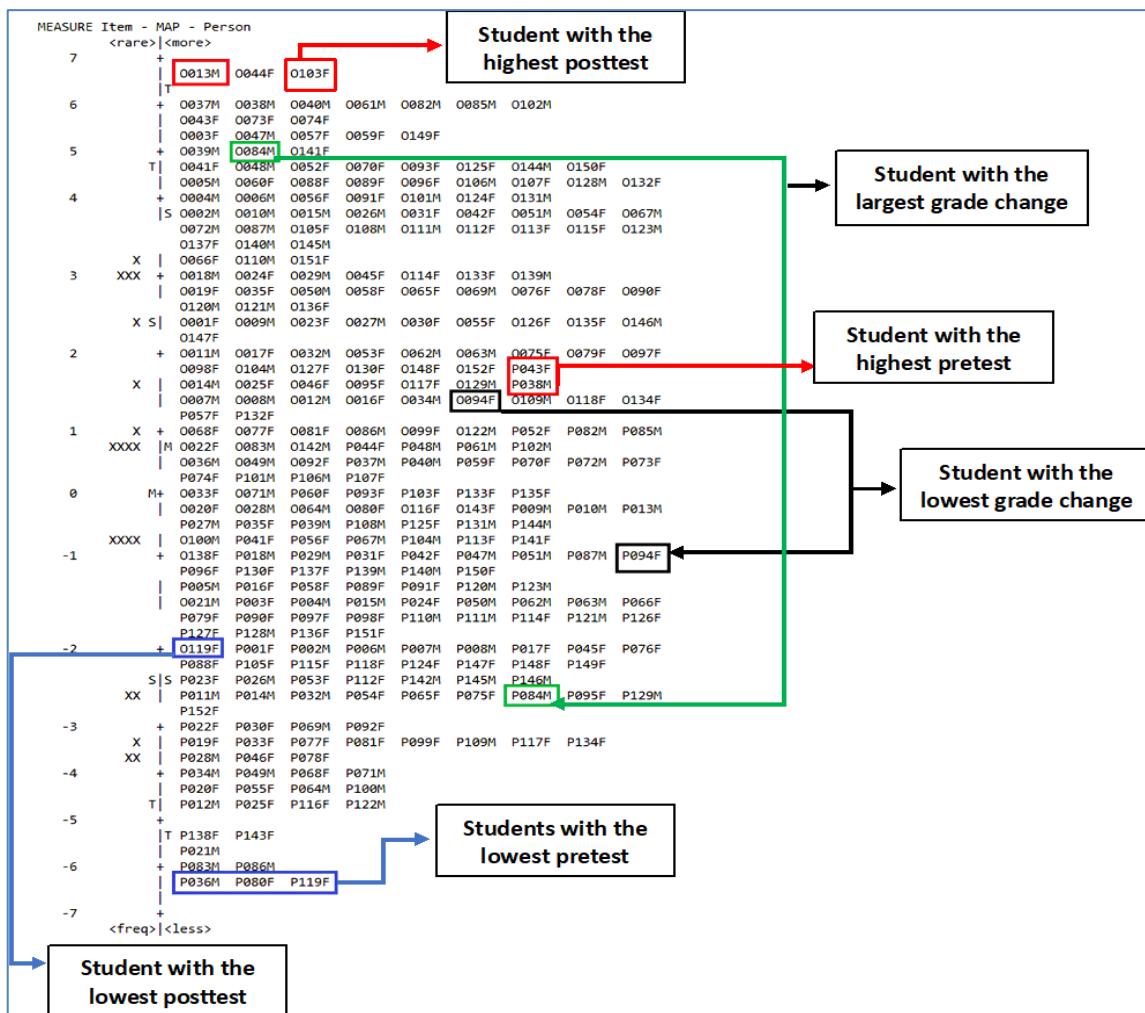
**Table 3.** Instrument reliability and model fit

	Persons	Items
N	235	5
Measure	0.06	0.00
Mean	14.9	698.4
SD	2.29	1.33
SE	0.77	0.11
Mean Outfit MNSQ	0.97	0.97
Mean Outfit ZSTD	-0.06	-0.26
Separation	2.78	11.81
Reliability	0.89	0.99
Cronbach's Alpha	0.90	
Raw variance of Unidimensionality	71.1%	

The test item coded ISR5 is a question with a very hard difficulty level. Furthermore, the question with code ISR3 is a question with a difficulty level of hard, the question with code ISR4 is a question with a difficulty level of Medium, and the question with code ISR1 and code ISR2 are questions with a difficulty level of Easy.

**RQ2: Changes in Students' Informal Statistical Reasoning Skills After Intervention Using Ethno-Flipped Classroom Model Based on Person Ability**

Changes in students' informal statistical reasoning skills were examined using pre-test and post-test scores collected from four learning classes, comprising a total sample of 152 students. The pre-test (Time 1) and post-test (Time 2) data were analyzed using the stacking analysis technique, which is incorporated within the Rasch model measurement framework. This analysis yielded logit values that were used to assess shifts in informal statistical reasoning skills before and after the implementation of the ethno-flipped classroom model, as illustrated in [Figure 2](#).



**Figure 2.** Wright map of changes in informal statistical reasoning ability in pre-test condition (time 1) and post-test condition (time 2) in terms of students' individual ability levels

Referring to [Figure 2](#), it was found that students with code 094F had the smallest change in logit value compared to the other students. Students with code 094F had a negative

logit value in the pre-test condition (Time 1) but increased to a positive logit value in the post-test condition (Time 2). In contrast to student code 084M, where the student has the largest change in logit value 7.56.

Figure 2 also shows that student code 119F had the lowest pre-test score (logit value -6.17), and in the post-test condition, student code 119F had the lowest logit value (logit value -1.92). In the post-test condition, student code 119F also had the lowest logit value. Students coded 043F and 038M as students with the highest pre-test scores. Different conditions were seen in the post-test scores, where students with codes 013M, 044F, and 103F were the group of students with the highest post-test scores. This finding shows evidence that students who have been given the intervention of an ethno-flipped classroom learning model show positive changes. The positive change is reinforced by the Logit Value Person value, which shows the change in the level of students' informal statistical reasoning ability from the pre-test condition (Time 1) to the post-test condition (Time 2), as shown in Tables 4 and 5.

**Table 4.** Logit value person in pretest condition (time-1)

Demographics	Very high, LVP > Mean Logit + 2SD (Level 5)		High, Mean Logit + SD < Mean Logit + 2SD (Level 4)		Moderate, Mean Logit < LVP ≤ Mean Logit + 2SD (Level 3)		Low, Mean Logit - SD < Mean Logit (Level 2)		Very Low, LVP ≤ Mean Logit - SD (Level 1)		Total
	Mean Logit + 2SD (Level 5)	%	Mean Logit + 2SD (Level 4)	%	Mean Logit + 2SD (Level 3)	%	Mean Logit (Level 2)	%	Mean Logit - SD (Level 1)	%	
Gender											
Male	0	0.00	1	1.54	24	36.92	28	43.08	12	18.46	65
Female	0	0.00	2	2.30	21	24.14	45	51.72	19	21.84	87
<b>Total</b>	<b>0</b>		<b>3</b>		<b>45</b>		<b>73</b>		<b>31</b>		<b>152</b>

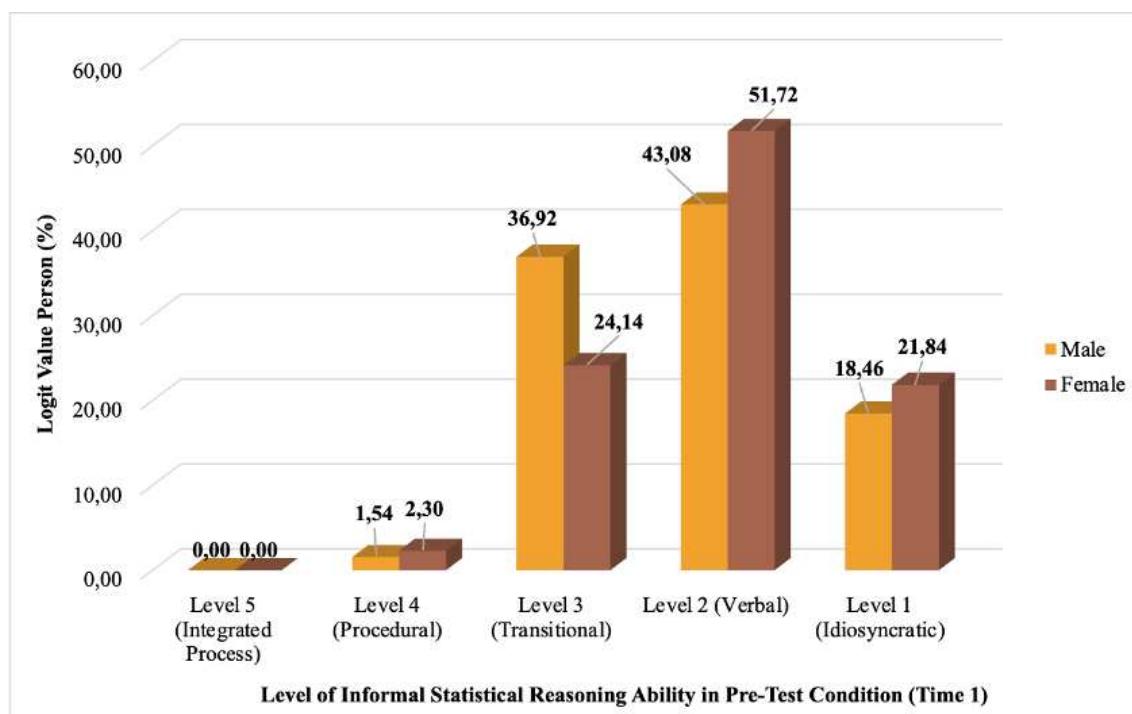
**Table 5.** Logit value person in posttest condition (time-2)

Demographics	Very high, LVP > Mean Logit + 2SD (Level 5)		High, Mean Logit + SD < Mean Logit + 2SD (Level 4)		Moderate, Mean Logit < LVP ≤ Mean Logit + 2SD (Level 3)		Low, Mean Logit - SD < Mean Logit (Level 2)		Very Low, LVP ≤ Mean Logit - SD (Level 1)		Total
	Mean Logit + 2SD (Level 5)	%	Mean Logit + 2SD (Level 4)	%	Mean Logit + 2SD (Level 3)	%	Mean Logit (Level 2)	%	Mean Logit - SD (Level 1)	%	
Gender											
Male	32	49.23	21	32.31	11	16.92	1	1.54	0	0.00	65
Female	34	39.08	39	44.83	12	13.79	2	2.30	0	0.00	87
<b>Total</b>	<b>66</b>		<b>60</b>		<b>23</b>		<b>3</b>		<b>0</b>		<b>152</b>

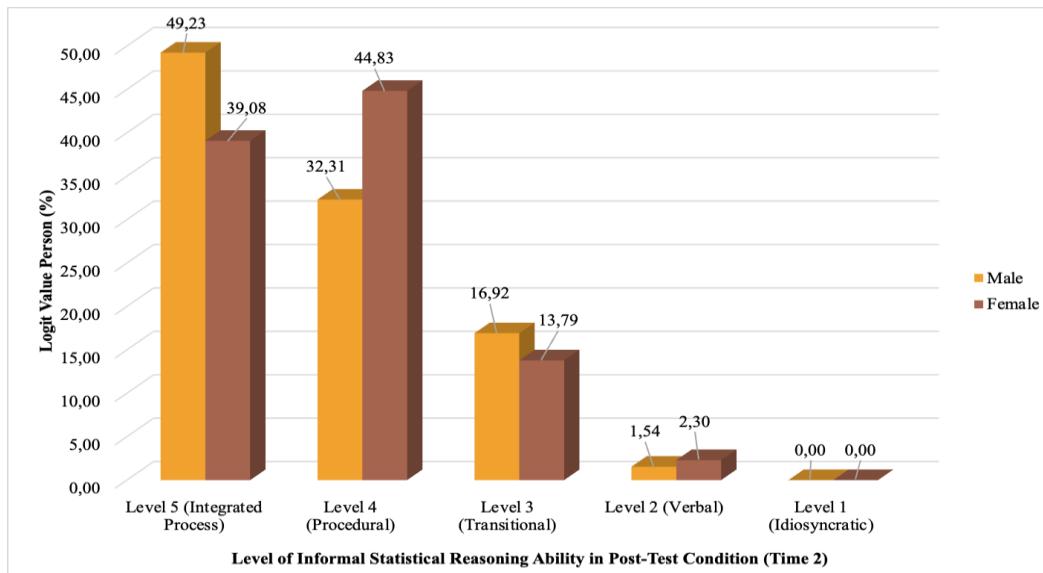
The level of students' informal statistical reasoning ability in the pretest condition (Time 1) was 48% at Level 2 (verbal). In terms of gender demographics, female students dominate Level 2 (Verbal), which is 51.72%, while 43.08% of male students have Level 2 (Verbal) informal statistical reasoning skills. After the intervention of the ethno-flipped classroom model, there was a change in the level of students' informal statistical reasoning skills, which were previously still dominant at level 2 (verbal) to level 5 (Integrated Process)

in the post-test condition (Time 2). As many as 43.42% of the students were at Level 5 (Integrated Process), and as many as 39.47% of the students were at Level 4 (procedural), and only 1.97% of the students remained at Level 2 (verbal) after receiving the intervention of the ethno-flipped classroom model. When reviewed from gender demographics, male students dominate the increase in level of informal statistical reasoning ability to Level 5 (Integrated Process), which is 49.23%, inversely proportional to the increase in level to Level 4 (procedural), students with female gender dominate, which is 44.83%.

Further analysis at the individual student level revealed that students with the lowest pretest and posttest logit scores (code 119F) who were originally in Level 1 (idiosyncratic) moved up to Level 2 (verbal). Another finding was that students with the highest pretest logit scores (codes 043F and 038M), who were originally at level 4 (procedural), moved to level 5 (Integrated Process). Similarly, students with the highest post-test logit scores (codes 013M, 044F, and 103F), who were originally at Level 3 (transitional), moved up to Level 5 (Integrated Process). The student with the highest change in ability level (code 084M) experienced a significant increase in ability level from Level 2 (verbal) to Level 5 (Integrated Process). Although the student with code 094F experienced the smallest increase in skill level, he still showed a positive level increase from Level 3 (Transitional) to Level 4 (Procedural). [Figures 3](#) and [4](#) illustrate the proficiency of the students' informal statistical reasoning skills in both the pre-test (Time 1) and post-test (Time 2) conditions.



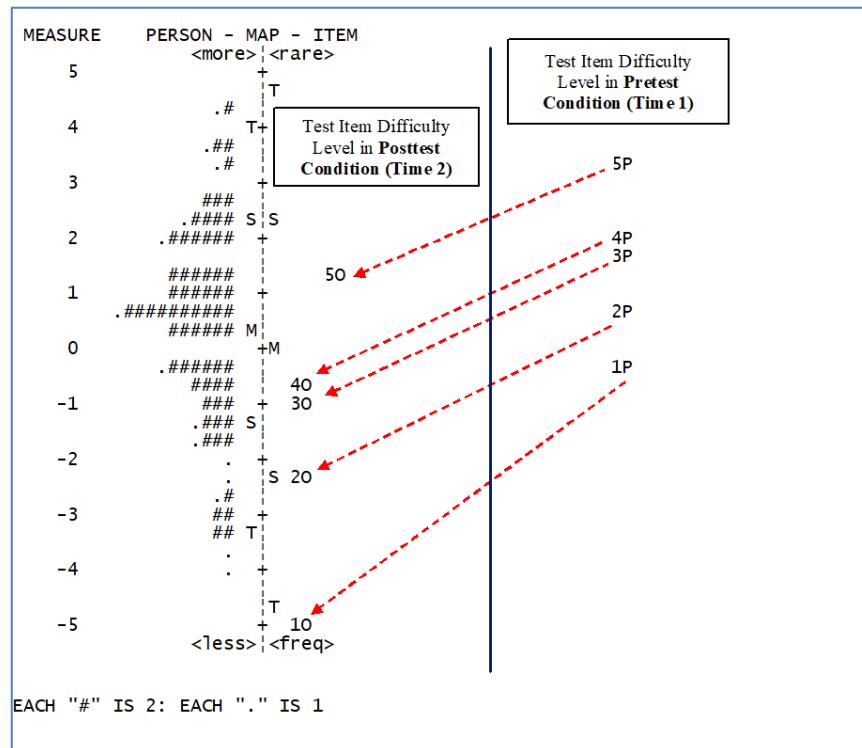
**Figure 3.** Level of informal statistical reasoning skills of the students in the pre-test condition (Time 1)



**Figure 4.** Level of informal statistical reasoning skills of the students in the post-test condition (Time 2)

### ***RQ3: Changes in Students' Informal Statistical Reasoning Skills After Intervention Using Ethno-Flipped Classroom Model Based on Test Difficulty***

Changes in students' informal statistical reasoning skills were analyzed using Racking Analysis. The Racking Analysis technique aims to identify changes in item difficulty that impact changes in students' informal statistical reasoning abilities, as presented in [Figure 5](#).



**Figure 5.** Wright map of changes in informal statistical reasoning ability in pre-test (time 1) and post-test (time 2) conditions in terms of test item difficulty levels

The test item with the code ISR5 in the pre-test condition(3.43 logit), which was the most difficult test item for students to solve, changed the logit value to(1.22 logit). Although there was a change in the difficulty level of item ISR5 in the pretest condition (code 5P) and posttest condition (code 5O), item ISR5 remained at the most difficult item difficulty level compared to other test items. Likewise, in item ISR1, where in the pre-test condition (code 1P) ISR1 was the easiest test item for students to solved (-0.39 logit) and in the post-test condition (code 1O) (-4.97 logit) experienced the greatest change in the level of difficulty of the test item,-4.58 logit. The Wright Map in [Figure 2](#) shows that all students in the field test activities were able to change their informal statistical reasoning. This can be seen from the change in the level of difficulty of the test items of the students' informal statistical reasoning ability. The average change in the difficulty level of the test items in the pre and post-test conditions was negative, that is,-3.09. The size of the difference in the difficulty level of the informal statistical reasoning test items, which is negative, indicates that in the post-test condition, the items become much easier to complete than in the pre-test condition, concluding that the intervention of the ethno-flipped classroom model has a positive impact on students in solving the informal statistical problems given.

***RQ4: Differences Between Students' Informal Statistical Reasoning Skills in Time 1 Condition (Pretest) and Time 2 Condition (Posttest)***

The results of data normality and homogeneity tests of informal statistical reasoning data show that the pretest and posttest data are homogeneous, but not normally distributed. Based on these results, non-parametric statistical tests using the Wilcoxon signed-rank test were performed to check for differences before and after the intervention. The average score increased 76.50 with a probability of less than 0.05 ( $0.00 < 0.05$ ). The results indicated that there was a significant difference between the pre-test and post-test scores of informal statistical reasoning skills.

### **3.2. Discussion**

The ethno-flipped classroom model found that all students experienced positive changes and were able to complete the informal statistical reasoning skill items well. The data confirmed a positive change in the average logit value from the pre-test to the post-test condition. Student scores in the post-test condition also obtained a positive average logit value. The magnitude of the positive change in logit value indicates that all students were able to improve their informal statistical reasoning skills well and positively after receiving the ethno-flipped classroom model intervention.

Changes in students' informal statistical reasoning skills can also be seen from the analysis of changes in the level of difficulty of test items in the pre-test and post-test conditions, as shown in [Figure 2](#). The analysis of changes in the level of difficulty of the test items was performed using the Racking Analysis technique. Based on the results of the analysis of changes in the level of difficulty of the test items, it can be concluded that students experience positive changes in informal statistical reasoning skills. Positive changes can be seen in the change in students' logit scores, which increased from the pre-test condition (Time 1) to the post-test condition (Time 2). The pattern of changes that occur in students'

logit scores can be seen in [Figure 1](#), where there are changes in the positions of students' abilities presented on the Wright Map. The position of students' informal statistical reasoning ability in the post-test condition (Time 2) changed and increased to a higher logit value position (towards a positive logit value) than the position of students' ability in the pre-test condition (Time 1). The change in students' increasing logit value was also reinforced by the change in the level of difficulty of the test items of students' informal statistical reasoning ability, as shown in [Figure 2](#). The nature of the change in the level of difficulty of the test items shown in [Figure 2](#) indicates that the students' informal statistical reasoning ability has changed. The change in the level of difficulty of the test items also shows that the ethno-flipped classroom model has a positive impact on students in solving informal statistical problems. The ethno-flipped classroom model intervention changed the item difficulty level from the pre-test to the post-test condition. The impact of the ethno-flipped classroom model gives students confidence to be able to solve informal statistical problems better in the post-test condition, thereby influencing both the development of item difficulty levels and the mastery of informal statistical reasoning indicators.

In addition to score improvements and item difficulty shifts, further insight can be drawn from the Wright Map and Person-Item Distribution, which also reveal. In line with these findings, the consistent increase in students' logit scores and the shift in item difficulty levels suggest that the ethno-flipped classroom model effectively facilitates deeper engagement with statistical concepts embedded in culturally relevant contexts. By incorporating ethnomathematical elements into learning activities, students were exposed to familiar yet meaningful real-life statistical problems that promoted intuitive understanding and improved reasoning. This approach aligns with constructivist principles, where students build new knowledge through active experiences, especially when those experiences are rooted in their cultural backgrounds (Ramadhani et al., [2023a](#)). The increase in student ability logit values also reflects that students were not only recalling procedures but were genuinely developing informal statistical reasoning skills through contextual interpretation, argumentation, and data based decision making. This supports the integration of culturally responsive pedagogy in 21st-century mathematics instruction.

In addition to score improvements and item difficulty shifts, the Wright Map and Person-Item Distribution also reveal meaningful cognitive progress. Moreover, the change in the difficulty levels of test items, as visualized in [Figure 2](#), supports the idea that the learning environment created by the ethno-flipped classroom model did not merely prepare students to perform well on a fixed set of problems (Andrade & Coutinho, [2016](#)). Instead, it nurtured transferable reasoning skills that could be applied across varying problem complexities. The changes in item difficulty distribution indicate that students' post-test abilities were no longer hindered by the cognitive demands of the tasks. This implies an increased confidence and competence in addressing tasks involving variation, distribution, and contextual interpretation, core components of informal statistical reasoning. Therefore, the model's impact is not only visible in score improvement but also in the shift of cognitive engagement and learning depth, emphasizing its relevance for culturally responsive and meaningful mathematics education.

Beyond the quantitative evidence, the Rasch analysis using Wright Map and Person-Item Distribution provided evidence of alignment between students' increasing abilities and the hierarchy of item difficulties (Andrich, 2017; Bond et al., 2020). This alignment suggests that the ethno-flipped classroom model did not merely enhance surface-level understanding, but also supported the progressive development of students' cognitive strategies for solving informal statistical tasks. Students began demonstrating the ability to compare distributions, reason with informal data trends, and justify conclusions using contextual references drawn from ethnomathematical problems. These are hallmarks of informal statistical reasoning, which require interpretation beyond computation. The fact that more students occupied higher logit positions in the post-test reinforces the claim that this model fostered not only knowledge retention but also reasoning sophistication in culturally grounded contexts. Such findings reinforce the transformative potential of integrating pedagogy and culture in statistical education.

Complementing the quantitative data, qualitative observations during in-class activities revealed that students were more engaged and collaborative when solving statistical problems based on their cultural environments. Through group discussions, contextual reflections, and peer teaching during class sessions, students exhibited increased autonomy and confidence in interpreting statistical data. This transformation suggests that the ethno-flipped classroom promotes both cognitive and affective gains, reinforcing students' sense of identity while also equipping them with critical reasoning skills. Such outcomes are especially important in multicultural classrooms, where the recognition of students' cultural assets can transform learning environments into spaces that are inclusive, responsive, and academically rigorous. Ultimately, the study demonstrates that combining flipped learning with ethnomathematical contexts is not merely a pedagogical variation, but a transformative approach that bridges culture and cognition in mathematics education.

#### 4. CONCLUSION

The results of this study show that changes in the informal statistical reasoning skills of students taught using the ethno-flipped classroom model provide positive changes with stacking analysis and the racking-rasch model. Stacking analysis offers insight into individual-level changes in ability by identifying whether a person's performance has improved, declined, or remained consistent over time, based on the assumption that item difficulty remains stable across measurement occasions. Racking analysis completes the change analysis process but in terms of item difficulty. Stacking and racking analyses within the Rasch model offer a viable approach for measuring change, serving as a complementary method to traditional assessment techniques. Measuring changes at the individual level and item difficulty allows researchers to identify competent and problematic students for intervention. Overall, the stacking and racking analyses indicate that the learning intervention implemented through the ethno-flipped classroom model effectively enhanced the informal statistical reasoning skills of senior high school students. The ethno-flipped classroom model intervention had an impact on positive changes in students' informal statistical reasoning skills in the pre-test (Time 1) and post-test (Time 2) conditions.

Despite the promising findings, this study has several limitations. First, the cultural context applied in developing informal statistical reasoning assessments was limited to the Nias cultural setting. This specificity may limit the generalizability of the findings to other cultural contexts within ethnomathematics-based learning. Second, the sample size was reduced from 235 to 152 participants due to incomplete participation in the full learning cycle, to maintain data validity in the Stacking and Racking analysis. This reduced sample may affect the representativeness of the results. Third, the primary focus of this study was limited to analyzing changes in students' informal statistical reasoning abilities through Stacking and Racking analysis, without further statistical testing or qualitative triangulation to validate or deepen the interpretation of these changes. Future research is encouraged to expand the cultural contexts involved, increase sample diversity, and integrate mixed-method approaches, combining quantitative analysis with qualitative data triangulation to strengthen the understanding of changes in informal statistical reasoning skills over time.

## **Acknowledgments**

The authors sincerely extend our gratitude to Universitas Potensi Utama, Universitas Negeri Medan, Universitas Pendidikan Indonesia, the Research Center for Education at the National Research and Innovation Agency (BRIN), and Johannes Kepler University for their invaluable contributions and support to this collaborative research. We also express our profound appreciation to the editors and reviewers for their insightful and constructive feedback, which has greatly enhanced the quality of this manuscript. This work was supported by Johannes Kepler University Open Access Publishing Fund and the Federal State Upper Austria.

## **Declarations**

Author Contribution	: RR: Conceptualization, Visualization, Writing - Original Draft, and Writing – Review & Editing; S: Formal Analysis, Methodology, and Writing – Review & Editing; FA: Formal Analysis, Methodology, and Writing – Review & Editing; ZL: Formal Analysis, Validation, and Supervision; EE: Validation, and Supervision; ES: Validation, and Supervision.
Funding Statement	: This research was supported by Johannes Kepler University Open Access Publishing Fund and the Federal State Upper Austria.
Conflict of Interest	: The authors declare no conflict of interest.
Additional Information	: Additional information is available for this paper.

## REFERENCES

- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Amin, N. F. M., Saud, M. S., Buntat, Y., & Latif, A. A. (2013). The application of rasch measurement model in measuring change in students self-directed learning readiness. *Advanced Science Letters*, 19(12), 3736–3738. <https://doi.org/10.1166/asl.2013.5242>
- Andrade, M., & Coutinho, C. (2016). Implementing flipped classroom in blended learning environments: A proposal based on the cognitive flexibility theory. In *E-learn: World conference on e-learning in corporate, government, healthcare, and higher education*, (pp. 1115–1125). <https://www.learntechlib.org/p/174050>
- Andrich, D. (2017). Advances in social measurement: A rasch measurement theory. In F. Guillemin, A. Leplege, S. Briancon, E. Spitz, & J. Coste (Eds.), *Perceived health and adaptation in chronic disease* (pp. 66–91). Routledge. <https://doi.org/10.1201/9781315155074-7>
- Attard, C., & Holmes, K. (2020). An exploration of teacher and student perceptions of blended learning in four secondary mathematics classrooms. *Mathematics Education Research Journal*, 34(4), 719–740. <https://doi.org/10.1007/s13394-020-00359-2>
- Bond, T., & Fox, C. M. (2015). Why Measurement Is Fundamental. In *Applying the rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315814698>
- Bond, T., Yan, Z., & Heene, M. (2020). *Applying the rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. <https://doi.org/10.4324/9780429030499>
- Combrinck, C., Scherman, V., & Maree, D. (2016). Evaluating anchor items and reframing assessment results through a practical application of the rasch measurement model. *South African Journal of Psychology*, 47(3), 316–329. <https://doi.org/10.1177/0081246316683569>
- Conway, B., Gary Martin, W., Strutchens, M., Kraska, M., & Huang, H. (2019). The statistical reasoning learning environment: A comparison of students' statistical reasoning ability. *Journal of Statistics Education*, 27(3), 171–187. <https://doi.org/10.1080/10691898.2019.1647008>
- D'Ambrosio, U., & de Campos Almeida, M. (2017). Ethnomathematics and the emergence of mathematics. In J. W. Adams, P. Barmby, & A. Mesoudi (Eds.), *The nature and development of mathematics* (pp. 69–85). Routledge. <https://doi.org/10.4324/9781315648163>
- Department for Education. (2014). *The national curriculum in England: Key stages 3 and 4 framework document*. Department for Education of United Kingdom. <https://education-uk.org/documents/pdfs/2014-nc-framework-ks3-4.pdf>
- Ekinci, F., Bektaş, O., Karaca, M., & Yiğit, K. N. (2023). The flipped learning perception scale: A validity and reliability study. *Education and Information Technologies*, 28(11), 14139–14166. <https://doi.org/10.1007/s10639-023-11731-7>

- Fielding, J., Makar, K., & Ben-Zvi, D. (2025). Developing students' reasoning with data and data-ing. *ZDM – Mathematics Education*, 57(1), 1–18. <https://doi.org/10.1007/s11858-025-01671-6>
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. American Statistical Association. [https://www.amstat.org/asa/files/pdfs/gaise/gaiseprek-12\\_full.pdf](https://www.amstat.org/asa/files/pdfs/gaise/gaiseprek-12_full.pdf)
- Furuto, L. (2016). Lessons learned: Strengths-based approaches to mathematics education in the Pacific. *Journal of Mathematics and Culture*, 10(2), 55–72.
- Furuto, L. (2018). Knowledge and Action for Change Through Culture, Community and Curriculum. In *the 13th International Congress on Mathematical Education (ICME-13 Monographs)*, Cham (pp. 103–114). [https://doi.org/10.1007/978-3-319-72170-5\\_7](https://doi.org/10.1007/978-3-319-72170-5_7)
- Ganesan, N., & Leong, K. E. (2020). Impact of fathom on statistical reasoning among upper secondary students. *Journal of Research in Science, Mathematics and Technology Education*, 3(2), 35–50. <https://doi.org/10.31756/jrsmte.321>
- Hasanah, U., Putrawangsa, S., Setiawati, F. A., & Purwanta, E. (2024). Statistical literacy in primary education: An analysis of Indonesian fifth-graders' data interpretation and analysis skills. *Journal on Mathematics Education*, 15(4), 1335–1356. <https://doi.org/10.22342/jme.v15i4.pp1335-1356>
- Hidetoshi, F., & Rothman, T. (2021). *Sacred mathematics: Japanese temple geometry*. Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691127453/sacred-mathematics>
- Hjelte, A., Schindler, M., & Nilsson, P. (2020). Kinds of mathematical reasoning addressed in empirical research in mathematics education: A systematic review. *Education Sciences*, 10(10), 289. <https://doi.org/10.3390/educsci10100289>
- Hu, X., Jiang, Y., & Bi, H. (2022). Measuring science self-efficacy with a focus on the perceived competence dimension: using mixed methods to develop an instrument and explore changes through cross-sectional and longitudinal analyses in high school. *International journal of STEM education*, 9(1), 47. <https://doi.org/10.1186/s40594-022-00363-x>
- Ilie, S., Forbes, K., Curran, S., & Vermunt, J. D. (2024). Higher education students' conceptions of learning gain. *Active Learning in Higher Education*, 1–17. <https://doi.org/10.1177/14697874241270461>
- Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing children's statistical thinking. *Mathematical thinking and learning*, 2(4), 269–307. [https://doi.org/10.1207/s15327833mtl0204\\_3](https://doi.org/10.1207/s15327833mtl0204_3)
- Kim, H., & Chae, D.-h. (2016). The development and application of a STEAM program based on traditional Korean culture. *Eurasia Journal of Mathematics, Science and Technology Education*, 12(7), 1925–1936. <https://doi.org/10.12973/eurasia.2016.1539a>
- Kollosche, D. (2021). Styles of reasoning for mathematics education. *Educational Studies in Mathematics*, 107(3), 471–486. <https://doi.org/10.1007/s10649-021-10046-z>

- Laliyo, L. A. R., Sumintono, B., & Panigoro, C. (2022). Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in rasch model. *Heliyon*, 8(3), e09126. <https://doi.org/10.1016/j.heliyon.2022.e09126>
- Ling, M.-T., Pang, V., & Ompok, C. C. (2018). Measuring change in early mathematics ability of children who learn using games: Stacked analysis in rasch measurement. In *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings*, Singapore (pp. 215–226). [https://doi.org/10.1007/978-981-10-8138-5\\_17](https://doi.org/10.1007/978-981-10-8138-5_17)
- Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 47(1), 121–150. <https://doi.org/10.3102/00346543047001121>
- Lipka, J., Hogan, M. P., Webster, J. P., Yanez, E., Adams, B., Clark, S., & Lacy, D. (2005). Math in a cultural context: Two case studies of a successful culturally based math project. *Anthropology & Education Quarterly*, 36(4), 367–385.
- 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>
- Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In *Teaching statistics in school mathematics-challenges for teaching and teacher education-New ICMI study series*, Dordrecht (pp. 347–358). [https://doi.org/10.1007/978-94-007-1131-0\\_33](https://doi.org/10.1007/978-94-007-1131-0_33)
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105. <https://doi.org/10.52041/serj.v8i1.457>
- Ministry of Education Culture Research and Technology. (2023). *Keputusan kepala badan standar, kurikulum, dan asesmen pendidikan kementerian pendidikan, kebudayaan, riset, dan teknologi nomor 008/H/KR/2022 tentang capaian pembelajaran pada pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah pada kurikulum merdeka* [Decree of the Head of the Education Standards, Curriculum, and Assessment Agency of the Ministry of Education, Culture, Research, and Technology Number 008/H/KR/2022 concerning learning outcomes in early childhood education, elementary education, and secondary education in the independent curriculum]. Kementerian Pendidikan Kebudayaan Riset dan Teknologi.
- Ministry of Education of People's Republic of China. (2003). *Ordinary senior secondary: Mathematics curriculum standards*. Ministry of Education of Republic of China.
- Ministry of Education Singapore. (2023). *Mathematics syllabuses secondary: One to four express course normal (academic) course*. Curriculum Planning and Development Division.
- Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical thinking and learning*, 4(1), 23–63. [https://doi.org/10.1207/s15327833mtl0401\\_2](https://doi.org/10.1207/s15327833mtl0401_2)
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. The National Council of Teachers of Mathematics, Inc.

- Pentecost, T. C., & Barbera, J. (2013). Measuring learning gains in chemical education: A comparison of two methods. *Journal of Chemical Education*, 90(7), 839–845. <https://doi.org/10.1021/ed400018v>
- Pfannkuch, M. (2006). Informal inferential reasoning. In *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*. [https://www.stat.auckland.ac.nz/~iase/publications/17/6A2\\_PFAN.pdf](https://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf)
- 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>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Prahmana, R. C. I., Yunianto, W., Rosa, M., & Orey, D. C. (2021). Ethnomathematics: Pranamatangsa 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>
- Ramadhani, R., Prahmana, R. C. I., Soeharto, S., & Saleh, A. (2024). Integrating traditional food and technology in statistical learning: A learning trajectory. *Journal on Mathematics Education*, 15(4), 1277–1310. <https://doi.org/10.22342/jme.v15i4.pp1277-1310>
- Ramadhani, R., Saragih, S., Maulida, R., & Simanjuntak, S. M. B. (2022). Measuring changes of students' statistical reasoning taught by ethnomathematics approach assisted TinkerPlots: A stacking analysis study. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, 6(3). <https://doi.org/10.31764/jtam.v6i3.8375>
- Ramadhani, R., Saragih, S., & Napitupulu, E. E. (2022). Exploration of students' statistical reasoning ability in the context of ethnomathematics: A study of the rasch model. *Mathematics Teaching Research Journal*, 14(1), 138–168.
- Ramadhani, R., Syahputra, E., & Simamora, E. (2022). Ethno-flipped classroom model: Sebuah rekomendasi model pembelajaran matematika di masa new normal [Ethno-flipped classroom model: A Recommendation for new normal mathematics learning model]. *AXIOM : Jurnal Pendidikan dan Matematika*, 10(2). <https://doi.org/10.30821/axiom.v10i2.10331>
- Ramadhani, R., Syahputra, E., & Simamora, E. (2023a). Ethnomathematics approach integrated flipped classroom model: Culturally contextualized meaningful learning and flexibility. *Jurnal Elemen*, 9(2), 371–387. <https://doi.org/10.29408/jel.v9i2.7871>
- Ramadhani, R., Syahputra, E., & Simamora, E. (2023b). *Model ethno-flipped classroom: Solusi pembelajaran fleksibel dan bermakna* [Ethno-flipped classroom model: A flexible and meaningful learning solution]. Indonesia Emas Group.
- Ramadhani, R., Syahputra, E., & Simamora, E. (2025). Enhancing students' informal statistical reasoning skills through the development of a worksheet rooted in the Nias culture. *Mathematics Teaching-Research Journal*, 17(2), 104–128.

- Rosa, M., & Orey, D. C. (2016). State of the art in ethnomathematics. In *Current and future perspectives of ethnomathematics as a program* (pp. 11–37). Springer International Publishing. [https://doi.org/10.1007/978-3-319-30120-4\\_3](https://doi.org/10.1007/978-3-319-30120-4_3)
- Sariningsih, R., & Herdiman, I. (2017). Mengembangkan kemampuan penalaran statistik dan berpikir kreatif matematis mahasiswa di Kota Cimahi melalui pendekatan open-ended [Developing statistical reasoning and mathematical creative thinking skills of students in Cimahi City through an open-ended approach]. *Jurnal Riset Pendidikan Matematika*, 4(2), 239–246. <https://doi.org/10.21831/jrpm.v4i2.16685>
- Schleicher, A. (2023). *PISA 2022: Insight and Interpretations*. OECD.
- Setianingsih, R., & Rahmah, D. A. (2020). Statistical reasoning levels of high school students in solving statistics related problems. *Journal of Physics: Conference Series*, 1569(4), 042064. <https://doi.org/10.1088/1742-6596/1569/4/042064>
- Smit, R., Dober, H., Hess, K., Bachmann, P., & Birri, T. (2023). Supporting primary students' mathematical reasoning practice: the effects of formative feedback and the mediating role of self-efficacy. *Research in Mathematics Education*, 25(3), 277–300. <https://doi.org/10.1080/14794802.2022.2062780>
- Soeharto, S., & Csapó, B. (2022). Assessing Indonesian student inductive reasoning: Rasch analysis. *Thinking Skills and Creativity*, 46, 101132. <https://doi.org/10.1016/j.tsc.2022.101132>
- Soeharto, S., Martono, M., Hairida, H., Akhmetova, A., Arifiyanti, F., Benő, C., & Charalambos, C. (2024). The metacognitive awareness of reading strategy among pre-service primary teachers and the possibility of rating improvement using Rasch analysis. *Studies in Educational Evaluation*, 80, 101319. <https://doi.org/10.1016/j.stueduc.2023.101319>
- Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch analysis to evaluate the psychometric properties of junior metacognitive awareness inventory in the Indonesian context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. <https://doi.org/10.15294/jpii.v10i4.27114>
- Sukoriyanto, S., & Khurin'in, K. (2023). Statistical reasoning of junior high school students in solving central tendency problems mean, median, and modus. *AIP Conference Proceedings*, 2569(1), 040020. <https://doi.org/10.1063/5.0112343>
- Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. In *1st International Conference on Education Innovation (ICEI 2017)*, (pp. 38–42). <https://doi.org/10.2991/icei-17.2018.11>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan* [Application of rasch modeling to educational assessment]. Trim Komunikata.
- Sunjaya, D. K., Herawati, D. M. D., Indraswari, N., Megawati, G., & Sumintono, B. (2021). Training and assessing model for the ability of community health volunteers in anthropometric measurement using the rasch stacking and racking analyses. *Journal of Environmental and Public Health*, 2021, 1–11. <https://doi.org/10.1155/2021/5515712>

- Uzun, Z., & Öğretmen, T. (2021). Test equating with the rasch model to compare pre-test and post-test measurements. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 12(4), 336–347. <https://doi.org/10.21031/epod.957614>
- Valenta, A., Rø, K., & Klock, S. I. (2024). A framework for reasoning in school mathematics: analyzing the development of mathematical claims. *Educational Studies in Mathematics*, 116(1), 91–111. <https://doi.org/10.1007/s10649-024-10309-5>
- Willoughby, S. D., & Metz, A. (2009). Exploring gender differences with different gain calculations in astronomy and biology. *American Journal of Physics*, 77(7), 651–657. <https://doi.org/10.1119/1.3133087>
- Wright, B. D. (2003). Rack and stack: time 1 vs. time 2. *Rasch measurement transactions*, 17(1), 905–906.