Survey Attitude Toward The Level Of Understanding Of Descriptive Statistics by Students in Higher Education: *Pair Comparison Methode*

Nusrotus Sa'idah[1]*, Jailani[2] , Sudiyatno[3]   iD

[1] Universitas Islam Nahdlatul Ulama Jepara, Indonesia
[2,3] Yogyakarta State University, Indonesia

A R T I C L E   I N F O

A B S T R A K

Banyak orang menganggap statistik itu sangat sulit. Hal ini berdampak pada siswa tidak hanya pada prestasi akademisnya tetapi juga pada pengaruh tidak menyenangkan atau persepsi negatif. Penelitian ini bertujuan menghasilkan penskalaan instrumen sikap terhadap tingkat statistika mahasiswa dengan metode pair comparison dengan karakteristik yang valis dan reliabel. Penelitian ini merupakan penelitian deskriptif kuantitatif dengan mengembangkan instrumen penelitian dengan skala pair comparison. Subjek penelitian ini adalah mahasiswa yang telah mendapatkan mata kuliah sejumlah 160 mahasiswa. Instrumen ini terdiri dari 11 dimensi yang mengukur sikap mahasiswa terhadap tingkat pemahaman statistika. Analisis menggunakan deskriptif kuantitatif. Hasil penskalaan instrumen sikap mahasiswa terhadap tingkat pemahaman statistika dengan metode pair comparison menunjukkan karakteristik setiap dimensi valid dan nilai Cronbach dengan kriteria reliabilitas tinggi. Hasil penskalaan instrumen sikap terhadap materi statistika dengan rata-rata urutan respon tertinggi adalah pada modus pemahaman materi," dengan nilai 0.695. Nilai reliabilitas yang tinggi ini dapat menyatakan rentang kemampuan yang lebih luas, artinya lebih heterogen. Dengan penskalaan pair comparison method ini dapat menangkap heterogenitas substansial dalam hirarki nilai pribadi seseorang dan lebih mudah responden dalam menjawab karena membandingkan dua objek.

A B S T R A C T

Many people find statistics very difficult. This impacts students not only on their academic performance but also on unpleasant influences or negative perceptions. This research aims to produce a scaling instrument for attitudes toward student statistics levels using the pair comparison method with valid and reliable characteristics. This research is a quantitative descriptive research by developing a research instrument with a pair comparison scale. The subjects of this research were 160 students who had received courses. This instrument consists of 11 dimensions that measure students' attitudes toward their level of understanding of statistics. The analysis uses quantitative description. The results of scaling the student attitude instrument towards the level of understanding of statistics using the pair comparison method show that the characteristics of each dimension are valid and the Cronbach value meets high-reliability criteria. The results of scaling the attitude instrument towards statistical material with the highest average response order were in the material understanding mode," with a value of 0.695. This high-reliability value can indicate a wider range of abilities, meaning more heterogeneous. With this pair comparison method scaling, it can capture substantial heterogeneity in a person's personal value hierarchy, and it is easier for respondents to answer because they compare two objects.

1. INTRODUCTION

In higher education, almost all disciplines require students to take statistics courses or some aspects of statistics. There have been several previous studies revealing factors that affect student performance in statistics courses. attitude towards statistics as one of the keys (Abbiati et al., 2021; Akour, 2022; Crettaz von Roten & de Roten, 2023; Xu, C., & Schau, 2019). Colleges prepare students to be able to interpret and critically evaluate data in statistical learning (Büscher, 2022; Crettaz von Roten & de Roten, 2023). This basic understanding of statistics is to prepare students for data literacy, or understanding their own research and evaluating the research that has been done. Statistics have an important role in learning, especially now that the new independent learning curriculum explains that learning is integrated with research. Therefore, statistics are very important to learn, especially for students. Many consider statistics to be very difficult; this has an impact on students not only on their academic performance but also on their unpleasant influence or negative perception of the application of statistical thinking or the

*Corresponding author.
E-mail addresses: nusrotussaidah@unisnu.ac.id (Nusrotus Sa'idah)

implications of statistics in everyday life (Akour, 2022; Sutter et al., 2022). Despite being a compulsory course at the university, students consider it a burden, underestimate its usefulness, and face difficulties when doing assignments, potentially leading to failure in learning statistics (Abbiati et al., 2021; Crettaz von Roten & de Roten, 2023). Statistical reasoning is needed by students or the wider community to be able to communicate with data appropriately and selectively because all policies are related to data talk, variations, and opportunities in this modern era (Büscher, 2022; Moore, D. S., McCabe, G. P., & Craig, 1997). Statistics have an important role in learning, especially now that the new independent learning curriculum explains that learning is integrated with research. Therefore, statistics are very important to learn, especially for students. Many consider statistics to be very difficult; this has an impact on students not only on their academic performance but also on their unpleasant influence or negative perception of the application of statistical thinking or the implications of statistics in everyday life (Akour, 2022; Sutter et al., 2022). Measurement of student attitudes towards statistics needs to be detected as a reference for student responses to learning that is considered difficult.

Measurement of student attitudes towards statistics already exists, namely the survey of attitudes towards statistics instrument (SATS-36) (Schau, 2003), where this instrument has six factors, namely effect, cognitive, value, difficulty, effort, and interest. This instrument has been widely used because it has proven its validity and reliability. The SATS-36 instrument has high internal consistency of each subscale with Confirmatory Factor Analysis (CFA) analysis (Ayebo et al., 2019; Nolan, M. M., Beran, T., & Hecker, 2012), while the use of SATS-36 in the Estonian context with CFA analysis does not support the six-factor model but with EFA supports a four-factor scale structure, where the three factors of affective, cognitive competence, and difficulty are combined into one (Hommik, C., & Luik, 2017). Research stated that the six factors in the survey of attitudes toward statistics (SATS-36) instrument have a good model fit, Cronbach's alpha, and reliability estimates showing good values (Shahirah & Moi, 2019; Xu, C., & Schau, 2019). Another study on student attitudes towards statistics in sports study programs is based on a gender perspective that investigates attitudes towards statistics courses, which shows that men have more positive attitudes towards statistics than women and have higher confidence (Crettaz von Roten & de Roten, 2023; Robert F. DeVellis, 2019). Measurement of non-cognitive attitudes is very important and should be researched in an effort to determine the attitude or response of students to learning. In various previous research studies, the disclosure of statistical attitudes was done using the Likert scale and rating scale. It is also different in this study, which describes students' attitudes towards statistics by disclosing their understanding of the material that has been studied. This non-cognitive attitude needs to be studied to find out the positive and negative responses of students.

Measurement of non-cognitive aspects by giving scores or numbers to the attributes to be measured (Setiawati et al., 2013). Non-cognitive psychological attributes are also known as personality and affective attributes. Attitude measurement is very important to find out the factors that contribute to learned abilities (Jones et al., 2022). The characteristics of psychological scales are stimulus responses, where the measured attributes reveal behavioral indicators (Saifuddin Azwar, 2020), These measurement instruments must fulfill quality criteria as well as disclose individual judgments (Jones et al., 2022). There are several levels of measurement, namely ordinal, ratio, interval, and nominal scales (Lord, F.M. & Novick, 1974; Setiawati et al., 2013). The quality of the instrument is evidenced in measuring certain factors by knowing its validity and reliability. Validity and reliability are proven to be good if, by increasing the sample size, further statistical measurements, different scale designs, etc., the validity and reliability of the instrument are proven to be good (Gradvohl et al., 2023). Therefore, the scaling process in developing an instrument greatly affects the characteristics of the items. Many studies use instruments with rating scales and Likert scales. There are still very limited research instruments using the thrustone scale.

The Thurstone scale is synonymous with involving experts or judgment to rate statement items and asking to put the favorability of the statement on a psychological number that can predict the outcome being measured (Guffey et al., 2007; Halik et al., 2022) and also assess not for approval (Syaifuddin Azwar, 2012). This shows the advantages of using the thrustone scale, namely that the involvement of panelists avoids subjectivity, authenticity, bias, and accuracy (Saifuddin Azwar, 2020). In addition, the thrustone scale has the advantage of discouraging social desirability (dishonest responses) and falsifying self-characterization (McDonald, 2013). The Thrustone scale also provides space for respondents to provide opinions about the statement items. There are three types of thrustone scaling techniques: 1) the pair comparison method; 2) the equal-appearing interval method; and 3) the excessive interval method (Bahar et al., 2021). Scaling in Thurstone to estimate preferences among objects based on the observed frequency of their pairwise comparisons (Lipovetsky, 2007). In planning the preparation of the instrument as a step to anticipate items that fall, consider the number of items. Thrustone scaling is done with the principle of converting the raw scores obtained from z scores that use a normal distribution. Therefore, the distance between scores is used with the same unit so that interval data can be

obtained. Thrustone scaling is widely used in the development of social science instruments, especially psychology, one of which measures student attitudes towards statistics. The pair comparison method is one of the thrustone-type scaling methods with a simple comparison and multi-attribute ranking process by comparing two of the object choices. (Pritikin, 2020; Ullah et al., 2020). The pair comparison method is suitable for measuring the abilities of heterogeneous individuals (Iijima et al., 2020), favorable due to the simplicity of the choice-of-object construct and incorporating comparative judgments (Kula Kartal & Gülleroğlu, 2015; Yurtcu & Dogan, 2015). The statistical ability of students in this study is heterogeneous; therefore, the hypothesis of measuring student attitudes towards understanding statistical solutions is appropriate if measured in response with the Pair comparison method.

Several previous studies have developed research with the Thrustone scale to measure police career success in Canada (Guffey et al., 2007), another with (Pratama, 2020) also developed the thrustone scale with the equal-appearing interval method to measure the attitude of religious moderation. Other research also developed a thrust-scale instrument on teachers' ability to utilize information technology (Arfandi et al., 2020). The results of the study show that the thrustone scale can be applied to various constructs or topics of discussion in compiling instruments. Therefore, in this study, attitudes were measured using a thrustone scale with the pair comparison method.

From the reality of previous research, measuring attitudes needs to be developed because, as an expression of individual judgment, the instruments used also need to have good quality and appropriate scales to avoid subjectivity and authenticity bias. Measuring statistical attitudes needs an appropriate scale to determine student responses and understanding of statistical material. In this study, measuring attitudes is more about how students understand the stages of descriptive statistical analysis. The ability of students in statistical literacy is quite high, but they are still poor at visualizing data (Jatisunda et al., 2020; Setiawan & Sukoco, 2021). Students find it difficult when reporting descriptive analysis data or with data interpretation due to a lack of prior knowledge of statistics (Maryati, 2017). This research aims to produce a scaling instrument for attitudes toward student statistics levels using the pair comparison method with valid and reliable characteristics.

2. METHODS

This research is quantitative descriptive research with a research design that is: 1) instrument development; 2) Scaling process on trial data; 3) analysis of psychometric characteristics. This research design begins with developing an instrument for student understanding attitudes towards statistical material to determine validity and reliability properly. The scaling process uses the pair comparison method scale, the scale for multi-attribute ranking process, and the multilevel order technique (Ullah et al., 2020). This method of pairwise comparison may be easier for participants to respond to (Pritikin, 2020). For example, comparing chocolate pudding with different flavors Each question involves two competing recipes (two objects) (Ullah et al., 2020). The subjects of this study were students who had received statistics courses at the Islamic University of Nahdlatul Ulama Jepara, totaling 160 students. These students are subjects who have received descriptive statistics courses, namely semester 4 of the Elementary School Teacher Education and Islamic Religious Education study programs.
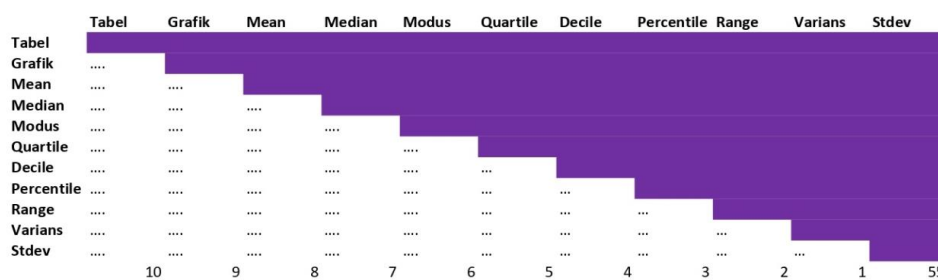

	Tabel	Grafik	Mean	Median	Modus	Quartile	Decile	Percentile	Range	Varians	Stdev
Tabel											
Grafik	....										
Mean	....	....									
Median	....	....	....								
Modus	....	....	....	....							
Quartile	....	....	....	....	....						
Decile	....	....	....	....	....	...					
Percentile	....	....	....	....	....	...	...				
Range	....	....	....	....	....	...	...	...			
Varians	....	....	....	....	....	...	...	...	...		
Stdev	....	....	....	....	....	...	...	...	...	...	
	10	9	8	7	6	5	4	3	2	1	55

Figure 1. Respons Pattern of a Statistical Attitude Instrument with a Pair Comparison Method

This instrument development research is to measure the level of understanding or ease with which student complete descriptive statistical analysis calculations. This instrument consists of 11 dimensions that measure student attitudes towards the level of understanding of statistics by comparing pairs. The question in this instrument is, "From descriptive statistical material, which do you think you understand more or is easier to solve?" This instrument is designed to identify attitudes toward understanding statistical material that has been learned. Each item has two options; choose one response

that matches the level of understanding or ease of learning of each subject. The 11 dimensions were paired into 55 questions. The pattern of subject selection in filling out the questionnaire showed in Figure 1. The data analysis used in this study used descriptive statistical analysis with Excel and instrument quality analysis with Jamovi software. Analysis of the characteristics of the instrument, calculating reliability with Cronbach alpha with an acceptable reliability estimate standard with a value of 0.80-0.90 (Crocker, L., & Algina, 2008), while SEM uses the equation $\sigma E = \sigma x \sqrt{1 - Pxx}$ (Retnawati, 2016). The higher the reliability, the smaller the measurement error. In addition, descriptive analysis to determine the group of materials that are considered easy to understand by students in statistical learning.

3. RESULT AND DISCUSSION

Results

Measurement in the realm of statistical attitudes has been done using the survey of attitudes toward statistics (SATS-36), but this study is different in measuring statistical attitudes, namely by student responses assessing their own understanding of statistical material. The items of this instrument were developed using the Thrustone type, which refers to the pair comparison method. Each item is paired with another item, and one of the appropriate items is selected. In this study, one item in one dimension was paired with another item in another dimension. This instrument has 11 dimensions; therefore, there are 55 paired items. Data analysis for the pair comparison scaling method was conducted using Microsoft Excel and Jamovi. The stages of analysis in pair comparison scaling are as follows: 1) make a frequency matrix of the subject's choice of each stimulus; 2) sort the stimulus from the smallest to the largest; 3) convert the frequency value into a proportion value (p) = f / N; 4) convert the price of p into a z score, which is the average deviation from the normal distribution; 5) calculate the average z score; 6) order the stimulus from the lowest to the highest. The questionnaire on students' attitudes towards the level of understanding of statistics was distributed in G-Form; the questionnaire was filled in by students who had received statistics courses, especially descriptive statistics. The following data analysis results are the original data that have been converted from frequency values to proportions. In addition, the data was converted into a z score (mean deviation or normal distribution). From the results of the instrument validity analysis showing the correlation matrix between dimensions, namely a p value < .001, each dimension is declared valid. Table 1 shows that the highest correlation value is between the "quartile-variance" dimension, which is 0.983, while the lowest correlation value is between the "Percentile-modus" dimension, with a value of 0.787. The correlation of the quartile-variance dimension is high because there is a relationship between the quartile and variance completion stages. In the development of this instrument, it shows the results of a high reliability analysis with a reliability value of> 0.6, namely 0.987. The calculation of the reliability value between dimensions also shows a very high value. The following are the results of the Cronbach's α analysis in each dimension: The reliability value based on internal consistency is related to the standard deviation of the variance of the measurement results. The variance of the measurement score greatly affects the size of the instrument reliability value (Mardapi, 2008). Formula SEM **or σe = σt $\sqrt{1-}$ ρxx'** can predict the magnitude of reliability if it is assumed that the SEM and the variance of the two groups are equal. Reliability for Each Dimension showed in Table 1.

Table 1. Reliability for Each Dimension

	Data Asli		Data Transformasi z	
	Cronbach	**SEM**	**Cronbach**	**SEM**
Tabel	0.986	0.049	0.971	0.196
Grafik	0.986	0.053	0.971	0.226
Mean	0.986	**0.038**	0.973	0.148
Median	0.986	0.049	0.972	0.191
Modus	0.987	0.034	0.976	0.120
Quartile	0.986	0.059	0.971	0.253
Decile	0.985	0.043	0.977	0.248
Percentile	0.986	0.037	0.977	0.233
Range	0.985	0.039	0.970	0.199
Varians	0.985	0.039	0.971	0.218
Stdev	0.986	0.030	0.972	0.189

Table 1 shows that the Cronbach's α reliability value for data that has been transformed into z scores has decreased from the original data. For the measurement of SEM, or standard error of measurement, there is an increase from the original data to the data that has been transformed into z

scores for all dimensions that measure statistical attitudes. The results of the data from the ordinal level to the interval do not increase the reliability value but increase the SEM value in each dimension. The next data analysis is a descriptive statistical analysis of each dimension to explain the results of pairwise scaling on the instrument of understanding attitudes towards statistical material presented in the mean, standard deviation, median, minimum, and maximum values. With the results of these descriptive statistics, it can be categorized by the attitude of students toward choosing paired dimensions that are easier to learn statistics. The results of descriptive data analysis for the pair comparison method instrument. showed in Table 2.

Table 2. Description of instrument data using pair comparison

Aspect	Original Data					Data after z transformation				
	Mean	**Stdev**	**Med**	**Min**	**Max**	**Mean**	**Stdev**	**Med**	**Min**	**Max**
Tabel	0.604	0.161	0.689	0.324	0.779	0.209	0.651	0.536	-0.966	0.904
Grafik	0.521	0.177	0.531	0.255	0.752	0.182	0.748	0.236	-0.929	1.167
Mean	0.658	0.125	0.607	0.489	0.848	0.239	0.492	0.000	-0.528	1.014
Median	0.635	0.165	0.669	0.324	0.827	0.212	0.634	0.314	-1,055	0.943
Modus	0.695	0.111	0.703	0.500	0.862	0.252	0.399	0.230	-0.595	0.908
Quartile	0.417	0.192	0.421	0.193	0.710	0.149	0.84	0.000	-0.789	1.495
Decile	0.317	0.143	0.317	0.131	0.524	0.155	0.824	0.000	-0.887	1.595
Percentile	0.313	0.124	0.289	0.179	0.500	0.174	0.774	0.000	-0.630	1.499
Range	0.493	0.129	0.500	0.289	0.689	0.206	0.662	0.073	-0.801	1.227
Varians	0.414	0.130	0.427	0.255	0.621	0.19	0.723	0.138	-0.644	1.430
Stdev	0.407	0.100	0.379	0.283	0.559	0.214	0.627	0.02	-0.517	1.276

Table 2. shows the mean, standard deviation, median, minimum and maximum results. The dimension for the subject matter "mode" gets the highest average value of 0.695 and the lowest order is the dimension that discusses "percentile" which is 0.313. This shows that the student's response in answering the instrument states that the material that is most easily understood is the "mode" material and the material that has not been understood is "percentile". The results of the original data analysis were transformed into z scores, the simplest approach to determine whether a person's score is in the output range with a z score (Kornak et al., 2019). Changes in data trends after being transformed into z scores can be seen in the mean and standard deviation values increasing for each dimension, but the median value decreases when the data is transformed into z scores. Students find it easier to work on the stages when completing mode data analysis because the stages in solving mode are very simple, both for solving single and group data. Students complete mode calculations more easily because the stages are very simple compared to working on data presentation, either tables or graphs (Masfingatin & Suprapto, 2020; Mulya, N., Nurlaelah, E., & Prabawanto, 2018). Descriptive statistics have been studied since elementary, middle, and high school; even at the undergraduate level, students are used to calculating descriptive statistics (Setiawan & Sukoco, 2021). The comparison of several dimensions that measure the statistical attitude of students using bloxpots The results show that students prefer the subject matter of the mode because the maximum value exceeds 800k but is not far from the subject matter mean," which is also close to 800k. It can also be seen that the average value from Table 1 shows that each dimension of the subject matter of the mode and median is almost the same, namely 0.695 and 0.658. The lowest dimension that students consider less understanding than others is the dimension on the subject matter of "percentile," with a value close to 600k or 0.313. Compared to the dimension of the subject matter of the mode, which is easier to step than the percentile, this finding makes it clear that the statistical formula to solve the subject matter of "mode" is easier than "percentile".

Discussion

This study analyzes a survey of student attitudes towards the level of understanding of descriptive statistics. This non-cognitive attitude is very important to examine because attitude determines the atmosphere of learning (Cahyawati et al., 2020). The results showed that the characteristics of the instrument using the pair comparison method showed that each dimension was valid with a p value < .001, and the reliability value of the 11 dimensions in pairs had a high reliability value with a reliability value > 0.6, namely 0.985–0.987. The size of the variance greatly affects the size of the reliability value and SEM (Bahar et al., 2021; Setiawati et al., 2013). With this pairwise comparison method, it shows high quality because students are easier to answer. This is similar to research shows that the paired comparison method can capture heterogeneity compared to the rating scale, which may be affected by response style bias (Iijima et al., 2020). High reliability results indicate a wider range of abilities, meaning

more heterogeneity (Ebel, R. L. & Frisbie, 1986), Conditions of heterogeneity lead to an overestimation of instrument reliability (Syaifuddin Azwar, 2012). A person's personal values have a hierarchical structure that is organized according to the level of importance. One value may be more important to one person than another, but it may be less important to another person. This pair comparison can capture the substantial heterogeneity in a person's personal value hierarchy (Iijima et al., 2020). The measuring process is an expression on an object that has a "yes" or "no" feature with a number symbol. This expression is very helpful in increasing the sensitivity to object measurement (Yurtcu & Dogan, 2015). This numerical expression is used to evaluate whether different responses in the same category form the same feature or not. The pair comparison method is a scaling process used for multi-attribute ranking, with a very simple way of comparing other objects to evaluate which object is better between two choices (Pritikin, 2020; Ullah et al., 2020).

Many studies have revealed attitudes using the Survey of Attitudes Towards Statistics (SATS-36). On a rating scale, there are six factors: effect, cognitive, value, difficulty, effort, and interest (Abbiati et al., 2021; Akour, 2022; Birbal et al., 2018; Crettaz von Roten & de Roten, 2023). This attitude refers to different interrelated dispositions that favor a response or not and is related to statistical learning. In contrast to these studies, this study measures the same attitude but on the response of students' ability to statistical learning, which is compared between statistical sub-materials using the pair comparison method attitude scale instrument. The response shows that students' understanding of the material that has been learned is in the material mode," with the highest average value of 0.695. This material is indeed very easy to learn because the steps in the formula are easily understood by students. It is not just interpreting data in the form of graphs and tables but know the mean and standard deviation of the data so as to analyze and interpret it appropriately (Emilia & Amir, 2022; Lukman et al., 2022; Masjudin et al., 2020). This is also made clear in the research stated that the initial level students in the classification of answers to the task of students calculating descriptive statistics other than the mean and median can answer correctly (Setiawan & Sukoco, 2021). Attitude towards understanding statistical material as a determinant for measuring student learning success (Kurniawan et al., 2019; Liaghatdar et al., 2011). Students can get satisfactory results. This attitude survey is very supportive of research that includes a person's emotions or feelings. A person's responses in psychology can affect human thinking and behavior (Musfira, 2022; Pudaruth et al., 2018). There are several cases of research where students in statistical analysis of data are always wrong in their analysis; they are still confused with the data obtained in accordance with the analysis carried out (Nurlita, 2018). The results of this study reveal that knowledge ability needs to be expressed through a survey of comparative responses to the level of ease of statistical data analysis from statistical materials taught by lecturers. Therefore, the results of scaling attitudes with the Pair comparison method in this study are very suitable for measuring comparisons in detecting the ability to complete the stages of heterogeneous statistical formulas.

4. CONCLUSION

The results of scaling the attitude instrument towards the statistical level of students with the pair comparison method show the characteristics of each dimension are valid and the reliability value of the 11 dimensions in pairs has a high reliability. The Cronbach's α reliability value for data that has been transformed into a z score has decreased from the original data. For the measurement of SEM, or standard error of measurement, there is an increase from the original data to the data that has been transformed into z scores for all dimensions that measure statistical attitudes. The results of scaling the instrument of attitude towards statistical material with the highest average response order are on the understanding of the material mode. This material is indeed very easy to learn because the steps in the formula are easy for students to understand. This high reliability value can state a wider range of abilities, meaning that it is more heterogeneous. The pair comparison method in this study can capture substantial heterogeneity in a person's personal value hierarchy, considering the ability of students to understand statistics is very diverse.

5. REFERENCES

Abbiati, N. N., Fabrizio, M. D. C., López, M. V., Pérez, A., Plencovich, M. C., & Cueto, G. (2021). Attitudes Related To Students' Performance in Statistics in University Programs in Argentina. *Statistics Education Research Journal*, *20*(2). https://doi.org/10.52041/SERJ.V20I2.356.

Akour, M. M. (2022). Rasch rating scale analysis of the survey of attitudes toward statistics. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(12). https://doi.org/10.29333/ejmste/12646.

Arfandi, A., Purnamawati, & Nurfaedah. (2020). The development of a Thurstone scale for identifying teacher ability in using information and communication technology. *Journal of Physics: Conference Series*, *1456*(1). https://doi.org/10.1088/1742-6596/1456/1/012038.

Ayebo, A., Bright, J., & Ballam, C. (2019). Examining the Factor Structure of the Survey of Attitudes Towards Statistics among Undergraduate Health Science Students. *International Electronic Journal of Mathematics Education*, *15*(1), 1–8. https://doi.org/10.29333/iejme/5942.

Azwar, Saifuddin. (2020). *Penyusunan Skala Psikologi Edisi 2*. Pustaka Pelajar.

Azwar, Syaifuddin. (2012). *Tes Prestasi*. Pustaka Pelajar.

Bahar, R., Setiawati, F. A., Sutarji, A., Hidayat, O., & Sudarna, N. (2021). *Measurement in Educational Research Comparison of the Thurstone Scale method in measuring teacher personality competence*. *1*(2), 97–103.

Birbal, D. R., Ramdass, D. M., & Harripaul, M. C. (2018). Student Teachers' Attitudes towards Blended Learning. *Journal of Education and Human Development*, *7*(2), 9–26. https://doi.org/10.15640/jehd.v7n2a2.

Büscher, C. (2022). Design Principles for Developing Statistical Literacy in Middle Schools. *Statistics Education Research Journal*, *21*(1), 1–16. https://doi.org/10.52041/serj.v21i1.80.

Cahyawati, D., Wahyudin, & Prabawanto, S. (2020). Students' statistical reasoning reviewed from academic internal factors in two learning classes. *Journal of Physics: Conference Series*, *1480*(1). https://doi.org/10.1088/1742-6596/1480/1/012051.

Crettaz von Roten, F., & de Roten, Y. (2023). Undergraduate sport sciences students' attitudes towards statistics: A gender perspective. *Journal of Hospitality, Leisure, Sport and Tourism Education*, *33*(May), 100452. https://doi.org/10.1016/j.jhlste.2023.100452.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning Pu.

Ebel, R. L. & Frisbie, D. A. (1986). *Essentials of educational measurement.* Prentice Hall. Inc.

Emilia, N. R., & Amir, M. F. (2022). *Treffinger Learning to Enhance Statistical Literacy Primary Student*. *11*(3), 522–532.

Gradvohl, E., Lukács, Á. J., Takács, J., Fritúz, G., Falus, A., & Feith, H. J. (2023). Development and validation of the questionnaire on resuscitation-related knowledge and attitude for adolescents. *Evaluation and Program Planning*, *100*(September 2021), 102338. https://doi.org/10.1016/j.evalprogplan.2023.102338.

Guffey, J. E., Larson, J. G., Zimmerman, L., & Shook, B. (2007). The Development of a Thurstone Scale for Identifying Desirable Police Officer Traits. *Journal of Police and Criminal Psychology*, *22*(1), 1–9. https://doi.org/10.1007/s11896-007-9001-8.

Halik, A., Khumairo, A., & Yawisah, U. (2022). The Instrument Construction for Students' Religious Moderation: Psychometric Analysis. *ENLIGHTEN (Jurnal Bimbingan dan Konseling Islam)*, *5*(2), 80–94. https://doi.org/10.32505/enlighten.v5i2.4895.

Hommik, C., & Luik, P. (2017). Adapting the survey of attitudes towards statistics (SATS-36) for Estonian secondary school students. *Statistics Education Research Journal*, *16*(1), 228-239. https://doi.org/https://doi.org/ 10.52041/serj.v16i1.229.

Iijima, Y., Okumura, Y., Yamasaki, S., Ando, S., Okada, K., Koike, S., Endo, K., Morimoto, Y., Williams, A., Murai, T., Tanaka, S. C., Hiraiwa-Hasegawa, M., Kasai, K., & Nishida, A. (2020). Assessing the hierarchy of personal values among adolescents: A comparison of rating scale and paired comparison methods. *Journal of Adolescence*, *80*(February), 53–59. https://doi.org/10.1016/j.adolescence.2020.02.003.

Jatisunda, M. G., Nahdi, D. S., & Suciawati, V. (2020). Literasi Statistika Mahasiswa Adminitrasi Publik. *SJME (Supremum Journal of Mathematics Education)*, *4*(2), 134–146. https://doi.org/10.35706/sjme.v4i2.3488.

Jones, M. G., Chesnutt, K., Ennes, M., Macher, D., & Paechter, M. (2022). Measuring science capital, science attitudes, and science experiences in elementary and middle school students. *Studies in Educational Evaluation*, *74*(April), 101180. https://doi.org/10.1016/j.stueduc.2022.101180.

Kornak, J., Fields, J., Kremers, W., Farmer, S., Heuer, H. W., Forsberg, L., Brushaber, D., Rindels, A., Dodge, H., Weintraub, S., Besser, L., Appleby, B., Bordelon, Y., Bove, J., Brannelly, P., Caso, C., Coppola, G., Dever, R., Dheel, C., … Rosen, H. (2019). Nonlinear Z-score modeling for improved detection of cognitive abnormality. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *11*, 797–808. https://doi.org/10.1016/j.dadm.2019.08.003.

Kula Kartal, S., & Gülleroğlu, H. D. (2015). Ölçekleme Yaklaşımlarının Madde ve Birey Sıralamaları Açısından Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *6*(1), 25–37. https://doi.org/10.21031/epod.53309.

Kurniawan, D. A., Astalini, A., Darmaji, D., & Melsayanti, R. (2019). Students' attitude towards natural sciences. *International Journal of Evaluation and Research in Education*, *8*(3), 455–460. https://doi.org/10.11591/ijere.v8i3.16395.

Liaghatdar, M. J., Soltani, A., & Abedi, A. (2011). A Validity Study of Attitudes toward Science Scale among Iranian Secondary School Students. *International Education Studies*, *4*(4). https://doi.org/10.5539/ies.v4n4p36.

Lipovetsky, S. (2007). Thurstone scaling in order statistics. *Mathematical and Computer Modelling*, *45*(7–8), 917–926. https://doi.org/10.1016/j.mcm.2006.09.009.

Lord, F.M. & Novick, M. R. (1974). *Statistical theories of mental test scores.* Addison-Wesley.

Lukman, L., Wahyudin, W., Suryadi, D., Dasari, D., & Prabawanto, S. (2022). Studying Student Statistical Literacy in Statistics Lectures on Higher Education Using Grounded Theory Approach. *Infinity Journal*, *11*(1), 163. https://doi.org/10.22460/infinity.v11i1.p163-176.

Mardapi, djemari. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Mitra Cendekia Press.

Maryati, I. (2017). Analisis Kesulitan Dalam Materi Statistika Ditinjau Dari Kemampuan Penalaran Dan Komunikasi Statistis. *Prisma*, *6*(2), 173–179. https://doi.org/10.35194/jp.v6i2.209.

Masfingatin, T., & Suprapto, E. (2020). Student's Statistical Literacy skills Based on the Reflective and Impulsive Cognitive Styles. *Al-Jabar : Jurnal Pendidikan Matematika*, *11*(2), 273–286. https://doi.org/10.24042/ajpm.v11i2.6902.

Masjudin, M., Muzaki, A., Abidin, Z., & Ariyanti, I. A. P. (2020). Analysis of student's statistical thinking ability in understanding the statistical data. *Journal of Physics: Conference Series*, *1521*(3). https://doi.org/10.1088/1742-6596/1521/3/032063.

McDonald, R. P. (2013). *Test theory: A unified treatment*. Pyschology Press.

Moore, D. S., McCabe, G. P., & Craig, B. A. (1997). *New pedagogy and new content: The case of statistics International statistical review*.

Mulya, N., Nurlaelah, E., & Prabawanto, S. (2018). Students ' statistical literacy on junior high school. *In International Conference on Mathematics and Science Education of universitas Pendidikan Indonesia*, *3*, 710–714.

Musfira, A. F. (2022). *Sentiment Analysis of Tweets Regarding Sri Lankan Crisis using Automatic Coding in ATLAS . ti 22 Presented by South Eastern University of Sri Lanka*. August.

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103–123. https://doi.org/https://doi.org/10.52041/serj.v11i2.333.

Nurlita, M. (2018). *Analisis Kemampuan Pemahaman Pengetahuan Mahasiswa*. *4*(November), 46–56. https://doi.org/10.17605/OSF.IO/JPZ2F.

Pratama, D. (2020). Pengembangan Skala Thurstone Metode Equal Appearing Interval untuk Mengukur Sikap Moderasi Beragama Siswa Sekolah Menengah Atas. *Jurnal Psikologi Teori dan Terapan*, *11*(1), 71. https://doi.org/10.26740/jptt.v11n1.p71-82.

Pritikin, J. N. (2020). An exploratory factor model for ordinal paired comparison indicators. *Heliyon*, *6*(9), e04821. https://doi.org/10.1016/j.heliyon.2020.e04821.

Pudaruth, S., Moheeputh, S., Permessur, N., & Chamroo, A. (2018). Sentiment Analysis from Facebook Comments using Automatic Coding in NVivo 11. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *7*(1), 41–48. https://doi.org/10.14201/adcaij2018714148.

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (panduan peneliti, Mahsiswa dan Psikometri)*. Parama Publishing.

Robert F. DeVellis. (2019). Frank D. Brown Hall Room 3208. *To Achieve Excellence by Guiding Individuals as They Become Professionals*", 1–7.

Schau, C. (2003). Students' attitudes: The "other" important outcome in statistics education. *Proceedings of the Joint Statistical Meetings*, 3673–3681.

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.

Setiawati, F. A., Mardapi, D., & Azwar, S. (2013). Penskalaan Teori Klasik Instrumen Multiple Intelligences Tipe Thurstone Dan Likert. *Jurnal Penelitian dan Evaluasi Pendidikan*, *17*(2), 259–274. https://doi.org/10.21831/pep.v17i2.1699.

Shahirah, S., & Moi, N. (2019). Investigating the Validity and Reliability of Survey Attitude towards Statistics Instrument among Rural Secondary School Students. *International Journal of Educational Methodology*, *5*(4), 651–661. https://doi.org/10.12973/ijem.5.4.651.

Sutter, C. C., Givvin, K. B., Tucker, M. C., Givvin, K. A., Leandro-Ramos, A., & Solomon, P. L. (2022). Student Concerns and Perceived Challenges in Introductory Statistics, How the Frequency Shifted during

COVID-19, and How They Differ by Subgroups of Students. *Journal of Statistics and Data Science Education*, 1–13. https://doi.org/10.1080/26939169.2022.2132325.

Ullah, K., Aslam, M., & Sindhu, T. N. (2020). Bayesian analysis of the Weibull paired comparison model using informative prior. *Alexandria Engineering Journal*, *59*(4), 2371–2378. https://doi.org/10.1016/j.aej.2020.02.032.

Xu, C., & Schau, C. (2019). Exploring method effects in the six-factor structure of the survey of attitudes toward statistics (SATS-36). *Statistics Education Research Journal, 18*(2), 39-53. https://doi.org/https://doi.org/10. 52041/serj.v18i2.139.

Yurtcu, M., & Dogan, N. (2015). Scaling with Paired Comparison Method for Reasons for Mathematics Anxiety of Secondary School Students. *Procedia - Social and Behavioral Sciences*, *197*(February), 851–857. https://doi.org/10.1016/j.sbspro.2015.07.209.