An Analysis of Statistical Reasoning Process of High School Students in Solving the Statistical Problem

Rosidah¹, I Ketut Budayasa², and Dwi Juniati³

¹Department of Mathematics and Science, State University of Makassar, Post Graduate Student of State University of Surabaya, Indonesia
²Department of Mathematics and Science, State University of Surabaya, Indonesia
³Department of Mathematics and Science, State University of Surabaya, Indonesia

¹rosidah@unm.ac.id, ²ketutbudayasa@unesa.ac.id, ³dwijuniati@unesa.ac.id

Abstract. Reasoning and problem-solving are interrelated competencies and are two important competencies in school math learning including statistics. Both competencies are targets in the school's mathematics curriculum. The purpose of this study is to investigate the statistical reasoning of high school students in solving statistical problems i.e. the size of data center and standard deviation. The subjects of this study are students with high and low mathematical abilities obtained from mathematics ability test results given to 450 students of class XI from three different schools in Makassar, Indonesia. This research is a descriptive research with the qualitative approach. The results showed that in solving the standard deviation problem students with high and low mathematical ability to know the algorithm well and procedural can finish the calculation well but do not know the concept of deviation standard so cannot interpret the results obtained. It is also found that students with high and low mathematical ability to solve problems well procedural on the material of data centering, but cannot interpret the context in context, especially on median materials. In short, the two students had poor conceptions and poor reasoning in descriptive statistics especially in determining median and standard deviations. Other findings of students with high mathematical ability may represent in the form of graphs, diagrams, histograms and bar charts, whereas students with low mathematical abilities can only represent data in the form of bar charts, graphs, and pie charts.

1. Introduction
In the National Council of Teachers of Mathematics (NCTM), 2000 stated that one of the basic mathematical competencies that must be had by students in learning mathematics in addition to problem-solving is reasoning. This is in line with the demands in the 2013 curriculum that the ability to reason is one of the mathematical skills which is the goal of learning mathematics that must be had by students in Indonesia after the learning of mathematics. This means that reasoning is substantial to be taught and cultivated to students and especially in the field of mathematical and statistical education research in the face of the challenges of globalization and the demands of the 21st century. In addition, Litner [1] argues that reasoning ability is essential to understanding mathematics and making mathematics meaningful. Therefore, reasoning can be trained from an early age to students by familiarizing students to provide arguments or logical reasons in solving mathematical problems and statistics, then according to Mueller et al, [2] the notion of understanding mathematics is meaningless if without emphasize reasoning. But the facts in the field show that many students have difficulties in learning mathematics because students learn by memorization without deep understanding and many
of them are less challenged to learn and solve mathematical problems, especially non-routine problems or problems that result in low student achievement in Indonesia. This fact is supported by the survey results from the Trends in International Mathematics and Science Study (TIMSS) survey in 2011, the achievement of Indonesian students in the competition on aspects of knowledge, applying, and reasoning aspects by 37%, 23%, and 17% [3]. From the data shows that the lowest average percentage achieved by Indonesian students is on the cognitive domain at the reasoning level of only 17%. The low achievements are worthy of attention, therefore in this article will discuss how the reasoning by high school students in statistics as one of the mathematics material.

At the school level in Indonesia, statistics is not a separate field of study, but are part of the mathematics lesson required to be followed by every student at every level of education, from elementary to higher education and even in the university level. Statistics have been established as an important part of the school's mathematics curriculum in various countries including Indonesia, and are widely used in all fields of science and even statistics especially the science of opportunities used by some countries in marketing its products e.g. cars and various other electronic products.

Mathematics education’s experts have recognized that the basis of statistical reasoning should be trained and grown since primary education [4]. These competencies need to be trained and developed so that students can have skills in managing information, interpreting data and critically evaluating and making decisions in utilizing information or data that exist in everyday life. To respond this, it required a science related to how processing information or data and analyzing data. According Garfield and Ben-Zvi [5] science related to the way of processing is statistics.

Therefore in this article will describe how the statistical reasoning of high school students in solving statistical problems namely the average, median, mode and standard deviation. Research on statistical reasoning has been studied by some experts mathematics education and statistics abroad, but in Indonesia research on statistical reasoning especially in high school students in terms of differences in students' math skills have not been studied. The statistical reasoning that has been studied by Chang and Ismail Z [6] is using the Jones et al [7] framework of describing data, representing data, organizing and reducing virgin, and analyzing and interpreting the data and data the students processed is the data presented. While in this article to assess the students' statistical reasoning used collecting data stage, presenting data, processing data, analyzing and interpreting data and data processed students is the data obtained by students in the school environment.

**Statistical Reasoning**

Statistical reasoning is defined as the way one reasoned with statistical ideas and understands statistical information. It includes making interpretations based on a set of data, data representation or statistical data summary [8]. While Chance & Garfield (2000: 101) states that statistical reasoning is defined as the way people reason with statistical information and make sense of it. It involves making interpretations based on data sets, or statistical summaries of data where students need to combine ideas about data and the chance to make an inference and interpret statistical results. In line with this, Olani et al [9] states that statistical reasoning refers to one's ability to understand and integrate statistical concepts and ideas to interpret data and to draw conclusions based on statistical contexts.

In identifying students 'reasoning abilities Garfield [5] specifies examples of statistical materials that can develop students' statistical reasoning abilities among others, reasoning of data that is to recognize and categorize qualitative or quantitative, discrete or continuous data and appropriate data types to be displayed on charts and tables, it is reasoning about the size of the statistics that is understanding the size of the data center, the size of the data location and the size of the spread of data further know why the recapitulation for predicting a data for large sample sizes is more accurate than the small sample size.

2. **Method**

The subjects of this study were students with high and low mathematics ability who have obtained the subject of statistical learning. The procedure of determining the subject: first, the researcher gave a
mathematics ability test to 450 high school students of class XI from three different schools in the form of mathematical problems that have been validated by mathematics education experts. The test results obtained are grouped according to predetermined criteria. In addition, the researcher was helped by the mathematics teachers at the school to recruit students with low, medium, and high maths skills. To obtain and explore the students' reasoning process each student is given the task of solving statistical problems i.e. data that can be obtained within his own school environment. Furthermore, interviews and observations are conducted to explore the depth the reasons why to take such conclusions and also ask the students to reveal what is thought and tell more detail the process of thinking. Interview session of a problem-solving task is recorded using a recording device, the result of the data then transcribed, the same thing was done on the second problem-solving the task and so on until obtained valid and consistent data. This research paper aims to investigate and analyze how the process of statistical reasoning of students in solving the problem of descriptive statistics is the measure of data centralization and standard deviation. Therefore, researcher used descriptive explorative and qualitative approaches

3. Results and Discussion

Based on the results of job-based interview transcripts conducted by the researcher to students after in-depth interviews to explore the process of statistical reasoning of students with high and low mathematical ability in solving problems are obtained as follows: The first stage of high students ability, is to read carefully the given problem carefully, after that the subject is given time to collect data related problems. The second stage after the students get the data then continued with the question how to collect data that has been obtained, the students responds that the data obtained by way of the table in the respondent to be filled related data about the height and weight of respondents, in addition, the students with high mathematical ability also states to obtain data related to problems could be made by directly measuring respondents using a measuring tool but they did not do with the complicated reasons and requiring a long enough time for measuring one person, they also represent to supply student data can do with authorized in the inquiries or can be done with requests existing student data at school. This means they understand the techniques of collecting data related to the problem by interview, using questionnaires or viewing school documents. While for subjects with low mathematical ability when they are given problems, they also read the given problem carefully. The next stage after acquired the data, they were asked questions related to how to collect data obtained. They stated that the data is obtained with a table member of the respondent, or it can be done by looking at the student's biography in the school. This means that students with the low mathematical ability only know the technique of data collection that is questionnaire or document.

Furthermore, the response of both subjects i.e. when they (students with high and low mathematical ability) asked to present data is high ability students’ states that this data needs to be presented in table form because the data is quite large and varied and to facilitate in the calculation process, for instance, the average. Besides, they represent the data in the form of bar charts, histograms, graphs and circuit diagrams. Before making a table thing done they are observing a set of random data that has been obtained by itself, then the next step students determine the largest and smallest data from a set of data and then sort the data followed by making a frequency table. When determining the average data center size, the median and high ability students’ modes prefer using the finish by using grouped data with the table reason already presented so that it can be used in determining average, median and mode, however when prompted to determine the data centering size by using a single data the student can also calculate. This means that high ability students know the algorithm and skillfully uses the formula in determining the data centering size. While the low ability students’ response when asked to present the data, the sequence data with the reason easier to find the largest data and the smallest data and more easily in making the frequency distribution table. The next step is they create graphs, bar charts and circle diagrams, while other dishes are not made with the reason to forget how. At the stage of data processing that determines the average students choose to use the concept of single data centralization with the reason more easily calculate that is by summing all the data then divided by the
number of data. However, when asked to use another way, they can calculate using the concept of group data. To determine the mode, they prefer to use single data because it only sees the most data appear while by way of group data cannot be done by them with the reason forgotten the formula. When asked to determine median, they can do in two ways that are with data group or single data. This means that the students know the algorithm and skillfully use the formula in determining the mean, median except mode with the group data is not done by them for the reason of forgetting the formula.

At the stage of analyzing and interpreting, students with high and low mathematical ability have the wrong reasoning on median materials. Likewise, when they provide an explanation of the representation made both have misconceptions in interpreting the centralized measure, especially in the median. This is in line with the 2008 Cooper & Smith [10] opinion that students have a lot of misunderstandings on the median data concentration. Furthermore, in the standard deviation material, they perform well calculations but when interpreting the results obtained, they cannot relate the statistical values obtained with the concept of statistics itself, in addition both of them cannot explain the definition of standard deviation and have a conception that wrong in analyzing and interpreting the results obtained. Here are excerpts from interviews students with high and low mathematical ability:

| Student’s excerpts with high mathematical ability | Student’s excerpts with low mathematical ability |
|--------------------------------------------------|-----------------------------------------------|
| **R**: What do you mean by the mean?              | **R**: What do you mean by the mean?           |
| **S**: Mean is the sum total of all data then divided by the amount of data | **S**: Mean is the sum total of all data then divided by the amount of data |
| **R**: What is the meaning of mean?              | **R**: What is the meaning of mean?           |
| **S**: It means mostly or divided equally         | **S**: The mean is divided equally             |
| **R**: What do you mean by median?               | **R**: What do you mean by median?            |
| **S**: It is the middle number                    | **S**: It is the middle number                 |
| **R**: How do you determine the median?          | **R**: How do you determine the median?       |
| **S**: If using single data, firstly the data is sorted | **S**: If using single data, firstly the data is sorted |
| **R**: And then?                                 | **R**: And then?                              |
| **S**: If the amount of even data then the two data in the middle number then divided into two, while if odd, the most middle data is the median | **S**: If the amount of data is odd then the most middle data is median and if amount even data then add two data which is divided then half median |
| **R**: Can you conclude the median value you have got? | **R**: Can you conclude the median value you have got? |
| **S**: Median for single data of high students’ body data class XI IPA 6 is 163 cm | **S**: Median for single data of high student body data class XI IPA 6 is 163 cm |
| **R**: What do you think is standard deviation? | **R**: What do you think is standard deviation? |
| **S**: Hemm I do not know mam, I still confused what is standard deviation | **S**: Hemm I do not know mam. |
| **R**: How do you determine standard deviation? | **R**: How do you determine standard deviation? |
| **S**: Using formula mam                          | **S**: With formula                            |

4. Conclusion
The result of the research shows that the ability of high school students with high and low mathematics ability is proficiently procedural in determining the data centering and the data dissemination measure that is the standard deviation but weak in conceptual understanding especially in median material and standard deviation. This is because students tend to learn by just remembering
the formula but not trained in interpreting the statistical values obtained, thus students are not trained in understanding the concepts associated with the main statistical problems associated with contextual data. The stages of analyzing and interpreting the two subjects are not able to relate the statistical values obtained with the statistical concept itself so that both cannot interpret well the results obtained in accordance with the context of the data it obtains or the problem given. By that understanding of the concept of statistics and reasoning skills is very important to be taught and trained and cultivated to in accordance with the context of data. The next stage of analyzing and interpreting the two subjects is not able to relate the statistical variability of data represented via histograms and stem-and-leaf plots. The stages of analyzing and interpreting the two subjects are not able to relate the statistical variability of data represented via histograms and stem-and-leaf plots.

References

[1] Littner, Johan. 2005. A framework for analyzing qualities of mathematical reasoning: version 3. Research Reports in Mathematics Education 3. Umea, Sweden: Departement of Mathematics, Umea University

[2] Mueller et al (2010) Rules Without Reason: Allowing Students to Rethink Previous Conceptions. The Montana Mathematics Enthusiast, vol 7

[3] Mullis, I. V. S., Martin, Michael. O, Foy, Pierre & Arora, Alka. (2012). TIMSS 2011 International Result in Mathematics, Chestnut Hill, MA, USA TIMSS & PIRLS International Study Center, Boston College.

[4] National Council of Teachers of Mathematics. Principles and Standards for School Mathematics. Reston, Va: NCTM, 2000

[5] Garfield, J. (2002). The challenge of developing statistical reasoning. Journal of Statistics Education, 10 (3).

[6] Chan, S.W, & Zaleha Ismail. Developing Statistical Reasoning Assessment Instrument for High School Students in Descriptive Statistics. Procedia-Social and Behavioral Sciences, 116, 4338-4343, 2014

[7] Jones, G. A, Thornton, C. A, Langrall, C. W, Mooney, E. S, Perry, B, & Putt, I. A. (2000). A Framework for characterizing children's statistical thingking. Mathematical Thinking and Learning, 2, 269-307.

[8] Ben-Zvi, D. and Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions and challenges. In Ben-Zvi, D & Garfield, J (eds.), The challenge of developing statistical literacy, reasoning, and thinking (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic.

[9] Olani et al (2011) Statistical Reasoning Ability, Self Efficacy, And Value Beliefs in A Reform Based University Statistics Course. The Nederland institute For Educational Research, University Of Groningen

[10] Cooper, L.L. and Shore, F.S. (2008). Students’ misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. Journal of Statistics Education, 16(2).

[11] Burrill, G. and Camden, M. (2005). Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable. Voolburg, The Netherlands: International Statistical Institute.