MATHE*dunesa*
Jurnal Ilmiah Pendidikan Matematika *Volume 9 No. 2 Tahun 2020*
ISSN:2301-9085

STATISTICAL LITERACY PROFILE OF HIGH SCHOOL STUDENTS IN SOLVING STATISTICS PROBLEM BASED ON MATHEMATICAL ABILITIES

Safiroh Fauziyah
Pendidikan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
E-mail: safirohfauziyah16030174001@mhs.unesa.ac.id

Abstract

Being someone literate by statistics is very important for each individual to be able to interpret statistical messages in various contexts. This research is a descriptive-qualitative study to describe the statistical literacy profile of high school students in solving statistical problems based on mathematical ability. The data collection method uses written tests and interviews. The research subjects were one student with a high mathematical ability category and one student with a moderate mathematical ability category. The results showed that: (1) students with high mathematical abilities can understand data, interpret data, and evaluate data well. Students with a high mathematical ability to communicate data by presenting data in the form of diagrams or graphs, but without considering the characteristics of the data in the presentation; (2) while students with moderate mathematical ability can understand data well. The results of data interpretation of moderate mathematical students' ability experience errors in terms of determining averages and data distribution and communicating data by presenting data in the form of diagrams or graphs that experience misconceptions related to the presentation of bar charts and histograms.

Keywords: Profile, Literacy, Statistical Literacy.

INTRODUCTION

In the current digital era, we unwittingly have been using statistics even though in a simple form, at home, at school, in the office, or elsewhere. For example, statistics relating to opinion polls, health, education, and others. Sharma (2012) states that someone needs statistical knowledge to explain statistical messages to support a wider range of information and social content, such as politics, health, education, and related issues. Therefore, statistical understanding needs to be improved by starting at the elementary school level and continuing to be strengthened and developed during the years of junior high school (SMP) and senior high school (SMA). Statistics and opportunities are one of the materials that are in the scope of mathematics for secondary education and upper secondary education. In the K-13 curriculum mentioned basic competencies for high schools (SMA) students must master in statistical material that is determining and analyzing the size of data concentration and data distribution presented in the form of distribution tables, frequencies, and histograms, as well as solving problems related to the presentation of measurement and enumeration data in the distribution table and histogram. Based on OECD (2018) concept of innovative literacy is related to the capacity of students to apply knowledge and skills in core subjects and to analyze, reason, and communicate effectively as they pose, solve and interpret problems in a variety of situations. Based on the explanation above, it can be concluded that literacy is the ability of a person to identify, understand, interpret, evaluate and communicate effectively both verbally and in writing that is connected with various contexts. Regarding statistical literacy, Wallman (1993) in Gal, I (2000) argues that statistical literacy is not only important for society as a whole. but also important for the community individually (personally). Because people make decisions in their personal lives based on the information and risk analysis provided by others in the community. As an example of decision making related to residence, the type of work to be sought, or the type of vehicle to be purchased, this is influenced by data provided from outside one's personal experience. According to Ben-Zhi & Garfield (2008), statistical literacy involves understanding and using basic language and statistical tools: knowing basic statistical terms, understanding the use of simple statistical symbols, and being able to recognize and interpret various data representations. Meanwhile, according to Gal (2004) Statistical literacy is described as the ability to interpret, evaluate critically, and communicate statistical information and messages. Based on some of the explanations above regarding the definition of statistical literacy from experts, in this study what is used as a definition of statistical literacy is the ability of a person or individual to understand concepts, basic statistical

terms, and the use of simple statistical symbols, and interpret, evaluate and communicate statistical data or information through various media.

Statistical literacy provides the opportunity for the community not only to read data but to evaluate and communicate statistical messages (Francois and Van Bendegem, in Yolcu 2012). Being someone literate by statistics is very important for each individual to be able to interpret statistical messages in various contexts. The reality on the ground shows that most students have difficulty in studying statistics and statistical literacy. Ayuningtias (2015) states that many students can solve statistical problems but only a few can explain the meaning of the results obtained. Besides, students can present data in the form of tables and graphs but have not been able to understand and interpret them properly. Meletiou and Lee (2002) stated that students had difficulty in understanding and presenting data in the form of a histogram. Jones et al (2002) state that students are very weak in describing data, intuitively they lack knowledge in terms of interpreting and understanding the concentration and dissemination of data. Additionally, Schield (1998) found that students had difficulty in evaluating arguments involving statistics and provided alternative explanations for observed statistical relationships.

The ability of mathematics is needed in statistical literacy, for example when reading tables, graphs or diagrams requires special reading skills. For example, to explain the contents of a table, we need knowledge related to information arranged in rows and columns. Meanwhile, to explain the contents of diagrams or graphs, knowledge of scale and coordinates reading is needed, and also to calculate the average and percentage. Based on the problems outlined above about the importance of statistical literacy and mathematical abilities possessed by students, the researcher is interested in researching with the title "High School Student Statistics Literacy Profile in Resolving Statistical Problems Based on Mathematical Capability".

METHOD

This study aims to describe the statistical literacy profile of high school students in solving statistical problems based on mathematical ability. Based on the research objectives, this research is a type of descriptive research with a qualitative approach. The research instrument consisted of the main instruments and supporting instruments. The main instruments of this study are the researchers themselves and the supporting instruments of this study, as follows: mathematics ability test questions, statistical problem test questions, and interview guidelines. Data obtained from this study are statistical problem tests and interview results. The statistical problem test is carried out by the chosen subject and then interviews are conducted on the subject. The purpose of the interview is to explore information from subjects not yet written in the written test. There is three data analysis in this research, they are mathematical ability test data analysis, statistical problem test data analysis, and analysis of interview results. Data analysis of mathematical ability tests is done using students grouped into three categories based on the results of mathematics ability test scores. The grouping of these three categories refers to the rating scale according to Purwaningtyas (2018) as follows: high ability if $80 \leq scores \leq 100$, medium ability if $60 \leq scores < 80$, and low ability if $0 \leq scores < 60$. In this study, the category of students with high and moderate mathematical ability was chosen, because they were considered to have better problem solving and communication skills verbally and in writing. From the category of high and moderate mathematical ability, two students were selected with details of one student with a high mathematical ability category and one student with a moderate mathematical ability category. Analysis of statistical problem test data based on statistical literacy indicators.

Table 1. Statistical Literacy Indicators

Statistical Literacy Aspects	Statistical Literacy Indicators	Code
Understanding statistical data or information	Read data and identify what is known and asked from problems presented in tables, diagrams, or graphs.	A1
	Look for information by referring to explicit observations on the data that is seen on the graph of the problem presented.	A2
	Determine the use of variables, symbols, or graphics that correspond to the problem presented.	A3
Interpret statistical data or information	Apply concepts and determine the value of centralizing or disseminating data to resolve the problem presented.	B1
	Explain the data presented in the form of diagrams or graphs that have been used in solving problems.	B2

	Make conclusions based on information available on the problem presented.	B3
Evaluate statistical data or information	Check for possible statistical error information presented.	C
Communicating statistical data or information	Make a diagram or graph following the information you want to communicate.	D

Analysis of the results of the interviews was carried out through the stages of data reduction, data presentation, and concluding (Miles & Huberman, 1992).

RESULT AND DISCUSSION

The statistical problem used in this study is as follows.

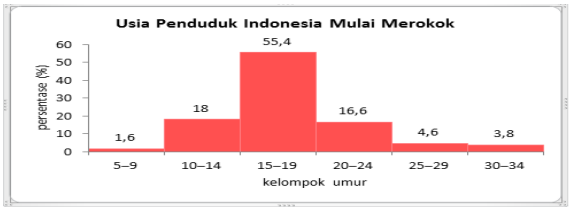
1. Pay attention to data on the growth of internet users from year to year based on the results of the 2017 survey.

Year	Growth of Internet Users (Million)
2006	20
2007	20
2008	25
2009	30
2010	42
2011	55
2012	63
2013	82
2014	88,1
2015	110,2
2016	132,7
2017	143,26

From the internet user growth data:

a. Serve as charts or graphs accordingly! Give the reason!
b. How big is the highest growth increase of internet users? What year happened?
c. Determine the average increase in internet users from 2006-2017!
d. Determine the mode of increase of internet users from 2006-2017!
e. Determine variants and standard deviations of an increase in internet users from 2006-2017!
f. In what year did the growth of internet users not increase?
g. Explain the information you get from the data!
h. What can be concluded from the data that has been presented?

2. Smoking habits in Indonesia are very high in various walks of life ranging from children, adolescents to adults. The following is data on the age of the Indonesian population who started smoking based on the 2014 Global Youth Tobacco Survey.



based on data above:

a. At what age is the highest percentage of Indonesians starting to smoke?
b. At what age is the lowest percentage of Indonesians starting to smoke?
c. Explain the information you got from the graph!

Based on the analysis of statistical problem test data and the results of interviews of subjects with high and moderate mathematical ability, obtained a description of the statistical literacy of high school students in solving statistical problems based on mathematical ability, as follows.

1. Description of high school students' statistical literacy in solving statistical problems with high levels of mathematical ability (S1).

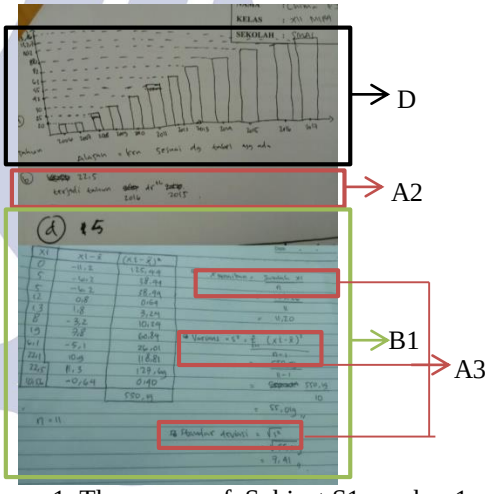


Figure 1. The answer of Subject S1 number 1a-1e

Based on statistical problem tests and interviews on item number one, subject S1 reads carefully the questions from beginning to end, then read questions to identify what was known and asked from the problem of growing internet users and the Indonesian population began smoking. In other words, subject S1 can read the data and identify what is known and asked from the problem presented in the form of tables and graphs.

On problem number 1.a aims to achieve the indicators make diagrams or graphs according to information to be communicated. Subject S1 creates a bar chart by drawing *the* x-axis and y-axis and then determine for the y-axis the number of internet users while for the x-axis is years. in

this activity, subject S1 prefers to use bar charts rather than line charts because she thinks that the use of the two diagrams is the same.

Based on data analysis and interview results, subject S1 presents data from a tabular form in the form of bar charts accompanied by clear arguments, but here subject S1 is still unable to distinguish the use of each diagram (bar and line diagrams) or without considering the characteristics of the data in data presentation. Because internet user growth data includes continuous data so it is more suitable to be presented using line charts. This is consistent with the results of research by Hafiyusholeh (2018) that students can build alternative views for a collection of data in the form of diagrams or graphs that refer to the purpose of communicating a data, without regard to the characteristics of the data. Subject S1 can only group data based on a single or group frequency distribution. In other words, Subject S1 can make diagrams or graphs following the information to be communicated but has not been able to determine the use of diagrams or graphs following the data presented.

In problem number 1.b aims to find a certain value from a graph. Subject S1 answered that the highest increase in internet user growth occurred in 2015-2016. This activity is carried out by Subject S1 by looking for differences in the number of growth of internet users each year.

Based on data analysis and interview results, subject S1 explains how to obtain data on the increase in internet user growth presented in the form of bar charts and determine the highest increase in internet user growth. In other words, subject S1 can explain the data presented in the form of diagrams that have been used in solving problems and can find information by referring to the visible data.

In problem numbers 1.c, 1.d and 1.e aim to achieve indicators applying the concept and determining the concentration value and data distribution and the use of variables and symbols. First, the subject S1 determines what is required to find the value of the mean, mode, variance, and standard deviation of the increase in internet user growth. After that, Subject S1 calculates the mean by adding up the data increase in internet users from 2006-2017 then divided by many data increases in internet users. To find the value of the mode, Subject S1 looks at the value that often appears from the increase internet growth data that is 5, so the mode value is 5. For the variance, Subject S1 adds up the square value of each data increase in internet growth minus the average increase in internet growth then divided by the amount of data minus one. Whereas the standard deviation is the square root of variance.

Based on data analysis and interview results, Subject S1 determines the mean, mode, variance, and standard deviation properly and correctly. Subject S1 can explain the definition of mean, mode, and each variable or symbol used to solve problems related to mean, mode, variance, and standard deviation appropriately. In other words, Subject S1 can apply the concept and determine the value of concentration and data distribution to solve the problem presented and be able to determine the use of variables or symbols following the problem presented.

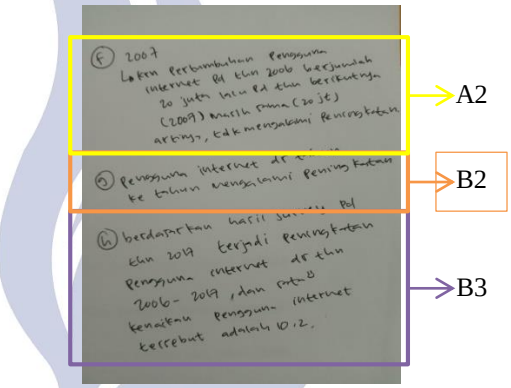


Figure 2. The answer of Subject S1 number 1f-1h

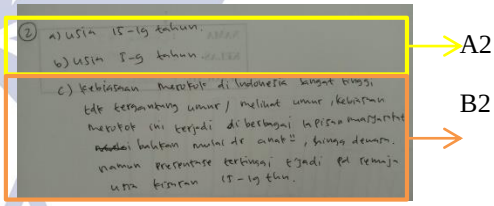


Figure 3. The answer of Subject S1 number 2a-2c

In problem numbers 1.f, 2.a and 2.b aim to find certain information from a graph. Subject S1 answered that in 2006-2007 the growth of internet users did not increase and the highest percentage of Indonesian started smoking at the age of 15-19 years, while the lowest percentage of started smoking at 5-9 years.

Based on data analysis and interview results, Subject S1 finds certain information that is in the diagram or graph of the problem presented, accompanied by clear and logical arguments. In other words, Subject S1 can search for information by referring to explicit observations of the data that is seen in the diagram or graph of the problem presented.

In problem numbers 1.g and 2.c, aim to achieve indicators explaining the data presented in the form of diagrams or graphs. First, Subject S1 replied that in 2006-2017 the growth of internet users had increased and there was no decrease at all even though in 2006 to 2007 internet users did not experience an increase. Second, Subject S1 answered that smoking habits in Indonesia are very high regardless of age, smoking habits occur in various walks of life even from children to adults, but the highest percentage occurs in adolescents aged around 15-19 years.

Based on data analysis and interview results, Subject S1 explains in general and details any information from diagrams or graphs such as stating the value of data that is explicitly seen or the value that is seen in a diagram or graph. In other words, Subject S1 can explain data in the form of diagrams or graphs that have been used in solving problems. This is consistent with what was described by Hafiyusholeh (2017) that students extract information directly from what is written explicitly based on existing diagrams or graphs.

In problem number 1.h aims to reach the indicators of conclusions and check the possibility of statistical information error. In this activity, the subject S1 concludes that based on a survey in 2017 that there was an increased Internet user from the year 2006 to 2017 and an average increase in the growth of Internet users is 11,2. Also, don't forget, the Subject S1 checks back on the answers that have been described.

Based on data analysis and interview results, S1 subjects make data conclusions by referring to data that is seen in the graph and based on one of the data centering values, which is the mean. This is following the results of Hafiyusholeh's research (2017) that students deduce data using trends general data to predict the likelihood of data to come and determine the value of the increase or decrease of data based on the average value. Subject S1 also rechecks the answers that have been described to ensure there is no misinformation, even though the calculation problem is not checked again. In other words, Subject S1 can draw conclusions based on information available on the problem presented, as well as checking for possible errors in the statistical information presented.

2. Description of high school students' statistical literacy in solving statistical problems with a moderate mathematical ability (S2).

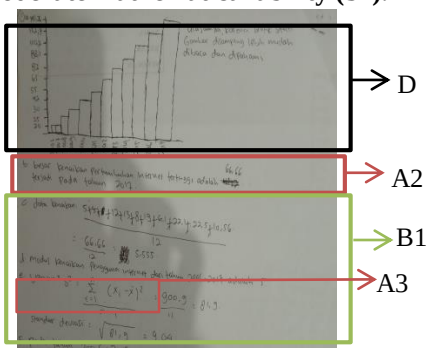


Figure 4. The answer of Subject S2 number 1a-1e

Based on data analysis and interview results, Subject S2 reads the problem randomly starting from the question to identify what was asked and reading as a whole to identify what is known from the problem of the growth of internet users and the Indonesian population started smoking. In other words, Subject S2 can read the data and identify what is known and asked from the problem presented in the form of tables and graphs. This is consistent with the results of research Sholikha (2018) that students can identify data or information provided using their language both written and oral.

In problem number 1.a aims to achieve the indicator to make a diagram or graph following the information you want to communicate. In this activity, Subject S2 draws *the* x-axis and y-axis, then draws the bar according to years and the internet user.

Based on data analysis and interview results, Subject S2 presents data from a tabular form into a bar chart accompanied by logical reasons, but Subject S2 experiences a misconception related to the image of the bar chart and histogram. Subject S2 considers the representation of a bar chart the same as a histogram so that Subject S2 draws a bar chart in the form of a rectangular arrangement which coincides on one side. Subject S2 in presenting data also does not consider the characteristics of the data, even though the data of internet user growth is continuous so it is more suitable to be presented in the form of a line diagram. In other words, Subject S2 has not been able to make diagrams or graphs following the information to be communicated and has not been able to determine diagrams or graphs following the data presented.

In problem number 1.b aims to find a certain value from a graph. Subject S2 answered that the highest increase in internet user growth occurred in 2017. This activity is carried out by Subject S2 by finding differences in the number of growth of internet users each year then adding up the entire data.

Based on data analysis and interview results, the Subject S2 experienced an error in explaining how to obtain value due to the growth of internet users. Subject S2 was correct by explaining looking for differences in internet user data from 2006-2017, but when asked to find the highest increase in internet user growth rates, Subject S2 added up all data increases in internet user growth, so that it could not find a certain value from a graph. In other words, Subject S2 can explain the data presented in the form of diagrams or graphs that have been used in solving problems but have not been able to find information by referring to the visible data.

In problem numbers 1.c, 1.d and 1.e aim to achieve indicators applying the concept and determining the concentration value and data distribution and the use of variables and symbols. First, Subject S2 determines what is needed to find the mean, mode, variance, and standard of the increase in internet user growth. After that, Subject S2 calculates the mean by adding up the data increase in internet users from 2006-2017 then divided by the data increase in internet users, but Subject S2 has an error in adding data. To find the value of the mode, Subject S2 looks at the value that often appears from the increase internet growth data that is 5, so the mode value is 5. For the variance, Subject S2 adds up the square value of each data increase in internet growth minus the average increase in internet growth then divided by the amount of data minus one. Whereas the standard deviation is the square root of variance.

Based on data analysis and interview results, Subject S2 experienced calculation errors in determining the mean value, so in determining the value of variance and standard deviation also experienced errors in terms of calculations, but in determining the mode value was correct. Subject S2 can explain the mean and mode definitions, as well as explain each variable or symbol used to solve problems related to mean, mode, variance, and standard deviation correctly. In other words, Subject S2 has not been able to apply the concept and determine the value of concentration and data distribution to solve the problem presented, and can determine the use of variables or symbols that correspond to the problem presented.

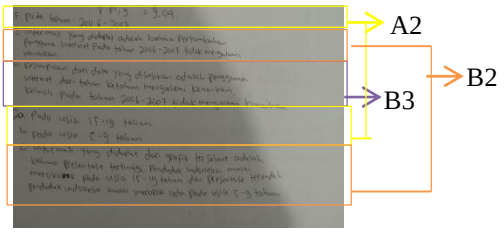


Figure 5. The answer of S2 number 1f-1h and 2a-2c

In problem numbers 1.f, 2.a and 2.b aim to find certain information from a graph. Subject S2 answered that in 2006-2007 the growth of internet users did not increase and the highest percentage of Indonesians started smoking at the age of 15-19 years while the lowest percentage of Indonesians started smoking at 5-9 years.

Based on data analysis and interview results, Subject S2 found certain information contained in diagrams or graphs of the problem of the growth of internet users and Indonesians who started smoking. In other words, Subject S2 can search for information by referring to explicit observations of the data that is seen in the diagram or graph of the problem presented.

In problem numbers, 1.g and 2.c, aim to achieve indicators explaining the data presented in the form of diagrams or graphs. First, Subject S2 answered that the growth of internet users in 2006-2007 did not increase. Second, Subject S2 answered that the highest percentage of Indonesians started smoking at the age of 15-19 years, and the lowest percentage of Indonesians started smoking at the age of 5-9 years.

Based on data analysis and interview results, Subject S2 explains in detail what information is obtained from diagrams or graphs, such as stating the value of data that is seen explicitly or the value that is seen on a diagram or graph. In other words, Subject S1 can explain data in the form of diagrams or graphs that have been used in solving problems.

In problem number 1.h aims to reach the indicators of conclusions and check for the possibility of statistical information error. Subject S2 concluded that internet users from year to year always increased without any decrease, except in 2006-2007 there was no increase, also Subject S2 did not re-check the answers that have been described.

Based on data analysis and interview results, Subject S2 makes data conclusions by referring to

data that is seen in the graph. This is consistent with the results of research Sholikha (2018) that students in make conclusions see the information. Subject S2 did not recheck the answers already described. In other words, Subject S2 can make conclusions based on the information available on the problem presented and does not fulfill the indicators to check for the possibility of incorrect statistical information presented.

CLOSING

Conclusion

Based on the results of data analysis and discussion related to the results of statistical problem tests, it can be described the statistical literacy profile of high school students in solving statistical problems based on mathematical ability, as follows.

Statistical Literacy Profile of High School Students in Solving Statistics Problems with High Mathematical Ability

Statistical literacy profile of a student with high mathematical abilities, in the aspects of understanding data or statistical information that is a student with high mathematical ability reads data carefully and systematically to identify what is known and asked, able to find certain information from a diagram based on explicit observations and be able to explain each variable or symbol which relates to the value of centralizing and distributing data in their language but has not been able to determine the use of diagrams according to the data presented. On aspects of interpreting data or statistical information, the student can define and determine the value of centralization and dissemination of data properly, capable to explain data in general and detail regarding the observations explicitly to specify a particular value of a chart or graph, as well as being able to make conclusions by referring to explicit observations and guided by one value of the concentration of data that is the mean. In the aspect of evaluating data or statistical information, the student checks back the answers that have been described. while on aspects communicate data or statistical information, the student can present data in the form of a chart or graph, but without considering the characteristics of the data in the presentation.

Statistical Literacy Profile of High School Students in Solving Statistics Problems with Moderate Mathematics Ability

Statistical literacy profile of a student with moderate mathematical abilities, in the aspect of understanding the data or statistical information that is the student with moderate mathematical ability reads random data to identify what is asked and known, capable to find certain information from a diagram and be able to explain each variable or symbol used by language itself, but have not been able to determine diagrams or graphs based on the data presented. In the aspect of interpreting data or statistical information, student experiences errors in determining the mean, variance, and standard deviation, but in determining the mode value is correct. The student explains data in detail by referring to explicit observations to determine certain values from a diagram or graph and can make conclusions by referring to explicit observations. In the aspect of evaluating data or statistical information, the student did not recheck the answers that have been described. While in communicating aspects of the data, the student can present data in the form of charts or graphs but had misconceptions associated with the image between bar chart and histogram, as well as the presentation of the data that does not consider the characteristics of the data presented.

Suggestion

Based on research that has been done, the advice given by researchers is as follows.

1. This study aims to describe the statistical literacy profile of high school students in solving statistical problems based on mathematical ability. In this study, in the aspect of communicating data or statistical information, students do not consider the characteristics of the data when presenting data in the form of diagrams or graphs. Therefore, it is important to emphasize the use and characteristics of data suitable for each diagram or graph.
2. When conducting interviews, researchers should look for a place away from the crowd, so researchers and students can concentrate on conducting interviews and get clear interview recordings.

REFERENCES

Ayuningtias, Diah. 2015. *Profil Literasi Statistis Siswa SMA/MA Ditinjau Dari Gaya Kognitif*. Tesis tidak diterbitkan. Surabaya: Universitas Negeri Surabaya.

Ben-Zhi, D and Garfield, J. 2008. *Developing Students' Statistical Reasoning: Connecting Research And Teaching Practice*. New York: Springer.

Gal, Iddo. 2002. *Adults' Statistical Literacy: Meanings, Components, Responsibilities*. International Statistical Review. Vol. 70(1): pp 1-51.

Gal, I. 2004. *"Statistical Literacy: Meanings, Components, Responsibilities"*. Dalam Ben-Zhi, D and Garfield, J (Ed.). 2004. *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. New York: Kluwer Academic Publisher. pp 47-48.

Hafiyusholeh, M. 2017. *"Literasi Statistik: Siswa SMA dalam Membaca, Menafsirkan, dan Menyimpulkan Data"*. Makalah disajikan dalam Seminar Nasional Integrasi Matematika dan Nilai Islami, Surabaya, Vol.1(1): pp 79-85, Juli 2017.

Hafiyusholeh, M. 2018. *Profil Literasi Statistis Siswa SMA dalam Menyelesaikan Masalah Statistika Ditinjau dari Perbedaan Gender*. Disertasi tidak diterbitkan. Surabaya: Universitas Negeri Surabaya.

Jones, G., Mooney, E., Langrall, C. & Thornton, C. 2002. *Student's Individual And Collective Statistical Thinking*, ICOTS6, (Online), (https://www.stat.auckland.ac.nz/~iase/publications/1/6a2_jone.pdf, diakses 1 Desember 2019).

Kemendikbud. 2016. *Permendikbud Nomor 24 Tahun 2016: Lampiran Kompetensi Inti dan Kompetensi Dasar Matematika SMA*. Jakarta: Kementrian Pendidikan dan Kebudayaan.

Meletiou, Maria and Lee, Carl. 2002. *Student Understanding Of Histogram: A Stumbling Stone To Development Of Intuition About Variation*. ICOTS6, (Online), (https://www.stat.auckland.ac.nz/~iase/publications/1/10_19_me.pdf, diakses 22 september 2019).

Miles, Matthew B. dan Huberman, A. Michael. 1992. *Analisis Data Kualitatif: Buku Sumber Tentang Metode-Metode Baru*. Jakarta: UI-Press.

OECD. 2018. *PISA 2021 Mathematics Framework (Draft)*. OECD Publishing.

Purwaningtyas, Rahmadita. 2018. *Profil berfikir reflektif siswa SMP dalam menyelesaikan masalah geometri ditinjau dari kemampuan matematika.* Skripsi tidak diterbitkan. Surabaya: Unesa.

Schield, M. 1998. *Statistical literacy and evidential statistics. ASA Proceedings of the section on statistical education*, (Online), (www.statlit.org/pdf/1998SchieldASA.pdf, diakses 1 Desember 2019).

Sharma, Sashi., Doyle, Phil., Shandil, Viney. and Talakia'atu, Semisi. 2012. *Developing Statistical Literacy With Years 9 Students: A Collaborative Research Project.* Dalam Smith, C (Ed). *Proceedings Of The British Society For Research Into Learning Mathematics*. Vol. 32 (3).

Yolcu, Ayse. 2012. *An investigation of eight grade students' statistical literacy, attitudes towards statistics and their relationship*. Published master's thesis. Turkey: Middle East Technical University