MCC Batch 7

by Ezra Putranda S

Submission date: 10-Dec-2020 05:38PM (UTC+0700)
Submission ID: 1470834007
File name: Statistical_literacy_rev_MCC.pdf (1.62M)
Word count: 6540
Character count: 37194

ISSN 2087-8885
E-ISSN 2407-0610

Journal on Mathematics Education
Volume xx, No. x, January xxxx, pp. x-xx

EXPLORING FIRST YEAR UNIVERSITY STUDENTS' STATISTICAL LITERACY: A CASE ON DESCRIBING AND VISUALIZING DATA

Ezra Putranda Setiawan[1], Heru Sukoco[1]

[1]Jurusan Pendidikan Matematika, Universitas Negeri Yogyakarta, Jalan Colombo 1, Yogyakarta, Indonesia
Email: ezra.ps@uny.ac.id

Abstract

Statistical literacy, which is the ability to use statistics in daily life, is an essential skill for facing society 5.0. This study aims to obtain some information about first-year university students' ability to use simple descriptive statistics and data visualization, which had studied in secondary school. Qualitative data were collected using a set of questions from 39 undergraduate students. The main results are as follows. The students were able to calculate various descriptive statistics, but some of them still unable to determine suitable statistics to describe the data clearly. Related to data visualization, many students drew the chart based on its easiness instead of the data or the purpose itself. Consequently, most of them produced an unsuitable or improper diagram. Improvement in statistical teaching – both in the university and the secondary schools – is needed so that the students can use descriptive statistics and data visualization correctly.

Keywords: Statistics education, data visualization, descriptive statistics, qualitative.

Abstrak

Literasi Statistika, yakni kemampuan untuk menggunakan statistika dalam kehidupan sehari-hari, merupakan salah satu kemampuan penting untuk menghadapi masyarakat 5.0. Penelitian ini bertujuan memperoleh informasi terkait kemampuan mahasiswa tahun pertama dalam mempergunakan statistik deskriptif sederhana dan visualisasi data yang telah mereka pelajari pada jenjang sekolah menengah. Data kualitatif dikumpulkan dari 39 mahasiswa menggunakan beberapa soal, dan diperoleh hasil sebagai berikut. Mahasiswa mampu menghitung bermacam statistik deskriptif, namun sebagian diantaranya tidak mampu menentukan statistik yang tepat untuk menggambarkan data secara jelas. Berkaitan dengan visualisasi data, lebih banyak mahasiswa yang menentukan bentuk visualisasi dari segi kepraktisan alih-alih bentuk data atau tujuan visualisasinya. Hal ini berakibat pada hasil diagram yang tidak sesuai atau tidak tepat. Diperlukan peningkatan kualitas pengajaran statistika – baik pada jenjang pendidikan tinggi maupun sekolah menengah – agar mahasiswa dapat mempergunakan statistik deskriptif maupun visualisasi data dengan tepat.

Kata kunci: Pendidikan statistika, visualisasi data, statistik deskriptif, kualitatif.

How to Cite: xxxxxxxx. & xxxxxxxx. (2020). ----------. *Journal on Mathematics Education, x* (x), xx-xx.

The 4.0 industrial revolution has occurred around the world, as indicated by the rapid growth of technology, the internet of things (IoT), artificial intelligence, and big data. As the volume of data grows at a rate of 50% per annum (Waal-Montgomery, 2016), understanding data is an essential skill for any data-driven society. Moreover, statistics can help people to filter the information by analysis of data and separate them from opinions (Widjajanti *et al.,* 2017). Based on this situation, the need for statistical literacy is increased (Frost, 2013). The definition and components on statistical literacy are vary among literatures (Sharma, 2017). First, it also can be defined as *transnumerative thinking* where people will be capable of making sense of and use a different representation of data to make sense of the situation among them (Chick *et al.,* 2005). Statistical literacy also can be seen in broader view, which consists of statistical understanding required in modern democracies and peoples dual role of statistical producer and consumer (Gould, 2017).

Due to its importance, statistical literacy competence should become a part of the school curriculum (Watson, 2003; Garfield & DelMas, 2010). In 2005, The Guidelines for Assessment and Instruction in Statistics Education (GAISE) framework was issued by the American Statistical Association (Metz, 2010). It is known that statistical teaching is rooted in the mathematics school curriculum at the K-12 school level (Weiland, 2017). In Brazil, statistics and probability are included in the mathematics curriculum (Campos *et al.,* 2011). Statistical literacy-related competencies areparts of the Indonesian mathematics curriculum from primary up to senior secondary school (Setiawan, 2019). Also the mathematical section of the Programme for International Student Assessment (PISA) test contains the 'data and uncertainty' part (Rahayu & Wijaya, 2018), which intersect with the components of statistical literacy. In the Trends of International Mathematics and Science Study (TIMSS), this section is known as 'data and chance' domain (Mills & Holloway, 2017).

Apart from the inclusion of statistical literacy competence on the school mathematics curricula, several studies showed that undergraduate students' had various level of statistical literacy. Yotongyos *et al.* (2005) presented that students from the Faculty of Education in a university in Thailand has moderate level of overall statistical literacy. A large-scaled assessment of statistical literacy involving more than 900 students in United States were done by Ziegler and Garfield (2018). Kim *et al.* (2019) did a test and observed the lesson plan to measure pre-service mathematics teachers' statistical literacy in Korea. In Pakistan, a survey revealed that BS students has low level of statistical literacy (Hassan *et al.,* 2020). Similar result were found in Indonesian students from bachelor of education in mathematics program and public administration program (Khaerunnisa & Pamungkas, 2017; Takaria & Talakua, 2018; Jatisunda *et al.,* 2020).

In general, statistical contents in undergraduate level are continuation of the secondary school level (Funny *et al.,* 2019). Following Indonesian curriculum, descriptive statistics and data visualization is the parts of statistical literacy related competence that appeared from the secondary and primary school level (Setiawan, 2019). Henceforth, it is reasonable to check whether the students are literate in these competencies before introducing the other aspects of statistics such as estimation, hypothesis testing, correlation analysis, and many more.

This study aims is to present undergraduate students' statistical literacy in terms of their ability to use descriptive statistics and visualize data appropriately. Description of statistical literacy could be used to check whether the students are literate in these topics.

METHOD

Approach and Subject

To extensively describe students' ability on descriptive statistics and data visualization, we used a qualitative approach. Following Creswell (2014), this type of study have natural situation and suitable to describe the actual result from the subjects. We provided no treatment or manipulation to thesubjects before and during the study.

The subject of this study was 39 students on first semester of study at the undergraduate/bachelor program in statistics of a public university in Indonesia. Most of the student were 17-18 year old, and9 (23,1 %) of them are males. One student had graduated from the pharmacy stream of vocational senior secondary school (*SMK*), where the other had graduated from the mathematics and science stream of general or Islamic senior secondary school (SMA/MA) in Indonesia. It is known that 32 (82%) students had finished their previous education in Java Island, whereas the other were came from outside Java (i.e. Sumatera, Bali, Nusa Tenggara Barat, and Kalimantan).

Data Collection

The data were collected using a test for students. On preparing the test, we assume that students have already known several descriptive statistics and data visualization, since these concepts arestudied in the primary and secondary school. Although some students have learned statistical inference in the twelfth grade of senior secondary school (Setiawan, 2020), we did not use this topic since it is taught only in the mathematics and natural science stream.

Formulation of the questions were inspired by a list of questions for assessing statistical education presented by Garfield and Ben-Zvi (2007) as well as Sharma (2017), with focus on using descriptive statistics and visualizing data. The questions given to the subjects (in Indonesian language) and its translation are presented in figure 1.

5. Anda telah mengenal berbagai macam statistik deskriptif (mean, modus, median, simpangan baku, dan lain lain). Misal dimiliki data sebaran **usia** penderita penyakit X di dua desa yang berbeda.

Desa A: 10, 12, 14, 35, 56, 58, 60

Desa B: 32, 34, 35, 36, 38

a) Seorang peneliti melaporkan bahwa *rata-rata dan median usia penderita penyakit X di desa A dan B adalah sama.* Setujukah anda dengan pernyataan tersebut? Jelaskan.

b) Apabila saudara diminta memberikan **satu** statistik deskriptif lain untuk memperjelas informasi yang diberikan peneliti pada nomor a di atas, statistik apa yang anda berikan? Hitunglah nilainya dan berikan alasannya.

c) Seseorang meminta anda menyajikan data tersebut di atas dalam bentuk diagram. Diagram apakah yang anda pilih? Buat dan jelaskan alasannya.

You know several descriptive statistics (eg. mean, mode, median, standard deviation, etc.). Suppose that we have the data of age (in years) of people with disease X from two different villages.

Village A: 10, 12, 14, 35, 56, 58, 60

Village B: 32, 34, 35, 36, 38

a) A researcher reported that the mean and median of age of peoples with disease X in village A is same with people in village B. Do you agree with that statement? Explain.

b) Suppose you are asked to give another descriptive statistics to explain the report given by the researcher mentioned in previous question, what statistics that you will give? Calculate it and give your reason.

c) Someone ask you to present the above data in form of a chart/diagram. What type of chart you will use? Draw it and give your reason.

Figure 1. Questions used in this study

As shown in figure 1, these questions had context, namely the patients' age which means that the data must be positive. The dataset consists of one variable with two categories or groups. Students with higher level of statistical literacy would aware to the presence of two groups and are able to show correct comparison between them.

The first question ensures that the respondents can calculate the mean and median from raw-ungrouped data. Since this is a closed question, students answer can be classified only into two groups, namely the correct and wrong answer. The correct answer for median and mean of these two groups were 35, which implies that they are equal.

The second question moves to the proper usage of descriptive statistics, while the last question checks the ability to visualize the data in a suitable form. On answering these questions, students were allowed to use calculator but not open any textbooks or references. Different from the test arranged in Garfield and Ben-Zvi (2007) or Jatisunda *et al.* (2020), we did not give any specific descriptive statistics nor chart type. Therefore, these questions can measure how respondents use their knowledge to decide the usage of descriptive statistics and chart types.

Data Analysis

The obtained data were analyzed using the descriptive-explorative method, which focused on the answer and the reason given by the subjects. Note that there are more than one correct answer for questions (b) and (c) since they are open question.

Table 1. Framework for classifying students answer in question (b) and (c)

Level	Answer of question (b)	Answer of question (c)
Low	Student provide non-sense statistics or repeat calculation of statistics that already used in (a).	Student provide incorrect chart or use wrong data to create the chart.
Middle	Student calculate descriptive statistics other than mean/median correctly, but failed to show the difference between the two groups *or* The calculation was incorrect.	Student create a chart but unable to show the difference between the two groups *or* Student failed to use a proper scale on the chart.
High	Student calculate descriptive statistics other than mean/median correctly and correctly presents the difference between the two groups.	Student create a chart using proper scale and clearly show the difference between the two groups.

We collected the answer sheet from all students, sorting alphabetically based on their name, then code the answer sheet as student #1, student #2, etc. The next step was the depth analysis of each students' answer on each question.

Regarding the descriptive statistics, we checked whether each student's calculation of descriptive statistics in question (a) and (b) was correct. These descriptive statistics answered on (b) then classified using the criteria presented Table 1. In the next step, we evaluate subjects' reason by made a list of all

reason used in each level of statistical literacy. Similar reason were grouped and were given same code. Representation of each group then presented in this paper.

Similarly, we examined the chart based on the criteria in Table 1, by paying more attention to the scale and the readability of the chart. Reason for drawing such chart from each subject were tabulated, coded, and grouped based on their content.

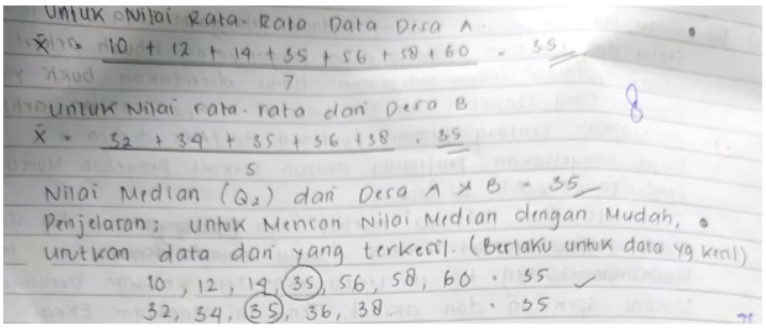
After all subjects' answer had tabulated, each researcher did the classifying and coding process independently. Later, the results were compared each other and different classification or code were discussed. Therefore, we can maintain the validity and objectivity of the process.

RESULTS AND DISCUSSION

Students' Ability on Using Descriptive Statistics

In general, any descriptive statistics can be calculated based on quantitative or numerical data manually or using an electronic calculator. Following Indonesian mathematics curriculum, students learn several descriptive statistics from primary school up to senior secondary school. In the primary school, they study how to calculate the mean, mode, and median, whereas, in secondary school, they explore the quartile(s) and range. In senior secondary school, students learned the absolute deviation, variance, and standard deviation. They also calculate each descriptive statistics studied before for the grouped data. Consequently, when starting their study at the undergraduate level, students are familiar and should be able to calculate and use various descriptive statistics.

From the answers to the first question, it can be seen that almost all respondents can calculate the mean and the median. They can show that both the mean and median of the patients' age from village A is equal to the mean and median of patients' age from village B. As seen in Figure 2, the calculation of these statistics are quite simple.



For the average of the village A data ....

For the average of the village B data ...

The median from village A and B ...
Explanations: to find the median, sort the data from the smallest

Figure 2. Sample correct answer for the first question

If the students recognize the difference between the data from these two villages, they should note that the second question asks them to give statistics that represent these differences. Based on this

idea, any statistics that the calculation based on the data from each village yields a different value is a correct answer.

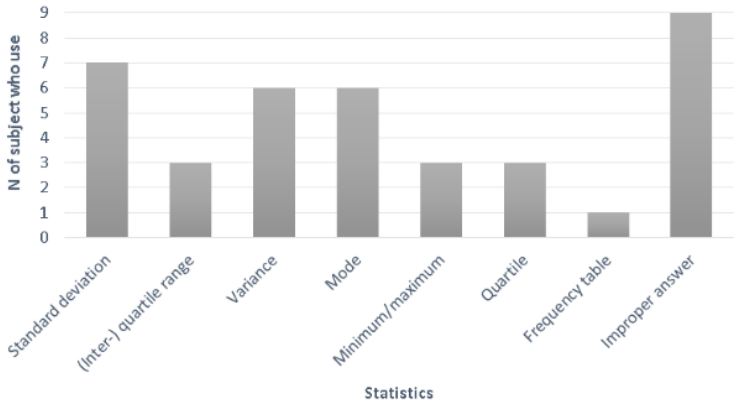


Figure 3. Descriptive statistics given by the respondent to answer the second question

Figure 3 shows that there are seven descriptive statistics given by the respondent to answer the second question, which were dominated by dispersion measures (i.e. standard deviation, interquartile range, and variance). Since all the data are different, the value of any dispersion measure between these two villages is different. Similarly, the first quartile, third quartile, minimum, and maximum of data from these two villages are different. More than half of all students belong to the high group, since they show the difference between these two groups of data using correct statistics. The most popular reason for this type of answer is to show the variability between the data, for example:

> *"Berdasarkan perhitungan dapat kita ketahui simpangan baku usia penderita penyakit X di desa A dan B tidak sama.* (From our calculation, we know that the standard deviation of age of patients with disease X in the village A and B are not equal)" (Student #9)

> "*Karena menunjukkan tingkatan keberagaman dari data tersebut.* (Because it presents the degree of variability of the given data)" (Student #30)

Students in the second group, namely in medium level of statistical literacy, could use other descriptive statistics but failed to show the difference of these group. This group dominated by students that (incorrectly) calculate the mode of the data, while the data for each village have no mode. Sample reason given by students in this group were as follows.

> "*Rata-rata, median, dan modusnya sama.* (The mean, median, and mode are same)" (Student #18)

> "*Karena nilai yang sering muncul hanyalah 35. 35 muncul paling banyak 2 kali, yang lain hanyalah sekali.* (Because the most frequent data is 35. 35 was appeared two times, while the others are only once)" (Student #21)

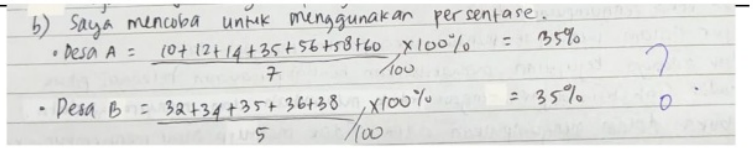
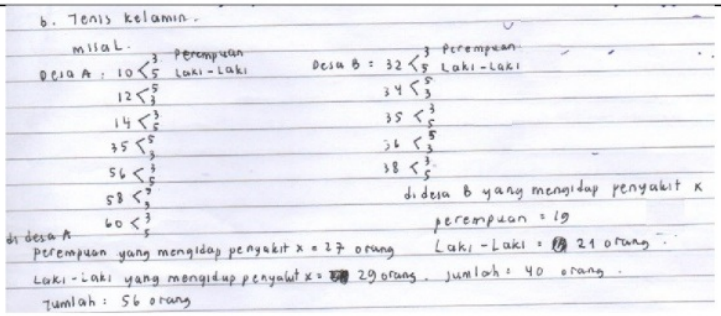
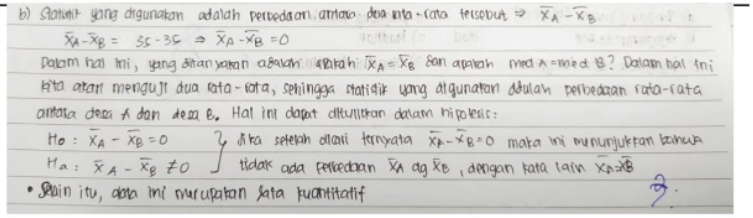
> "*Karena dengan banyaknya modus dapat mengetahui nilai yang sering keluar dan bisa memperjelas informasi.* (Because by using mode we are able to know the most occurred value and can get clearer information)" (Student #3).

It can be seen that the both Student #18 and Student #21 calculate the mode incorrectly, which cause them stated that the mean, median, and mode were same. On the other hand, Student #3 did not calculate the mode so that he/she might not realize that there were no mode in the data. Another type of students on this group provided wrong calculation of variance (i.e. did not taking the square of difference) which resulting in zero variance and did not provide another statistics.

In the third group, namely the lowest statistical literacy, we found several students that repeat the calculation of mean or median, or transforming the data into table. These students might not aware to several descriptive statistics mentioned in the questions, as presented by Student #28. Further classification of improper answer are presented in Table 2.

> *"Memberikan rata-rata merupakan informasi yang tepat dan pasti karena sudah diketahui berapa rata-ratanya.* (Present the mean is a correct and certain information because its value has been known)" (Student #28, present the mean which already calculated).

Table 2. Incorrect answer of question (b) represent low statistical literacy

Types of improper answer	Sample answer
Non-sense statistics: calculating the mean and present them as a percentage.	
Assuming presence of another information (*jenis kelamin* = sex) that did not given nor asked in the question.	
Doing hypothesis testing.	

From Table 2, an interesting result is that several students are unable to use the given dataset. They ask

more information, or, add some information that were not presented nor asked by the question, with some reason as follows.

> "*Statistik yang perlu ditambahkan seperti: jumlah penduduk desa A dan desa B; jumlah penduduk desa A dan desa B yang sehat* (Added statistics should be: the number of population in village A and village B; number of healthy people in village A and village B)" (Student #30)

> "*Dengan adanya jumlah penderita di tiap umur, maka kita dapat mencari nilai dan membandingkannya* (If the number of patients in each level of age was available, we can find the value and compare them)" (Student #33)

Reason given by Student #33 show that she did not understand that the presented data was the overall raw data instead of arranged data in the frequency distribution. Student #10, who purposed a hypothesis testing procedure to answer this question, might think that hypothesis testing is the part of descriptive statistics instead of inferential statistics.

Based on these results, it can be inferred that mostly students were know and able to calculate various descriptive statistics. Meanwhile, they might have poor competence in determining suitable statistics to describe and compare the raw data. Some of them also unaware that not all statistics (ie. mode) can be used in any dataset.

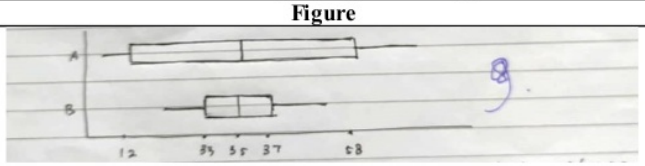
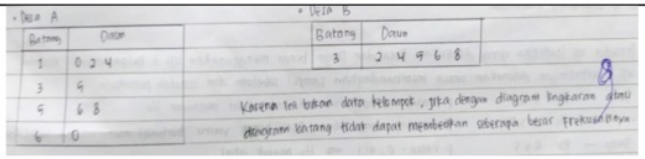
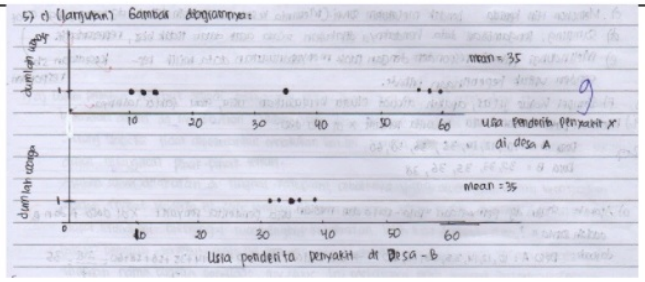
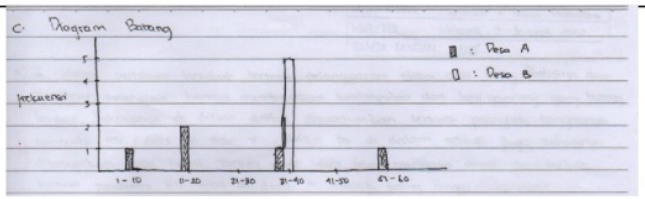
Students Ability on Using Data Visualization

As presented in Figure 1, the data in the question was ages of peoples with certain disease, which can be classified into interval or ratio scale. Theoretically, the suitable data visualization might be bar charts, dot plot, histograms, box (-and whisker-) plot, or stem-and-leaf plot.

From 39 participants of this study, only 37 participants that provide an answer for this questions. Most of the students (40%) use bar chart to display the data, which seem reasonable since this type of data visualization has been used from the primary school level up to the higher secondary school level. The box plot as well as the stem-and-leaf plot were used by one student each, while the dot plot used by four students. Most of the students had created two separated chart, or one chart for each villages, instead of combining these information into one chart. As a consequence, it is possible that same chart type (eg. bar chart) looks very different: one was present the difference between the two villages clearly, whereas the other contains various mistakes and difficult to understand.

Following the classification in Table 1, students' answers showing high statistical literacy are presented in Table 2. These charts clearly show the difference of patients' age from these two villages and use correct scale in the axes. Some of the students in this group presented all data in one chart, whereas the other used two separate chart. When the latter method was chosen, statistical-literate students must create same scale for each chart so that the reader can compare the data between these two groups easily.

Table 3. Sample proper chart types with clear/correct graphing

Figure	Explanations
	A boxplot clearly show the same median and the different variability of the data between these two groups.
	A stem-and-leaf plot was good for presenting the difference of variability between the two villages.
	A dot plot comparing the patients age, clearly show the variability of age between these two villages. The central tendency can be simply guessed from the display.
	A bar chart made by grouping the patients' age into five classes. The difference of patients' age among the two groups can be clearly seen.

In the group of students with medium level of statistical literacy, we find two kinds of answer as follows. Despite presenting the data in a suitable chart, several students did not give much attention to the scale on the axis. As presented in Figure 4, they only put the value of the data below the axis without see the difference between them. The chart still be able to read, but the comparison between those two groups would be difficult to do. Compared to charts on Table 2, these charts did not represent the difference of variability between these two groups.

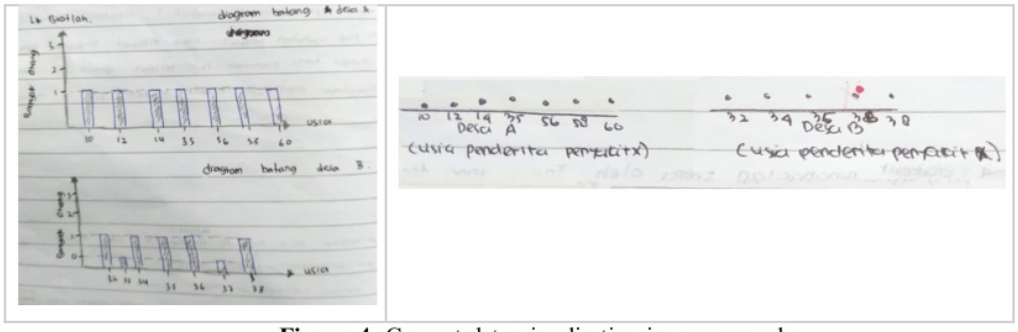
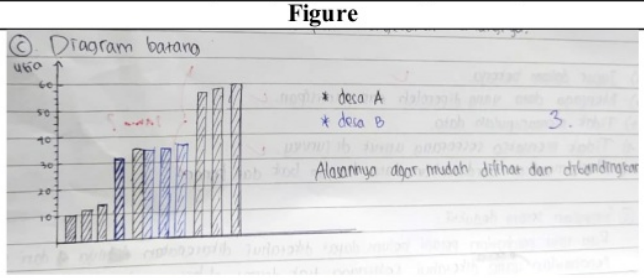
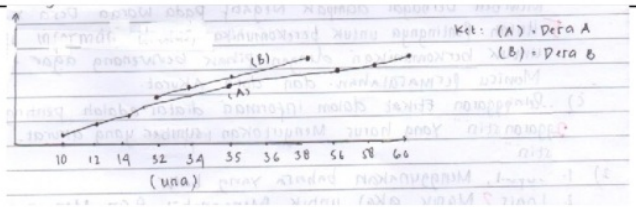
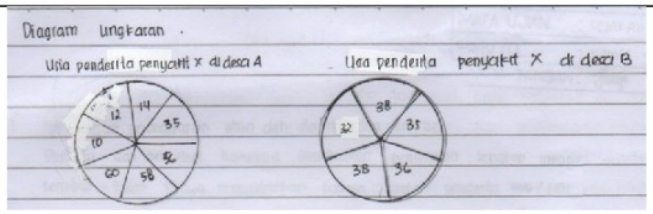
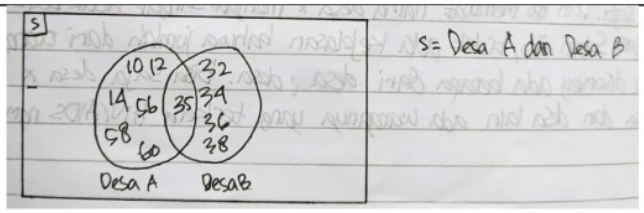


Figure 4. Correct data visualization in wrong scale.

Second, we found that several students use improper chart types to display the data, as presented in Table 4. Similar to those in Figure 4, these charts were unable to present the different variability of the data between these two villages.

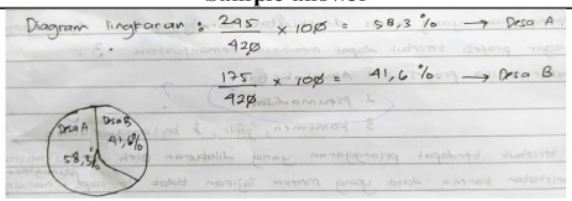
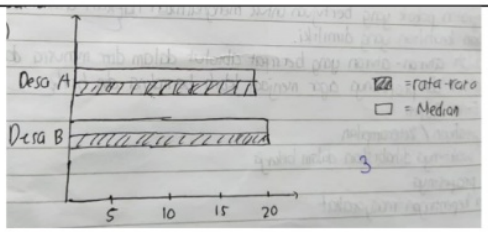
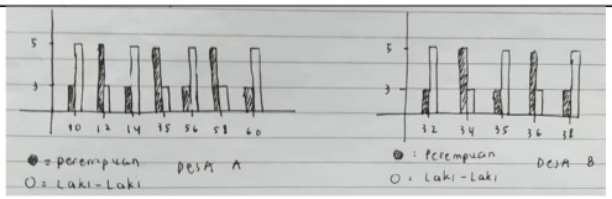
Table 4. Sample improper chart representing middle level of statistical literacy

Figure	Explanations
	Only convert the raw data into bars; no explanations on the horizontal axis. Although the difference between the two villages is represented by the colour, this chart is unacceptable.
	Since the data consists of only one variable, there was no reason to made two dimensional plot.
	A pie chart might represent the grouped data, but it is very difficult to compare the patients' age between these two groups.
	Although the content is correct, Venn diagram is used in set theory and not proposed to display numerical data. The variable also not mentioned in the diagram.

The correct parts of charts in Table 4 were only the variable and groupings of data. Instead of make the reader understand the data easily, these chart types might cause the reader confused. Almost all chart types there will be difficult to create and read if the number of subject is very large.

Students with low level of statistical literacy created several charts that did not represent the original data. These charts might use irrelevant data or unimportant variable. Such types of chart, which displayed in Table 5, represent the lowest competence on data visualization among all students participated on this study.

Table 5. Sample improper chart representing low statistical literacy on data visualization

Types of mistakes	Sample answer
Failed to identify the proper variable	
Wrongly present the descriptive statistics (which were same) instead of the original data.	
Add more information, namely the patients' sex that were not available in the original data.	

Why students choose a chart type to represent the dataset? Our study finds that seven respondents give no reason for their data display types. However, various reasons given by the students can be classified into four categories, as shown in table 6.

Table 6. Classification of student reason when visualize the data

Types of reason	Explanation	Sample answer
Data-based reason	Mention data properties, eg. number of variable, groups, types, etc.	"*Karena data yang diberikan merupakan data tunggal dan terdiri dari dua kelompok A dan B* (Because the data are single-valued and consist of two groups A and B)" (Student #16)
Chart-based reason*	Related to the chart types or chart usage.	"*Informasi dari data kuantitatif dapat terangkum dalam diagram ini.* (Information from quantitative data can be presented in this diagram)" (Student #10)
Purpose-based reason	Mention the purpose, ie. what are need to show or present from the data.	"*Dapat menampilkan hubungan jumlah penderita dengan umurnya.* (Can represent the relations between the number of patients and their age) (Student #31).
People-based reason	Related to the people who make and/or will read the diagram.	"*Agar mudah terbaca sehingga dapat dengan mudah menentukan simpulan mana yang mudah diambil.* (Will be easier to read, so that the conclusion can be taken easier)" (Student #13)

*In this study, all chart-based reason were combined to the data-based reason.

Among these types, more than half of the participants wrote people-based reason, while therarest was the purpose-based reason. How students' reason are related to the data visualization? Almost all students that draw correct diagram (perform high level of statistical literacy) wrote data-based reason, which sometimes combined to the other types of reason. In contrast, most of students with incorrect diagrams were only able to give people-based reason or give no reason at all.

Discussion

The goal of this study is to present undergraduate students' statistical literacy in terms of their ability to use descriptive statistics and visualize data. Usage of ill-structured essay question yield some benefit over the multiple choice question such as the ARTIST test (Garfield *et al.,* 2002). Various types of students answer can be found and classified to measure students' ability on using descriptivestatistics and data visualization. Compared to the questions provided by Sharma (2017), our test not onlymeasure students' ability to read and interpret the statistics or chart, but also to measure students' ability to use proper statistics or chart which should be interpretable and meaningful.

All respondent of this study were undergraduate students in Statistics major in their first semester. Therefore, it can be deduced that they were aware (and might has some interesting) to the statistical content in the senior secondary school mathematics. However, this study reveal that most of themhave medium to high level of statistical literacy in using proper descriptive statistics, but low to medium level of visualizing data. Although this study does not account students from other study programs, first-year university level with similar background might have similar profiles in statistical literacy before obtaining a statistical course in the undergraduate level.

Comparing to the other research on statistical literacy with similar subject, this study give an emphasis to the usage of describe statistics and data visualization. Some benefits of this focus is as follows. First, the students are asked to use higher order thinking skills, since they must evaluate which descriptive statistics or charts that should be used instead of just calculating a proposed statistics. As a result, students' difficulties to solve this problem correctly might also reflects their low ability on statistical reasoning, which needs to be confirmed using the other test (Sabbag *et al.,* 2018). Second, this result implies that teaching of statistics in undergraduate level needs to review the concepts on statistics that had been studied in the previous level, with emphasis on higher order thinking skills such as evaluating the usage of descriptive statistics and data visualization. Last, this study can reveal students' reasoning behind data visualization which can be grouped into four types (data-based, chart-based, purpose-based, and people-based). However, further research is needed to reveal the relations between the competence on data visualization and statistical reasoning.

How the problem of undergraduate students' statistical literacy can be solved? Numerous study show that sufficient ability in statistical literacy could not be developed by statistical teaching that focused on gathering statistical knowledge, learning facts and formulas, and obeying standard procedures (Schield, 2004). As a consequence, improvement in statistical teaching – both in the

university and secondary level education – is needed so that students become able to use the proper descriptive statistics and data visualization. Statistical learning in schools, which is dominated by computational aspects of statistics instead of conceptual understanding (Tiro, 2018), should be synchronized (Ridgway *et al.*, 2011).

Regarding the data visualization, teaching strategy using various modern tools has been developed, for example Nolan and Perret (2016) or Gelman and Nolan (2017). Following Wolfe (2015), several textbooks on communication courses can help us to find completed guidelines for determining the visualization types. Numerous literatures on students' difficulties (Boels *et al.*, 2019; Dewi *et al.,* 2020), misconceptions (Zaidan *et al.*, 2012; Chan & Ismail, 2013; Ismail and Chan, 2015; Yusuf *et al.,* 2017), as well as learning obstacles (Sotos *et al.,* 2007) can be used as references to develop proper strategies in teaching descriptive statistics. By noting that students' interest in statistical literacy competence is affected by prior mathematics achievement (Carmichael *et al.,* 2010, Jatisunda *et al.,* 2020) and self-efficacy (Carmichael *et al.,* 2010), the strategies used for teaching statistics should give emphasis on these variables. For example, teacher can use the statistical reasoning learning environment (Hidayah *et al.*, 2018), project-based learning strategy (Zhang & Fang, 2019), scaffolding approach (Murod *et al.,* 2019), guided discovery learning (Hariyanti & Wutsqa, 2020), as well as various technology (Suhermi & Widjajanti, 2020). Last, the multi-literacy model that frequently used in developing primary and secondary school curricula (Abidin, 2017; Nurgiyantoro *et al.,* 2020) should consider the statistical literacy.

The growth and presence of Big Data around the world can be used to develop further research as follows. Since many spreadsheets and statistical software are available, a survey can be done using larger dataset. Students can be asked to calculate the descriptive statistics and visualize the data using this software. More variables and categories can be implemented to the problem so that a higher level of statistical literacy can be evaluated.

CONCLUSION

This study analyzed the statistical literacy of first-year students from major statistics at a public university, with a focus on their ability to use descriptive statistics and to visualize the data. Most of the student are able to use descriptive statistics, but less of them are able to visualize the data properly. Students also use various reason in determining the suitable visualization of a given dataset. Although these competencies has been taught in the previous level, more reinforcement on descriptive statistics and data visualization should be given for the students so that they would be able to use them properly.

ACKNOWLEDGMENTS

The author thanks to Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Negeri Yogyakarta for gave some suggestion to improve the clarity of this manuscript in the Manuscript Coaching Clinic (MCC) Batch 7 2020.

REFERENCES

Abidin, Y. (2017). Developing Literacy Learning Model Based on Multi Literacy, Integrated, and Differentiated Concept at Primary School. *Jurnal Cakrawala Pendidikan*, *36*(2), 156-166. https://doi.org/10.21831/cp.v36i2.13283

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, *28*, 100291. https://doi.org/10.1016/j.edurev.2019.100291

Campos, T. M., Cazorla, I. M., & Kataoka, V. Y. (2011). Statistics school curricula in Brazil. In *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 5-8). Springer, Dordrecht.

Carmichael, C., Callingham, R., Hay, I., & Watson, J. (2010). Statistical literacy in the middle school: The relationship between interest, self-efficacy and prior mathematics achievement. *Australian journal of educational & developmental psychology*, *10*, 83-93.

Chan, S. W., & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia-Social and Behavioral Sciences*, *93*, 1478-1483. https://doi.org/10.1016/j.sbspro.2013.10.067

Chick, H. L., Pfannkuch, M., & Watson, J. M. (2005). Transnumerative thinking: finding and telling stories within data. *Curriculum matters*, *1*, 87-109.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE publications.

Dewi, D. K., Khodijah, S. S., & Zanthy, L. S. (2020). Analisis kesulitan matematik siswa smp pada materi statistika. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *4*(1), 1-7. https://doi.org/10.31004/cendekia.v4i1.148

Frost, J. (2013) Why statistics is important. *The World of Statistics*. Retrieved from http://www.worldofstatistics.org/2013/03/04/why-statistics-is-important/

Funny, R. A., Ghofur, M. A., Oktiningrum, W., & Nuraini, N. L. S. (2019). Reflective thinking skills of engineering students in learning statistics. *Journal on Mathematics Education*, *10*(3), 445-458. https://doi.org/10.22342/jme.10.3.9446.445-458

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International statistical review*, *75*(3), 372-396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Garfield, J., & DelMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics*, *32*(1), 2-7. https://doi.org/10.1111/j.1467-9639.2009.00373.x

Gelman, A., & Nolan, D. (2017). *Teaching statistics: A bag of tricks*. Oxford University Press.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22-25.

Hariyanti, F., & Wutsqa, D.U. (2020). Pengembangan perangkat pembelajaran statistika dan peluang untuk mengembangkan statistical literacy siswa SMP. *Jurnal Riset Pendidikan Matematika*, *7*(1), 46-58. https://doi.org/10.21831/jrpm.v7i1.14997

Hassan, A., Ghaffar, A., & Zaman, A. (2020). An Investigative Study on University Students' Statistical Literacy in Pakistan. *Sir Syed Journal of Education & Social Research*, *3*(1), 159-164. https://doi.org/10.36902/sjesr-vol3-iss1-2020(159-164)

Hidayah, N., Wahyudin, W., & Turmudi, T. (2018, December). The sport student's statistical literacy through statistical reasoning learning environment (SRLE). In *International Conference on Mathematics and Science Education of Universitas Pendidikan Indonesia* (Vol. 3, pp. 665-670).

Ismail, Z., & Chan, S. W. (2015, February). Malaysian students' misconceptions about measures of central tendency: An error analysis. In *AIP Conference Proceedings* (Vol. 1643, No. 1, pp. 93-100). American Institute of Physics. https://doi.org/10.1063/1.4907430

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

Khaerunnisa, E., & Pamungkas, A. S. (2017). Profil Kemampuan Literasi Statistis Mahasiswa Jurusan Pendidikan Matematika Universitas Sultan Ageng Tirtayasa. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *6*(2), 246-255. http://dx.doi.org/10.24127/ajpm.v6i2.970

Kim, D. E., Kang, P. L., & Lee, M. H. (2019). Analysis on preservice mathematics teachers' statistical literacy in lesson plan. *East Asian mathematical journal*, *35*(4), 429-449. http://dx.doi.org/10.7858/eamj.2019.036

Metz, M. L. (2010). Using GAISE and NCTM standards as frameworks for teaching probability and statistics to pre-service elementary and middle school mathematics teachers. *Journal of Statistics Education*, *18*(3). https://doi.org/10.1080/10691898.2010.11889585

Mills, J. D., & Holloway, C. E. (2013). The development of statistical literacy skills in the eighth grade: Exploring the TIMSS data to evaluate student achievement and teacher characteristics in the United States. *Educational Research and Evaluation*, *19*(4), 323-345. https://doi.org/10.1080/13803611.2013.771110

Murod, R. R., Priatna, N., & Martadiputra, B. A. P. (2019). The scaffolding approach to enhancesenior high school student's statistical literacy ability. *Journal of Physics: Conference Series, 1227*(1), 012028. https://doi.org/10.1088/1742-6596/1227/1/012028

Nolan, D., & Perrett, J. (2016). Teaching and learning data visualization: Ideas and assignments. *The American Statistician*, *70*(3), 260-269. https://doi.org/10.1080/00031305.2015.1123651

Nurgiyantoro, B., Lestyarini, B., & Rahayu, D. H. (2020). Mapping junior high school students' functional literacy competence. *Jurnal Cakrawala Pendidikan, 39*(3), 560-572. https://doi.org/10.21831/cp.v39i3.34061

Rahayu, N. A., & Wijaya, A. (2018). Hypothetical Learning Trajectory: A Case of Statistical Thinking. *Journal of Physics: Conference Series*, *1097*(1), 012109. https://doi.org/10.1088/1742-6596/1097/1/012109

Ridgway, J., Nicholson, J., & McCusker, S. (2011). Developing statistical literacy in students and teachers. In *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 311-322). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-1131-0_30

Schield, M. (2004). Statistical literacy curriculum design. *IASE Curriculum Design Roundtable.* Retrieved from *http://www.StatLit.org/pdf/2004SchieldIASE.pdf*

Sabbag, A., Garfield, J., & Zieffler, A. (2018). Assessing statistical literacy and statistical reasoning. *Statistics Education Research Journal*, *17*(2), 141-160.

Setiawan, E. P. (2019). Analisis muatan literasi statistika dalam buku teks matematika Kurikulum 2013. *Pythagoras: Jurnal Pendidikan Matematika*, 14(2), 163-177. https://doi.org/10.21831/pg.v14i2.28558

Setiawan, E. P. (2020). Introducing statistical inference to senior high school students: a textbook analysis. *Journal of Physics: Conference Series*, *1663*(1), 012014. https://doi.org/10.1088/1742-6596/1663/1/012014

Sharma, S. (2017). Definitions and models of statistical literacy: a literature review. *Open Review of Educational Research*, *4*(1), 118-133. https://doi.org/10.1080/23265507.2017.1354313

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational research review*, *2*(2), 98-113. https://doi.org/10.1016/j.edurev.2007.04.001

Suhermi & Widjajanti, D. B. (2020). What are the roles of technology in improving student statistical literacy?. *Journal of Physics: Conference Series*, *1581*(1), 012067. https://doi.org/10.1088/1742-6596/1581/1/012067

Takaria, J., & Talakua, M. (2018). The ability of statistical literacy student teacher candidate in terms of prior-ability on mathematics. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, *2*(2), 395-408. https://doi.org/10.21831/jk.v2i2.18768

Tiro, M. A. (2018). National Movement for Statistical Literacy in Indonesia: An Idea. *Journal of Physics: Conference Series*, *1028*(1), 012216. https://doi.org/10.1088/1742-6596/1028/1/012216

Waal-Montgomery, M.D. (2016). *World's data volume to grow 40% per year & 50 times by 2020* Aureus. Retrieved from https://e27.co/worlds-data-volume-togrow-40-per-year-50-times-by-2020-aureus-20150115-2/.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1-8. https://doi.org/10.1080/01621459.1993.10594283

Watson, J. M. (2003). Statistical literacy at the school level: What should students know and do. *The Bulletin of The International Statistical Institute, Berlim*, *54*, 1-4.

Widjajanti, D. B., Listyani, E., Susanti, M., & Setyaningrum, W. (2017, August). Examining Prospective Teachers' Mathematical Communication Skills in Statistics. In *1st Annual International Conference on Mathematics, Science, and Education (ICoMSE)*. Atlantis Press. https://doi.org/10.2991/icomse-17.2018.41

Wolfe, J. (2015). Teaching students to focus on the data in data visualization. *Journal of Business and Technical Communication*, *29*(3), 344-359. https://doi.org/10.1177/1050651915573944

Yotongyos, M., Traiwichitkhun, D., & Kaemkate, W. (2015). Undergraduate Students' Statistical Literacy: A Survey Study. *Procedia-Social and Behavioral Sciences*, *191*, 2731-2734. https://doi.org/10.1016/j.sbspro.2015.04.328

Yusuf, Y., Titat, N., & Yuliawati, T. (2017). Analisis hambatan belajar (learning obstacle) siswa SMP pada materi statistika. *Aksioma: Jurnal Matematika dan Pendidikan Matematika*, *8*(1), 76-86. https://doi.org/10.26877/aks.v8i1.1509

Zaidan, A., Ismail, Z., Yusof, Y. M., & Kashefi, H. (2012). Misconceptions in descriptive statistics among postgraduates in social sciences. *Procedia-Social and Behavioral Sciences*, *46*, 3535-3540. https://doi.org/10.1016/j.sbspro.2012.06.100

Zhang, H., & Fang, L. (2019). Project-based learning for statistical literacy: a gamification approach. In *Digital Turn in Schools—Research, Policy, Practice* (pp. 3-16). Springer, Singapore. https://doi.org/10.1007/978-981-13-7361-9_1

Ziegler, L., & Garfield, J. (2018). Developing a statistical literacy assessment for the modern introductory statistics course. *Statistics Education Research Journal*, *17*(2), 161-178.

MCC Batch 7

ORIGINALITY REPORT

7%	5%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Islam Bandung Student Paper	3%
2	www.tandfonline.com Internet Source	1%
3	repository.uhamka.ac.id Internet Source	1%
4	E P Setiawan. "Introducing statistical inference to senior high school students: a textbook analysis", Journal of Physics: Conference Series, 2020 Publication	<1%
5	Mohammad Palesh, Sten Fredrikson, Hamidreza Jamshidi, Goran Tomson, Max Petzold. "How is magnetic resonance imaging used in Iran?", International Journal of Technology Assessment in Health Care, 2008 Publication	<1%
6	Travis Weiland. "Problematizing statistical literacy: An intersection of critical and statistical literacies", Educational Studies in Mathematics,	<1%

2017
Publication

7	Failasuf Fadli, Siti Irene Astuti D., Rukiyati Rukiyati. "Techno - Resilience for Teachers: Concepts and Action", TEM Journal, 2020 Publication	<1%
8	Dhoni Hartanto, Asalil Mustain, Febry Dwi Nugroho. "Prediction of vapor-liquid equilibria for the alcohol + glycerol systems using UNIFAC and modified UNIFAC (Dortmund)", AIP Publishing, 2017 Publication	<1%
9	hal.archives-ouvertes.fr Internet Source	<1%
10	188.40.90.205 Internet Source	<1%
11	mpra.ub.uni-muenchen.de Internet Source	<1%
12	umaine.edu Internet Source	<1%
13	files.eric.ed.gov Internet Source	<1%
14	www.pmena.org Internet Source	<1%
15	ir.lib.hiroshima-u.ac.jp	

	Internet Source	<1%
16	Firda Hariyanti. "Statistical Literacy Siswa SMP dalam Pembelajaran Matematika", Ekspose: Jurnal Penelitian Hukum dan Pendidikan, 2020 Publication	<1%
17	mafiadoc.com Internet Source	<1%
18	"Third International Handbook of Mathematics Education", Springer Science and Business Media LLC, 2013 Publication	<1%
19	New ICMI Study Series, 2011. Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

MCC Batch 7

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17