Statistical reasoning levels of students in prediction tasks

Intan Sari Rufiana\textsuperscript{1,2}, Cholis Sa'dijah\textsuperscript{2}, Subanji\textsuperscript{2}, and Hery Susanto\textsuperscript{2}  
\textsuperscript{1,2}Mathematics Education, Universitas Negeri Malang, East Java, Indonesia  
\textsuperscript{3}Universitas Muhammadiyah Ponorogo, East Java, Indonesia

Email: intan.sari.1603119@students.um.ac.id

Abstract. Statistical reasoning is a major requirement in the era of big data. Therefore, there is a need to research this statistical reasoning. One important study in statistical reasoning is the levels of statistical reasoning. The levels of statistical reasoning are based on the construction of a conceptual framework. This research aimed to describe seven levels of students’ statistical reasoning in predicting data. The subjects of this research were 40 students of the second semester taken from two different classes. This was conducted to fulfill all levels of reasoning. At first, the subjects were given a test, and then seven students were selected to be interviewed related to the levels of statistical reasoning, respectively. Data triangulation was done to maintain the validity and reliability of the data generated. There are seven levels of statistical reasoning in predicting data, namely, the levels of pre-idiosyncratic, idiosyncratic, verbal, transitional 1, procedural, transitional 2, and integrated processes.

1. The first section in your paper

Nowadays, in the era of big data, statistics are needed in various fields [1–4]. The complexity of data dimensions has an impact on the increasing need for statistics [1]. Statistical reasoning is one of the important things in studying statistics in the class [5,6]. Besides, statistical reasoning is also important in problem-solving [7]. Moreover, [8] also explained that to be able to interpret, understand, and make good decisions regarding data, statistical reasoning is needed.

The researchers here limit this article only to the discussion of students’ reasoning related to prediction tasks, which are a part of the fourth statistical process of interpreting data. By only using this construct, it is expected that more detailed characteristics will be known for each level of statistical reasoning. In predicting, students do not have to give definitive answers to events that will occur, but try to find answers as close as possible [9]. This predictive ability is important for statistical applications in the real world, as described by [4,10].

One important study in statistical reasoning is the levels of statistical reasoning. There are five levels of statistical reasoning, namely levels of idiosyncratic, verbal, transitional, procedural, and integrated processes [6,11]. Students’ reasoning about elements of distribution and divide three levels of reasoning, namely nonstructural responses, multi structural responses, and relational responses [12], [16] examine the statistical reasoning of elementary and secondary schools students [6,13]. The research framework of [6] is based on two theoretical views. First, students who have statistical reasoning need to understand the concept of how to handle data that are diverse and evolve. Another is following the SOLO Taxonomy view that the development of students’ reasoning reflects a shift from their reasons complexity. This is the consideration of the researchers to reconstruct the levels of statistical reasoning among university students.
The researchers divided the complexity of students' reasoning into two components, namely the accuracy of identification and integration. [11] also use integration capabilities, but do not consider the component of accuracy in leveling. As they explain, students at the highest level of statistical reasoning can integrate the whole concept correctly. In prediction tasks, it is very important to consider the accuracy of identification. This accuracy consideration can be used to detect predictions with actual values. The more incoming information used for prediction, the prediction is given will tend to be better [14].

2. Method

2.1. Research Subjects and Instruments

This research involved subjects of 40 second-semester students from two different classes so that the overall levels of statistical reasoning could be fulfilled. Initially, all subjects were asked to work on the questions related to predictions. Questions were taken from the research conducted by [13]. The subjects were chosen based on the criteria shown in Table 1.

| Reasoning Levels | Description |
|------------------|-------------|
| Idiosyncratic (IR) | Students can use several statistical words/symbols but cannot fully understand and connect with the appropriate information. |
| Verbal (VR) | Students can find out the definitions of some statistical ideas, but they still fail to apply them correctly. |
| Transitional (TR) | Students can recognize one or two aspects of the statistical process but cannot include some concepts to answer questions. |
| Procedural (PR) | Students can identify statistical processes accurately but are still weak in their ability to fully understand and are unable to integrate them. |
| Integrated Process (TPR) | Students have full knowledge in the statistical process and can integrate it. |

Based on the criteria in Table 1, the answers of forty students were grouped into five levels of statistical reasoning related to prediction tasks. Distribution of students’ answer in five levels of statistical reasoning is as follows Pre-Idiocypratic Reasoning (PIR) as much 6 students, Idiosyncratic Reasoning (IR) 5 students, Verbal Reasoning (VR) as much 7 students, Transitional Reasoning I (TR 1) as much 6 students, Procedural Reasoning (PR) as much 7 students, Transitional Reasoning II (TR II) as much 5 students, Integrated Process Reasoning (IPR) as much 4 students.

According to the distribution of answers, the researchers interviewed students of each level of reasoning, which had good communication skills. The students taken were Subject 1 (S1) for PIR, Subject 2 (S2) for IR, Subject 3 (S3) for VR, Subject 4 (S4) represents TR 1, Subject 5 (S5) for PR, Subject 6 For TR II and Subject 7 for IPR.

2.2. Data Collection

The researchers conducted data collection through question instruments and interview guidelines. The researchers also prepared a holistic rubric before giving assignments to students. The researchers used the characteristics of the five levels of statistical reasoning used [6,11] as an assessment rubric. In [7], it is explained that there are two kinds of rubrics to assess students’ work, namely holistic rubric, and analytical rubric. A holistic rubric is used to describe the quality of performance [15], whereas the analytical rubric is used to assign values to the tasks component [16].

Questions were given to 40 students done individually. From the results of the students' works, task-based interviews were then conducted. Task-based interviews can be used to assess subjects [17]. At the interview, the questions instrument was given to students. Then, students were asked to explain the results of their works using a computer if necessary. The recordings were then transcribed to be encoded.
Triangulation of data from the interview video display and the answers results of the students' written test were conducted to maintain the validity and reliability of the data [18,19].

2.3. Data Analysis
The results of interviews through videos were viewed several times. These results were then coded based on the description of the levels of statistical reasoning as written in Table 1. The results of students' written tests were also examined using the assessment rubric that has been made. The coding results were then used to examine the relations with the results of the written test, and then to be used to determine each category of reasoning level. After the categories of each student's levels of statistical reasoning were obtained, they were then used to carry out the statistical reasoning levelling process. This category was carried out starting from the beginning, namely through reviewing the literature then proceeding to analyse the answers from the results of interviews and tests.

3. Result and Discussion
3.1. Result
The followings are the results of the research described by the researchers by explaining each student's answers group showing the levels of statistical reasoning in making predictions. Students who were at a pre-idiosyncratic level could be seen when they could not provide predictions, as shown in the student's answer, as shown in Figure 1.

![Figure 1](image1.png)

**Figure 1** The student’s answer of pre-idiosyncratic level

Figure 1 indicated that the student could not give the reasoning related to prediction tasks. The student said that the prediction of consumption of instant noodles in Malaysia was 9.29 while in Taiwan was 9.2 because there were data whose values were greater than other data so that they affected the values of other data. Moreover, the years before 2018 could not be used to find predictions in 2018.

When the researchers asked the student the questions why the student was not able to give conclusions, the student explained that because in the data there were years where consumption of instant noodles increased and decreased as shown in the following interview snippet:

\[ P : "Why \ did \ you \ answer \ the \ question \ by \ stating \ that \ the \ data \ could \ not \ be \ used \ to \ predict?" \]

(The researcher pointed to student’s answers.)

\[ S1 : "If \ you \ see \ the \ data, \ there \ is \ a \ drop \ (pointing \ to \ data \ on \ instant \ noodles \ consumption \ in \ 2014 \ until \ 2017), \ and \ there \ is \ a \ rise \ (pointing \ to \ data \ on \ the \ consumption \ of \ instant \ noodles \ in \ Taiwan \ in \ 2012 \ and \ 2013)." \]

\[ P : "Does \ it \ mean \ that \ predictions \ cannot \ be \ made?" \]

\[ S1 : "Uhhh ... No, it can't, Ma'am." \]

\[ P : "Did \ you \ perhaps \ consider \ a \ certain \ size?" \]

\[ S1 : "Hmmm ... there's nothing ... there's none, Ma'am (shaking head)." \]

The second level is idiosyncratic. Students at the idiosyncratic level know some words and symbols but use them without really understanding. This level can be shown in Figure 2.

![Figure 2](image2.png)

**Figure 2** The idiosyncratic level of student’s answer

From this data, we can draw conclusions about the mean consumption of instant noodles for Malaysia and Taiwan in 2018 in the previous 6 years. Based on the conclusions, we can predict the consumption of instant noodles for Malaysia and Taiwan in 2018.
Referring to Figure 2, the student wrote that from this data, we could conclude the mean consumption of instant noodles for Malaysia and Taiwan in 2018 in the previous six years. Based on the conclusions, we can predict the consumption of instant noodles for Malaysia and Taiwan in 2018. From these conclusions, it can be used to predict the consumption of instant noodles in 2018. But, this student used the mean term without really understanding the definition. In the interview process, the student reasoned that when there were data like these, the most frequently done was to calculate the mean as described in the following interview snippet:

P : "Why did you give that answer?" (The researcher pointed to the term "mean" used by the student.)
S2 : "We can count the predictions if the mean has been determined."
P : "Why should you use the mean?"
S2 : "Well... for this kind of data, the mean is usually calculated, Ma'am."
P : "Does it have to be like that?"
S2 : "Yes, Ma'am, my high school teacher, said so."
P : "If there are other problems, for example, the data of consuming instant noodles in the first few years increased, and went down, is it necessary to use the mean?"
S2 : "Yes, Ma'am. If there are numbers, the final mean will be calculated."

The third is the verbal level. The characteristic of the students at this level is the ability to understand verbally. The students cannot connect with actual behavior. Often students give their reasoning ambiguously and unclearly. This level can be seen in Figure 3.

![Figure 3 The student’s answers of verbal level](image)

From Figure 3, the student calculated the mean. Based on these answers, the student was able to state that the mean can be used to predict the consumption of instant noodles in 2018. However, the student viewed the mean here, not as a measure of concentration. This student thought that the mean could be searched by adding the entire consumption data and then dividing it by the number of years, as stated in the following interview snippet:

P : "Why did you give that answer?" (The researcher pointed to the mean calculation value done by the student.)
S3 : "We can count the predictions if the mean has been determined."
P : "Why should you use the mean?"
S3 : "Well... by using the mean, we will be able to add the overall data later and then divide it by the number of years."
P : "Did not you use other measures?"
S3 : "No, I did not, Ma'am (seeing the work while shaking the head)."
P : "Were there other considerations that you used to predict the consumption of instant noodles in 2018?"
S3 : "Hmmm ... there were no others, Ma'am. (shaking the head again)
Figure 4 The student’s answer of Transitional Level 1 by first calculating the mean then deducing Malaysia having more consumption on its mean.

The next level is the transitional level 1. The students at this level have characteristics that they can give and choose the correct definition but do not understand the concept completely, as written in Figure 4a and Figure 5.

Figure 5 The student’s answer of transitional level 1 by calculating the median after the mean and then concluding Taiwan having more consumption.

Figures 4a and 5 indicated that the students calculated the mean. In addition to calculating the mean value, the students at this level also considered the median value. Based on these answers, the student wrote that it was not enough to determine the mean value to predict the data; the median value must also be seen. The student concluded that the consumption of instant noodles in Malaysia was bigger than that of Taiwan if viewed from the mean. But the student also concluded that the consumption of instant noodles in Taiwan was bigger than that of Malaysia when viewed from the mean. The student finally concluded that the median size was more accurate than the mean. The student reasoned that this means would be greatly influenced by extreme values. But this student could not understand the actual mean concept because the student did not connect the answer with the data. It can be seen from the student’s inability to associate the form of distribution and concentration value through graphs/diagrams as stated in the following interview snippet:

\[ P : \quad \text{“Why are there two different answers? When using the mean, the consumption of instant noodles in Malaysia was greater, but when using the median, the consumption of instant noodles in Taiwan is greater?”} \]

\[ S4 : \quad \text{“It is not enough to use the mean itself, Ma’am. We must also consider the median value to conclude.”} \]

\[ P : \quad \text{“Then the conclusions are different between calculations using mean and median?”} \]
S4 : "Yes, Ma'am. They are different. When calculated by the mean, the prediction of instant noodles consumption in 2018 in Malaysia is greater. Meanwhile, if calculated by the median, the prediction will tend to be greater of instant noodles consumption in Taiwan."

P : "At last, the consumption of instant noodles in Taiwan was bigger than Malaysia. Why?"

S4 : "Well ... (thinking) ... because ... (while reading his work)"

P : "Are you in doubt with the answers?"

S4 : "Hmm ... Malaysia should be bigger, Ma'am. (smiling)"

The fifth level is the level of procedural reasoning. The students having the characteristics in this level include those who understand the concept procedurally. But the students are unable to integrate fully. This level can be seen in Figure 6.

![Figure 6 The student’s answers of procedural level](image)

Figure 6 indicates that the student calculated the quartile values 1 and 3 and the median data on the consumption of instant noodles. The student also calculated the minimum and maximum values of instant noodle consumption. These values were then used to describe the box plot form of consumption of instant noodles from each country. The student made predictions by stating that the consumption of instant noodles in Taiwan in 2018 was around 8.5 to 10.0, whereas in Malaysia, it might be able to increase or decrease. From the answers above, the student was seen procedurally indicated how to describe the box plot. The box plot was then used to make predictions. But this student could not associate the form of distribution, the size of the concentration and the size of the spread, as explained in the following interview excerpt:

P : "How did you compare the distribution of the box plot?"

S5 : "By looking at the box plot, Ma'am"

P : "How did you see the box plot?"

S5 : "I made the box plot from the lowest value, quartile 1, quartile 2, quartile 3, and maximum value"

P : "How could you make a final prediction that the prediction of instant noodles in Taiwan ranged from 8.5 to 10.0?"

S5 : "Maximum and minimum values, Ma'am."

P : "What about the predictions of instant noodles in Malaysia?"

S5 : "The minimum and maximum values, Ma'am."

P : "There were no other reasons for consideration?"

S5 : "Hmmm ... what was it? (see the results of the work) ... nothing."
The student’s answers of Transitional Level 2 firstly calculated the maximum, minimum, and quartile values then used those values to draw a box plot.

Henceforth is transitional level 2. The students at this level have the same characteristics as procedural level, but the students have been able to connect between several related, but incomplete rules as written in Figures 7 and 8.

**Figure 7** The student’s answers of Transitional Level 2 firstly calculated the maximum, minimum, and quartile values then used those values to draw a box plot.

The student’s answer of transactional level 2 in making a conclusion.

The graphs between Malaysia and Taiwan were almost the same because the graphs both tended to form a distribution with a positive skewness. The consumption of positive instant noodles in Taiwan in 2018 ranged from 8.5 to 10.0. But if it was only in Malaysia, it could possibly experience an increase/decrease in the consumption of instant noodles seen from the minimum and maximum values.

**Figure 8** The student’s answer of transactional level 2 in making a conclusion.

From the student’s answers above that, the graphs between Malaysia and Taiwan were almost the same because the graphs both tended to form a distribution with positive skewness. The consumption of positive instant noodles in Taiwan in 2018 ranged from 8.5 to 10.0. But if it was only in Malaysia, it could possibly experience an increase/decrease in the consumption of instant noodles seen from the minimum and maximum values, it can be seen that the student calculated the quartile value 1, quartile 2, quartile 3, and the minimum and maximum values of data on the consumption of instant noodles in two countries as procedural students. These values were then used to describe the box plot form of consumption of instant noodles from each country. The difference with procedural students, transitional level 2 students were able to state that the graph distribution between Malaysia and Taiwan was almost the same, which tended to have positive skewness. The only difference was in measures, as explained in the following interview excerpt:

\[ P : \] "How did you compare the distribution of the box plot?
\[ S6 : \] "Well ... it was from the form of the distribution ... (looking at the work)
The distribution was both equally positive skewness."
\[ P : \] "If the distribution was both positively slant, how could you predict that the consumption of instant noodles in Malaysia was higher than in Taiwan?"
\[ S6 : \] "heee ... (while laughing) Finally, I compared the maximum and minimum values of the graphs because they both have positive skewness."
\[ P : \] "There were no other reasons for consideration?"
\[ S6 : \] "Hmmm I do not know ... what is it then? (shaking the head). I did not compare it."

The next level is the integrated process level. The students at this level have the characteristics that can control rules and behavior in a complete and sure manner, as written in the answer as shown in Figure 9.
Malaysia's Box Plot, the consumption of instant noodle was more than Taiwan’s. The measures of concentration, the mean is certainly greater than the median because the boxplot has a positive slope. The measures of the spread and the consumption of instant noodles in Malaysia spread more than in Taiwan. Likewise for Taiwan, this is because basically the distribution was the same with Malaysia. The measure of the concentration mean was greater than the median. The measures of the spread used the interquartile range not the deviation standard. Therefore, the predicting of 2018, the consumption of instant noodles in Malaysia would be greater than in Taiwan, given the interquartile range of 10.6-8.3 = 2.3 (Malaysia) and 9.5 -8.7 = 0.8 (Taiwan)

Figure 9 The student’s answers of integrated process level previously started from calculating the maximum, minimum, and quartile values, and then used those values to draw a box plot then compared to make a complete conclusion

From the student’s answers as shown in Figure 9, it can be seen that the student calculated the quartile value 1, quartile 2, quartile 3, minimum and maximum values of data on the consumption of instant noodles. These values were then used to describe the box plot form of the consumption of instant noodles from each country. The student stated that the distribution of the graph between Malaysia and Taiwan was almost the same, which tended to have a positive slope. This student explained the measures of the concentration that the mean was greater than the median because the shape of the distribution positively slanted. Meanwhile, the measures of the spread, the consumption of instant noodles in Malaysia was more widespread than in Taiwan, as explained in the following interview excerpt:

P : ”How did you compare the distribution of the box plot?”
S7 : “The box plot was almost the same, Ma’am. It was skewness positively. But the one showing the consumption of instant noodles in Malaysia was longer.”
P : ”How did you know if it was longer?”
S7 : “Well... the one regarding Malaysia was from number 7, it was a little more. and from 12, it was a little less. Meanwhile, the one regarding was just this.”(while pointing to the box plot in the results of the work)
P : ”What about the measures of the central tendency?”
S7 : “The measures of the central tendency, which was similar between Malaysia and Taiwan, was the median. It was because the mean was greater than the median drawn by the large value here.” (while pointing to the tail from the box plot that sticks to the right)
P : ”What about the measures of the spread?”
S7 : ”The measures of the spread of instant noodles consumption in Malaysia were more widespread than in Taiwan because the range of Malaysia was bigger.”
P : ”Range? Why was Malaysia's range bigger?”
S7 : ”That is, Ma’am. The difference between the maximum value and the minimum value of the consumption of instant noodles in Malaysia was greater than in Taiwan.”

3.2. Discussion
The researchers found the development of five levels found by [6,11] becomes seven levels of statistical reasoning on prediction tasks. This addition is the pre-idosyncratic level before the idiosyncratic level and the transition level 2 before the integrated process level. The students with pre-idosyncratic reasoning level provide predictions without using statistical words/symbols. These students tend to only use subjective reasons or informal reasons. The tendency of students to imagine data as a collection of individuals is not as a collection that has certain characteristics. In the early stages, students have not used thinking awareness, are intuitive, tend to use all available data in the questions regardless of their relevance or focused on irrelevant aspects [20,21]. In a previous study, the researchers named [22] found that students were not able to make conclusions. The use of these students’ informal
reasoning was the students’ efforts in the search structure. As described by [23] that informal reasoning is characterized as a search structure while formal reasoning is a structure of use. This is in line with the statement of [24] that intuitive thinking is usually used as a hypothetical determinant to develop knowledge and be used for analytical and procedural thinking.

The students with an idiosyncratic level of reasoning already know several words and symbols but use words/symbols without really understanding them. This error is similar to the [23,25] that students tend to only memorize statistical terms or symbols but cannot relate them to data/graphs. A similar error was also found by [23] that at the end, students made wrong predictions. [25] also states that when entering the beginning of learning, students usually enter the class using alternative conceptions that are counter-intuitive to what is in learning.

The students with a level of verbal reasoning understand verbally. However, the students at this level cannot connect with actual behavior. The students only memorize concepts not understand concepts as explained by [11] that the students at this level describe predictions from data/graphs as incomplete.

However, the students with transitional reasoning level 1 have the characteristics that they can give and choose the correct definition. These students do not understand the concept completely only partially [26]. It can be seen from the students’ inability to associate the form of distribution and concentration value through graphs/diagrams as explained by [11] that students are not able to associate their answers with data/graphs when comparing a set of data. The ability to use this diagram is considered as the basis of reasoning [27].

The students with procedural reasoning level can manipulate data to be able to predict correctly [28]. However, these students are not able to integrate it. As explained by [11] that students are proficient in making comparisons of data sets but cannot fully explain. This procedural reasoning ability is usually obtained because of the existence of training with guidance and appropriate feedback [29,30].

The students with transitional reasoning level 2 can manipulate data to be able to predict correctly. But, the difference between transitional reasoning level and procedural level is that the students can explain some of the relations among concepts. The students at this level cannot be included in the previous level, namely the procedural level because the students can associate between concepts, even if in part. The students at this level also cannot be included in the level afterward, namely the integrated process level because the concept of integration ability is not yet full. The existence of this level shows a shift from the increasingly complex reasons of students, as explained by the SOLO Taxonomy. The accuracy of students’ predictions in this level is higher than in the procedural level because the students have been able to integrate the concept even though it is not fully. The consideration of accuracy in identifying this is one of the important points in prediction tasks. This accuracy consideration can be used to detect predictions with actual values. Good predictions are predictions that use the linkages between information with one another so that if more information is entered, the predictions produced will be even better [14].

The students with integrated process level of reasoning can make predictions in full with confidence. These students can associate all the concepts underlying the predictions made. Following the Solo Taxonomy hierarchy at the extended abstract level, the students at this level can think inductively and deductively and make connections between information to conclude. This is done to build a new concept and apply it. This is supported by [11,12] that students at this level make comparisons of data as a whole by associating with actual data and graphs.

4. Conclusion
It can be concluded that there are seven levels of university students’ statistical reasoning in prediction tasks, namely pre-idiiosyncratic level, idiiosyncratic level, verbal level, transitional level 1, procedural level, transitional level 2 and integrated process level. The students with pre-idiiosyncratic reasoning level provide predictions without using statistical words/symbols. These students do not use reasoning or tend to use only subjective reasons or informal reasons. Meanwhile, the students with an idiiosyncratic level of reasoning already know several words and symbols but use the words/symbols without really understanding them. The students in verbal level are only able to understand verbally. The students at this level cannot connect with actual behavior. Often the reasoning is ambiguous and unclear. The
students only memorize concepts not understand concepts. The students with transitional level 1 show characteristics of being able to give and choose the correct definition. Meanwhile, the students in procedural reasoning level can manipulate data to be able to predict correctly. Furthermore, the students with a transitional level of reasoning 2 show the characteristics of being able to manipulate data and to predict correctly. The difference between transitional level 2 and procedural level is the students can explain some of the relations between concepts. Lastly, the students at integrated process level can make predictions in full with confidence. These students can associate all the concepts underlying their predictions.

Acknowledgments
I want to thank the Lembaga Pengelola Dana Pendidikan (LPDP), Departemen Keuangan Indonesia as a funder in publishing this article.

References
[1] Adler R J, Agami S and Pranav P 2017 Modeling and replicating statistical topology and evidence for CMB nonhomogeneity Proc. Natl. Acad. Sci. 114 11878–83
[2] Chan J M, Carlsson G and Rabadan R 2013 Topology of viral evolution Proc. Natl. Acad. Sci. 110 18566–71
[3] Hiraoka Y, Nakamura T, Hirata A, Escolar E G, Matsue K and Nishiura Y 2016 Hierarchical structures of amorphous solids characterized by persistent homology Proc. Natl. Acad. Sci. 113 7035–40
[4] Kauser Ahmed P 2017 Analysis of data mining tools for disease prediction J. Pharm. Sci. Res. 9 1886–8
[5] Garfield J and Chance B 2004 Assessment in Statistics Education: Issues and Challenges Math. Think. Learn. 2 99–125
[6] Garfield J 2002 The challenge of developing statistical reasoning J. Stat. Educ. 10
[7] NCTM 2000 Principles and Standarts for School Mathematics (Reston: The National Council of Teachers of Mathematics, Inc.)
[8] Yusuf Y 2017 Statistical Reasoning Construction On Statistics Research Statistics Scholaria 7 60–9
[9] Heizer J and Render B 2008 Operation Management (New Jersey: Personal International Edition)
[10] Haddawy P, Hasan A H M I, Kasantikul R, Lawpoolsri S, Sa-angchhai P, Kaewkungwal J and Singhasivanon P 2018 Spatiotemporal Bayesian networks for malaria prediction Artif. Intell. Med. 84 127–38
[11] Chan S W, Ismail Z and Sumintono B 2016 A framework for assessing high school students’ statistical reasoning PLoS One 11 1–32
[12] READING C and REID J 2006 An Emerging Hierarchy of Reasoning about Distribution: From a Variation Perspective Stat. Educ. Res. J. 5 46–68
[13] Doyle K M, Dias O, Kennis J R, Czarnocha B and Baker W 2015 The Rational Number Sub-Constructs as a Foundation for Problem Solving Adults Learn. Math. An Int. J. 11 21–42
[14] Fathony R Z A Statistika dan Machine Learning: Satu Ilmu Dua Wajah 5
[15] Utami A D 2018 Six Levels of Indonesian Primary School Students’ Mental Model in Comprehending the Concept of Integer 11 29–44
[16] Sa’dijah C 2009 Asemen kinerja dalam pembelajaran matematika J. Pendidik. Inov. 4 92–5
[17] Goldin G A 1997 Observing Mathematical Problem Solving through Task-Based Interviews Merga 9 303–9
[18] Fauziyah N, Lant C Le, Budayasa I K and Juniati D 2019 Cognition Processes of Students with High Functioning Autism Spectrum Disorder in Solving Mathematical Problems Int. J. Instr. 12 457–78
[19] Patton L L 2002 Jewels of authority: women and textual tradition in Hindu India (Oxford University Press)
[20] Sujadi I 2008 Rekonstruksi Tingkat-Tingkat Berpikir Probabilistik Siswa Sekolah Menengah
Pertama Uns 187–208

[21] Wijaya A 2015 Kesalahan siswa dalam memilah data relevan pada soal matematika berbasis konteks Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015 461–9

[22] Rufiana I S, Sa‘jidah C, Subanji S, Susanto H and Abdur Rahman Asari 2018 Informal Statistical Reasoning of Students Taken Int. J. Insight Math. Teach. 01 130–40

[23] Zieffler A, Garfield J, delMas R and Reading C 2008 A Framework to Support Research on Informal Inferential Reasoning. Stat. Educ. Res. J. 7 40–58

[24] Abdillah A, Subanji S, Nusantara T and Susanto H 2018 Proses Berpikir Siswa dalam Menyelesaikan III Structured Problems Matematis PROSIDING Seminar Nasional Matematika dan Pembelajarannya 2017 ed A Qohar, S Sukoriyanto, I N Hidayah, S H Nasution, D Satyananda and Rustanto Rahardi (Malang: Jurusan Matematika FMIPA Universitas Negeri Malang)

[25] Nicetas G and Villarino B 2018 College Students ’ Conceptual Understanding of the Concepts of Force and Energy : Comparisons with Earlier Studies Int. J. Innov. Sci. Math. Educ. 26 22–61

[26] Irfan M, Sa‘jidah C, Ishartono N, Widodo S A, Rahman A A and Hudha M N 2019 Interference in Solving Mathematical Problems JCSI 2018, October 19-20, Yogyakarta, Indonesia pp 1–10

[27] Giacomone B, Beltrán-Pellicer P and Godino J D 2019 Cognitive analysis on prospective mathematics teachers’ reasoning using area and tree diagrams Int. J. Innov. Sci. Math. Educ. 27 18–32

[28] Irfan M, Nusantara T, Subanji S and Sisworo S 2019 Direct Proportion Or Inverse Proportion ? The Occurrence Of Student Thinking Interference Int. J. Sci. Technol. Res. 8 587–90

[29] Hegarty M and Kozhevnikov M 1999 Types of visual-spatial representations and mathematical problem solving. J. Educ. Psychol. 91 684–9

[30] Kirschner P A, Sweller J and Clark R E 2006 Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching Educ. Psychol. 41 75–86