

THE EMERGENCE OF PROBABILISTIC LANGUAGE FROM STUDENTS' UNCERTAINTY IN LEARNING INFORMAL INFERENCE REASONING

Lukman Fakhmi, Hening Windria, And Mika Meitriana Manurung

STKIP Surya

Tangerang, Indonesia

e-mail: lukman.fakhmi@students.stkipsurya.ac.id

Abstract. In our daily life, we often draw conclusion regarding some phenomenon to get new information from sample of the population. By doing so, we unconsciously learn the inferential statistics informally or known as Informal Inferential Reasoning (IIR). Therefore, it can be said that IIR is close to our life. As a result, it is wise to use IIR in introducing inferential statistics to students. In introducing IIR, a set of learning design needs to be made. In this research, the design would use Realistic Mathematics Education (RME) as the approach of the learning and dot plot as a visual representation. This paper describes the probabilistic language students used in expressing some uncertainty when making generalization. The method used is design research. The involved subjects in this research were six students of VII-8 at SMPN 7 Tangerang Selatan. The results showed that the probabilistic language appears when students generalize the data. Moreover, probabilistic language also appears when students added samples in growing sample activity. Probabilistic language which appeared in this research are "mungkin", "hampir sama", "kurang sama", and so on.

Keywords : *design research; informal inferential reasoning; probabilistic language; uncertainty.*

Statistical inference is a part of statistics that aims to infer those conclusions based on evidence which provided by sample to the population [1]. We underline the definition of statistics inference is generalize a data based on a sample, not from the population. So, it shows that exist an uncertain to infer those conclusions. It was supported too by [1], [2] and [3] that one of ingredients of statistical inference have a degree of certainty. However, to appear the degree of certainty is difficult for children.

The uncertainty will not be appeared by children more formally using confidence interval of p -value [4]. On the other hand, it should be noted that the inferential statistics need to be learned early grade by students because drawing conclusions is needed by students in an introductory statistics course [5]. Therefore, we have to research deeper about inferential statistics teaching at schools.

Informal Inferential Reasoning (IIR) is one of the hot topics which is discussed by researchers in order to change the aims in learning statistics education especially inferential statistics. The term of informal is only as starting point, because to develop the formal understanding is needed informal knowledge as a starting point [5]. IIR consists of three features [3], those are:

- (1) Making a generalization (like a predictions, parameter estimates, and conclusion that goes beyond the data)
- (2) Data as evidence for those generalizations
- (3) Using probabilistic language for stating the uncertainty which has been made in describing the generalization.

However, this research focuses on the emergence of the probabilistic language from students' uncertainty in learning IIR.

The probabilistic language would show if exist a set of learning. So, we need an approach/ method/ model of learning toward the goal. In this research, the instructional material designed using *Realistic Mathematics Education* (RME). It refers to the approach of mathematics education that have initiated by Freudental to reform the way mathematics in Netherland [6]. One reason of choosing RME, because there is a progressive mathematizing which is a difference with the others. Progressive mathematizing in RME is a mathematization process from a phenomenon[7].

In *Realistic Mathematics Education* (RME), there are several characteristics that we need. One of them is using model to mathematize process. Using the "model" does not lead to props[7]. Then, he stated that the function of it is as a bridge from concrete level math knowledge to formal level math knowledge. Therefore, researchers used dot plot as a model in the instructional material.

Dot plot is a simple visual representation for quantitative information. This display is a two-dimensional graphical which is very flexible and easy to interpret. Researchers used a dot plot as a visual representation of the lesson plan because it

has already been succesfully carried out by Padmi in 2015 to depeloved IIR on the 7th grade students in Indonesia. Then, we can draw the dot by hand without a computer, and able to analyze the data easily. On the other hand, using dot can appear the probabilistic language of the student.

Based on the explanation above, we planned to look the probabilistic language which was appeared by the student. So, we had a question for this research, that is "What are the probabilistic languages that emerge from the students in learning Informal Inferential Reasoning?".

METHOD

In general, this research aims to look and know the emergence of probabilistic language from students' uncertainty in learning Informal Inferential Reasoning (IIR).

The method used is design research with a qualitative approach. Design research is one of many ways to develop local instruction[8]. Design research consists of cycles of three phases, they are: (1) preparation and design phase, (2) teaching experiment, and (3) retrospective analysis [9].

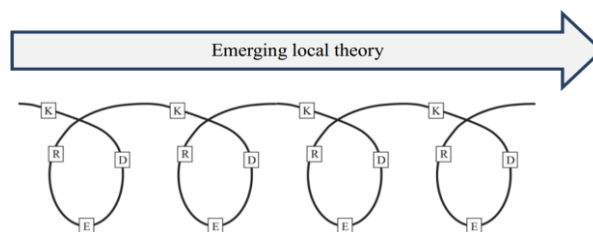


Figure 1. Cyclic Process Of Knowledge, Design, Experiment, Reflection, Reflection And (New) Knowledge

In preparation and design phase exists a Hypothetical Learning Trajectory (HLT) which has to be designed. Before it, researchers will conduct a literature review on learning Informal Inferential Reasoning (IIR) and looking for the relation of IIR towards curriculum 2013 (in Bahasa: K13). After we did it, there is a learning trajectory to researchers for making HLT (look at **Table 1**). HLT consist of three components, namely: (1) direction of learning goals, (2) learning activities, and (3) allegations of understanding or thinking of students in a learning context[9]. Beside the HLT, there are pretest, interview guides to teacher, and observation guides to class

TABLE 1. LEARNING TRAJECTORY

Meeting	The Learning Contexts and Activities of The Students
1	• Introduction about <i>dot plot</i> (using weight of people on learning as context) Context: <i>Duration of using smartphones in hours</i> • Concept of Sample • Concept of Population
2	Context: <i>Duration of using smartphones in hours</i> • Compare the dots • Growing sample and probabilistic language
3	Context: <i>Duration of using smartphones in hours (Are girls longer time using smartphone than boy?)</i> • Unfairly taking samples (take data from all students in the school) • Bias (randomly taking the data)
4	Context: <i>The ideal weight at the age of 13 years using the Body Mass Index (BMI)</i> • Making a generalizations • Data as evidence fot those generalizations • Using probabilistic language for the uncertain when was making those generalizations.

The second phase of this research is teaching experiment. This phase aims to test the HLT which we have designed. In learning process, HLT has to be the main guidelines. So, teacher has to act according to HLT.

The last phase is retrospective analysis. The teaching experiment will be analyzed by researchers to compare whether or no, all of the HLT which has been tested corresponding.

The cyclic process can be ocured repeatedly until we find the learning trajectory which we revised had been tested. However, cycles in this research used only one cycle where the teaching phase of an experiment consisting of four activities. The result of this cycle will emergence local theory.

FINDING AND DISCUSSION

Based on the teacher interview and observation class guidelines, the design of Hypothetical Learning Trajectory (HLT) was not needed a significant modification. Researchers only changed the technical teaching in the classroom like the arrangement of the U-shaped seat, determining member of the group randomly (the adjacent seat), vocabulary or the term in mathematics is an important part to be known by students, using K13 revision 2016 as

the instructional material, and using conventional method on learning. Furthermore, the teacher did not know about IIR and RME consequently we introduced it to her. On the other hand, researchers also selected 6 students to follow the research.

In general, the overview of the student's initial capability can be said to have reached the expected standard. The students election is based on middle ability and discussion with their mathematics teacher. Although, that way

picked the students, we kept asking them about their answers one by one. If there are mistake, we were helping to justify their answers.

The design of learning Informal Inferential Reasoning (IIR) here only to look the probabilistic language which is appeared by students. After we have known their prior knowledge about statistics through pretest and students' interview, teacher's interview, and observation the class, the next step will be analysis and discussion of the teaching experiment results to the research question.

(1) Meeting 1

First meeting consists of two activities. For early apperception, researchers just introduce dot plot in advance. After the students understand the dot plot, they would be formed into three groups and taught the concept of the sample (the first activity) and the population (the second activity). However, in this section we will discuss the first activity only because there are

Probabilistic languages from students' uncertainty. The context would using "*duration of using smartphones in hours*". Of course, this research has to need data. So, a few days before, researchers had been asking data about students' duration in using smartphones in the class.

In the first activity, we learned about concept of sample. Researchers issued an online newspaper entitled "Kebiasaan orang Indonesia, Pelototi 'Smartphone' Selama 5,5 Jam Sehari" and were sharing the students' worksheet 1.1 to each groups. In english we can said "Habits of Indonesia, Looked at 'Smartphone' for 5,5 Hours a Day". They were unconsciously acts like a sample, whether or no themselves like the online newspaper. Therefore, they would investigate the truth using their data class but only taken a little data and using dot plot as visual representation of the data. The following figure is one of the dot plot created by the students (see Figure 2).

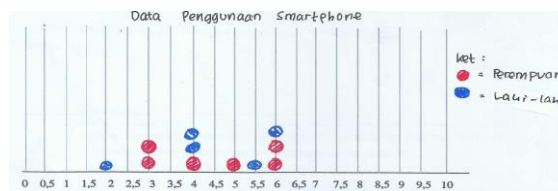


Figure 2. Dot plot of group 3

Students' worksheet only consist of 4 questions. The last question asked students to estimate or predict whether the information obtained is appropriate or not the actual information of data class. From the students' worksheet, they indicated an uncertainty when asked to predict a conclusion. The following are answers to each of the groups regarding the probabilistic language students show about uncertainty.

4. Prediksi atau perkiraanlah tentang grafik yang kalian buat apakah akan sama seperti grafik dari seisi kelas? Kemukakan alasanmu.
Jawab: Dari grafik di atas dan grafik seisi kelas jika dibandingkan Pasti ada yang Sama dan ada yang berbeda, karena tidak semua pengguna smartphone memiliki waktu yang sama.

Figure 3. The Answer of group 1

Jawab: Hampir, karena tidak ada siswa kelas yang menggunakan smartphon dalam waktu tersingkat yaitu 0 jam / sama saja seperti tidak bermain atau menggunakan smartphone

Figure 4. The Answer of group 2

Jawab: mungkin karena kami hanya mendata 11 orang anak, dan setiap anak berbeda-beda waktu penggunaan smartphone. Jadi akan berbeda grafiknya antara 11 orang anak dengan selsi kelas.

Figure 5. The Answer of group 3

The group 1 (see figure 3) stated that if their dot graph compared with data class (dot) graph definitely there are same and different (hour or information) because everyone has not the same duration when was using the smartphone. Based on their answer, they are implicit information which is a form of doubt that the information from the sample would be equal to the population. Besides this, there are different answers to the group 2 and 3. For the group 2, they stated “hampir ...” or in english “close ...” or “almost ...”. Then, for group 3, they stated “mungkin ...” or in english “may be ...”. From their answers we can conclude that when they asked to make those conclusions, they showed a probabilistic language for the uncertainty from the sample they made.

(2) Meeting 2

In the second meeting, they would move a little bit to the inferential statistics. They would learn about comparing the dots (the first activity) and growing sample based on one of the dot which has chosen by them would appear the probabilistic language (the second activity).

The first activity, the dots of each groups would put in the white board. Then, teacher shared students' worksheet (worksheet 2.1) to each groups. There are two questions in the worksheet. The first, they asked to write the similarities and differences of the dots but we do not discuss about this one. The last, they would choose which one the best represents the dot for the graph class. The following are the students' worksheet about the dot which represents.

Grafik yang paling menyerupai grafik kelas adalah grafik 2.
 Karena grafik tersebut Setiap angka atau waktu memiliki jumlah yang hampir sama dengan grafik kelas.

Figure 6. The Answer of group 1

Grafik yang paling menyerupai grafik kelas adalah grafik 2.
 Karena grafik tersebut Memiliki 1 orang yang tidak menggunakan smartphone di kelas, ada yang tidak mempunyai smartphone. Jadi bisa dibilang mirip dengan grafik kelas.

Figure 7. The Answer of group 2

Grafik yang paling menyerupai grafik kelas adalah grafik 2.
 Karena grafik tersebut memiliki kemiripan lebih banyak dengan grafik kelas. Alasan nya kebanyakan anak perempuan tidak memiliki smartphone dan anak laki-laki nya memiliki smartphone semua, dan memiliki nilai tertinggi yaitu 4 jam (salah satu).

Figure 8. The Answer of group 3

From the question about the represent graph, group 1 and 3 have a same answer namely dot plot of group 2. The group 1 stated that every duration (using smartphone) “almost the same” with the dot plot of their class. Then, the group 3 stated that the dot plot of group 2 to the most represent because the dot “has similarities more” with the dot plot of their class. However, the different answer with the group 2. They answered the the dot plot of group 1 “it can be said to be similiar” with the dot plot of their class. From their answers, researchers look them seem doubtful and show the uncertainty on stated a infomation from sample to draw the population.

After they have done the first activity, they would choose which one of the dot to add

the sample (growing sample). Finally, they agreed that the dot of graph 2 which would be added the data as much as 7 samples. Then, the other dot plot released by the teacher. After that, the teacher was back to given students’ worksheet (worksheet 2.2) to each groups. There are two questions in the students’ worksheet.

The first question is about the information of the dot has chosen by them whether or no same with after added the sample. In this question, they would add the sample one by one in alternately and directly noted the changes. However, based on their answered about probabilistic language, only group 3 stated about it. The following is the probabilistic language of group 3 (see Figure 9)

Jawab: Masih sama karena kemungkinan ada sekitar 3 orang yang memakai smartphone 6 jam.
 > Masih sama karena nilai tertinggi masih menang. Pagi 4 jam
 > Kurang sama karena kebanyakan perempuan yang ada di 4 jam di grafik kemarin lebih banyak laki-laki
 > Kurang sama karena di 2 jam kebanyakan anak perempuan di grafik kemarin
 > Kurang karena di 5,5 jam hanya ada 1 orang anak laki-laki
 > Kurang karena di 3 jam melebihi 2 orang
 > Kurang karena hanya ada 2 orang anak di 2 jam di grafik kemarin

Figure 9. The Probabilistic language of group 3

They were showing that if the dot added some data, it would change the information. Some probabilistic languages which were in their worksheet namely “masih sama ...” or “still the same ...”, “kurang sama ...” or “less the same ...” and “hampir ...” or “almost ...”. The other groups did not show the probabilistic language when added the data to the dot plot. They were only stating that different information when the growing sample happened.

The second questions have asked to student whether or no after the data added to the dot would same with the dot graph class. Moreover, they have asked to conclude it. The following are their answer about it.

Jawab: Hampir sama karena pada grafik kelas, waktu yg paling banyak digunakan adalah 4 jam, dan yg paling lama adalah 7 jam yang ada pada grafik baru yg ditam bahkan.

Figure 10. The Conclusion of group 1 after adding sample

Jawab: Hampir, karena tidak semua grafik yg baru ditambahkan sama seperti grafik kelas yg kemarin kita buat

Figure 11. The conclusion of group 2 after adding sample

Jawab: mungkin hampir sama, karena banyak kemiripan yang ada di grafik tersebut, seperti nilai tertinggi, nilai yang paling lama menggunakan Smartphone (7 jam), dan nilai penggunaan Smartphone (kebanyakan perempuan tidak memakai smartphone dan anak laki-laki semua memakai smartphone)

Figure 12. The conclusion of group 3 after adding sample

Based on their answers, they were showing that exist a uncertainty when they conclude a (growing) sample to the population. The probabilistic appeared like “hampir ...” or “almost ...”, “hampir sama ...” or “almost the same ...”, and “mungkin hampir sama ...” or “my be almost the same ...”. Moreover, there is a conversation between teacher and students. The following are the conversation 1.

Conversation 1	
Teacher	: According to you, how do the conclusions if the data that you have added one by one but adding it was not all, how about the information?
Student 1 and 2	: Getting different information.
All students	: It would be almost a whole.

From the conversation, it seems researchers look that the students have known if the effect of sample size would resemble the whole data (population).

(3) Meeting 3

In this section, they would learn deeper about inferential statistics. The third meeting consisted of two activities. The aim of the first activity is to look how the students take a data (sample) from the smartphone user at SMPN 7 Tangerang Selatan. So, they would be faced with a large-sized data. Whereas, in the second activity they would investigate claims that the girls time inusing smartphones longer than the boys at the school and to look the students understand that taking the data randomly is allowed. However, here we were only discuss

about the first activity because in the second activity, we did not find the probabilistic language from the learning.

In the first activity, the students seemed very strive and difficult to put the data into their dot plot because they only given 5 minutes to put it into their dot scale. They were unconscious with their act that may to select the sample. However, in this section we also looked the probabilistic that was appeared by the students. There were some conversations between teacher and the students that showed the probabilistic language to state the uncertainty

Conversation 2

Teacher	: From the dot which you have created it, did it already represents a grapo of the school?
Student2	: Not represent.
Student 1	: Not half.
Teacher	: According to you, is it already actually resembles graph school or not?
Student 1	: Not yet, because not all.
Teacher	: If like that, is it not yet or nearly resemble?
Student 1	: I think almost.
Student 2	: Eeehh...
Student 1 & 2	: Yet, it should half before.
Teacher	: If it exceeds the half, did the graph resembles a graph class?
Student 2	: Getting resemble.
Teacher	: If less than a half?
Student 2	: Almost resemble.
Student 1	: Less resemble.

From their conversation, it seems they were still doubtful to state that their dot would same with the dot of the school. Based on their activity, researchers conclude that the students were able to give opinions about the bias or predictions of the data collected. Moreover, they were using the probabilistic language (uncertainty).

(4) Meeting 4

This meeting was the last meeting of the sequences of Informal Inferential Reasoning (IIR) learning. The activity in this section only discuss about their homework about three princples about IIR which has been explained in the chapter of introduction. They would present the results of their findings in front of the class about the ideal weight using the Body Mass Index (BMI) to the students aged 13 years old.

The students recognized the difficulties to find data. The first, they stated that they did

not have enough time to find and process the data. It might be happened because it looked for students whom were 13 years old. The second, they needed a different data between the boys and the girls. Then, they also need to find the height and weight of them. Then, calculate it used the formula of BMI. In fact, those were not all of the obstacles, but also making a decimal one digit behind the comma. However, they finished their homework although facing some obstacles.

Based on their worksheet, there was no probabilistic language when they were concluding their finding on the ending of their question in worksheet. Moreover, when they were asked explain their worksheet in front of the class, they seemed not enthusiastic to conduct discussions. So, the teacher took the initiative to encourage discussion to appear the probabilistic language from the students. The following were our conversation.

Conversation 3

Teacher	: If anyone asked, does the girls in this school has the ideal body? How do you think?
Students 3 and 4 (chuckle)	

<i>Student 3</i>	: <i>Not necessarily the ideal of all.</i>
<i>Teacher</i>	: <i>What is the evidence you show?</i>
<i>Student 4</i>	: <i>Using the dot and data.</i>

From the conversation, actually they unconsciously knew that they stated a probabilistic language to the uncertainty from sample to the population although it is not the formal way. In this meeting, the probabilistic language was not overstated because this meeting was to look what the students have been found during they did the mini-research.

CONCLUSION

This research is a design research which only to find the emergence of probabilistic language from students' uncertainty in learning Informal Inferential Reasoning. On the learning side, we can conclude that probabilistic language can appear from the students like "hampir ..." or "almost ...", "mungkin ..." or "may be ...", "mungkin hampir sama ..." or "almost the same ...", "bisa dibilang mirip" or "it can be said to be similar", "masih sama ..." or "still the same ...", "kurang sama ..." or "less the same ..." etc. It was appeared when they asked to make inference and the growing sample happened. Although, they expressed a doubt that is still uncertain but it is not wrong for their express because it is a processor bridge to learn a degree of uncertainty from informal to formal. At least, the students knew that to inference a sample to the population there is a degree of uncertainty or commonly called probabilistic language.

REFERENCES

- [1] D. S. Moore and W. I. Notz, *Statistics: Concepts and Controversies Seventh Edition*, New York: W. H. Freeman and Company, 2009.
- [2] A. Bakker, P. Kent, J. Derry, R. Noss and C. Hoyles, "Statistical Inference at Work: Statistical Process Control as an Example," *Statistics Education Research Journal*, pp. 130-145, November 2008.
- [3] K. Makar and A. Rubin, "A Framework for Thinking about Informal Statistical Inference," *Statistics Education Research Journal*, vol. 8, no. 1, pp. 85-105, Mei 2009.
- [4] K. Makar, A. Bakker and D. Ben-Zvi, "The Reasoning Behind Informal Statistical Inference," *Mathematical Thinking and Learning*, vol. 13, no. 1-2, pp. 152-173, 2011.
- [5] A. Zieffler, J. Garfield, R. Delmas and R. Chris, "A Framework to Support Research on Informal Inferential Reasoning," *Statistics Education Research Journal*, vol. 7, no. 2, pp. 40-58, 2008.
- [6] K. Gravemeijer, *Developing Realistic Mathematics Education*, CD-[beta] Press, 1994.
- [7] A. Wijaya, *Pendidikan Matematika Realistik: Suatu Alternatif Pendekatan*, Yogyakarta: Graha Ilmu, 2012.
- [8] I. G. A. R. S. Padmi, "Developing 7th Grade Students' Informal Inferential Reasoning," *Mathematics Education Research Program, Postgraduate Program of Surabaya State University*, 2015.
- [9] D. van Eerde, "Design Research: Looking into The Heart of Mathematics Education," in *Proceeding The First South East Asia Design/Development Research (SEA-DR) International Conference*, Palembang, 2013.
- [10] M. A. Simon, "Reconstructing Mathematics Pedagogy from a Constructivist Perspective," *Journal for Research in Mathematics Education*, vol. 26, no. 2, pp. 114-145, 1995.
- [11] A. Bakker, "Design Research in Statistics Education: On Symbolizing and Computer Tools," 2004.