Exploring informal inferential reasoning: the case of comparing two data sets problem

Exploring informal inferential reasoning: the case of comparing two data sets problem

P Andriani[1,2,*], C Sa'dijah[2], Subanji[2] and H Susanto[2]

[1] Program Studi Tadris Matematika, Universitas Islam Negeri Mataram, Jalan Gajah Mada No. 100, Mataram 83127, Indonesia
[2] Program Pascasarjana, Universitas Negeri Malang, Jalan Semarang No. 5, Malang 65141, Indonesia.

*parhaini.andriani@uinmataram.ac.id

Abstract. Informal inferential reasoning as part of statistical reasoning is extremely important in learning. However, this reasoning has not been widely studied in Indonesia. This article explores informal inferential reasoning of students in solving comparing two data sets problem. This qualitative research was involved 118 junior high school students in Mataram West Nusa Tenggara. Instrument in this study is Assessment of Informal Inferential Reasoning (AIIR) which was modified and validated by mathematics and mathematics education experts. Student responses are classified based hierarchical cognitive framework. The findings showed that the existing levels was still relevant for the students of junior high school in Mataram. However, there were subjects which could not be classified in the levelling. The characteristic reasoning of the subjects are tends to ignore the variability and distribution aspects of the data in making predictions, but they can indicate significant value differences between two groups of data. They encountered an error in certain aspects, but they are able to produce an almost right conclusions. This study recommends the need for a new levelling to assess informal inferential reasoning of junior high school students by considering the ability of local and global view of data.

1. Introduction

The uncertain of complexity in the world requires schools to prepare their students to be flexible thinkers, lifelong learners [1]; and to be competent in creating and evaluating data-based claims critically [2]. Such competence can be developed through statistical reasoning. This is in line with the objectives of mathematics learning in the Curriculum 2013 of Indonesia, which requires students to be able to use reasoning in patterns and properties, perform mathematical manipulations in generalizing, compiling evidence, or explaining mathematical ideas and statements [3]. The National Council of Teachers of Mathematics (NCTM) also established statistical reasoning as one of the skills that needs to be developed at the school level [4].

Informal inferential reasoning is a part of statistical reasoning which is frequently analyzed in recent statistic education studies [5]. This logical reasoning is important for students at any level because it can help their ability in predicting and deciding which data they can use in daily life [6]. This reasoning is able to develop students' ability in understanding formal statistical inference [6-8]. However, this reasoning is still not much analyzed both at schools or universities in Indonesia.

There are various definitions of informal inferential reasoning. Zieffler et al defines this reasoning as "the way students use their informal statistical knowledge to make arguments in favor of conclusions

about unknown populations based on observed samples"[7]. Another definition expresses informal inferential reasoning as a cognitive activity that includes generalizing of population based on data (samples) informally [8]. A more general view is made by Jacob which expresses informal inferential reasoning as the statistical reasoning of students used when completing the task of informal statistical inference [9]. Thus, it can be understood that informal inferential reasoning is defined as probabilistic generalization about population based on data (sample) by using informal knowledge (without formal statistical test), intuition, context, argument, data representation, and tools.

There are three types of tasks that can be used to develop informal inferential reasoning, ie: (1) estimate and draw a populations' graph based on a sample; (2) compare two or more sets of data; and (3) consider two opposing models [7]. These types of tasks then were developed by Goss [10] into Assessment of Informal Inferential Reasoning (AIIR) and are being used to assess informal inferential reasoning of junior high school students. In order to measure characteristic of students' informal inferential reasoning, Goss [10] has modified SOLO model of Biggs and Collin [11]. This hierarchical cognitive framework is used to develop informal inferential reasoning levels in junior high school students. The levels consist of: (1) Pre-Structural (No cycle); (2) Pre-IIR (Cycle-1); (3) Naïve-IIR (Cycle-2), and (4) Appropriate-IIR (Cycle-3). Each level is in form of learning cycle based on SOLO model which consists of Unstructured, Multi-structural, and Relational (U-M-R cycle). The characteristics of each level contain three fundamental aspects: use of variability, use of context, and certainty and argumentation (see table 1).

Table 1. Characteristics of the informal inferential reasoning levels developed by Goss [10].

Level	Characteristics of the reasoning
Pre-Structural (No-Cycle)	Modifying context to answer question
Pre-IIR (Cycle-1)	Not being able to understand the concept of variability, simply answering based on what is in context, using a deterministic language.
Naïve-IIR (Cycle-2)	Having a limited understanding about variability, using more context than data or otherwise, using probabilistic language that has little impact on conclusions.
Appropriate-IIR (Cycle-3)	Able to understand the concept of variability appropriately, considering context and data in formulating strategies and solutions, using probabilistic language with uncertainty expressions in solutions

In this study, type of task given is comparing two sets of data problem. Comparing two data sets is a type of task which is frequently used to measure informal inferential reasoning [9], [12], [13]. This type of task is important as a basic for the learners to understand formal statistical inference which generally taught at universities. Comparison between two or more groups of data is usually seen as formal inference which includes significance test, confidence interval, p-value, and other ideas in drawing conclusion about population based on the observed sample. According to Pfannkuch [12], there are four important aspects which are needed to compare some data groups which are: central tendency, comparison of difference of central tendency relative to variability, data distribution checking (normality assumption, outliers, clusters), and sample size effect.

Taking into account the importance of the informal inferential reasoning in statistical learning, hence the need for study of this reasoning at school level in Indonesia. Therefore, this study explores the informal inferential reasoning of junior high school students in solving the problem of comparing two groups of data.

2. Method

This study is a qualitative research with case study type. The study was conducted in June 2017 which involves 118 students of grade VIII (13 – 14 years old) at two state junior high schools and 1 Islamic junior high school in Mataram West Nusa Tenggara. This location has been chosen because these schools are the best junior high schools in Mataram.

Each participant was given a test in form of instrument which had been adapted from Assessment of Informal Inferential Reasoning (AIIR) which was developed by Goss [10]. This instrument was modified based on the context in Indonesia and validated by mathematics and mathematics education experts. The test was about comparing two data sets which was presented in form of dot plot. The test consisted of two parts which were related to testing the effectiveness of the use of textbooks by comparing two groups of data about the score of students who were given textbook and who were not.

At the first part, the participants were asked to predict two data which showed that the scores of those who were given textbook were better than the scores of those who were not. At the second part, the participants were given three pairs of data where each pair of data (class) showed a comparison of scores of those who were given text book and those who were not. Then, the participants were asked to make ranks on each pair of data. Class with rank 1 showed strong evidence that the students who were given textbook get better test scores than those who were not and class with rank 3 showed weak evidence that students who were given textbook get the better test scores. Then the students were asked to explain the reasoning of their answers.

The responses of the participants were then analyzed based on hierarchy cognitive design which the informal inferential reasoning has been developed by Goss [10]. This description of indicator can be made as standard to classify the participants' ability after given test. Data triangulation was done with a task-based interview to clarify the participants' answers on the answer sheets.

3. Results and discussion

Based on the analysis on the answer given by the students, it can be seen that there were 17 students (14%) on pre-structural level, 50 students (42%) on pre-IIR level, 44 students (37%) on Naïve-IIR level and 3 people (3%) on appropriate-IIR level. Four of them are difficult be classified in the levelling. Some forms of students' reasoning are explained as follows.

3.1. Students' reasoning on pre structural level

Subject 1 (S1) which is on pre structural level makes wrong data prediction by determining that the students who were not given books were better than those who were. It is obvious that S1 was wrong to understand the question by drawing two groups of data which the distribution was almost the same. S1 made data which stated that the students who were not given books were better than who were. S1ordered data pairs based on criteria "better", "more good scores", "the least of good scores" without making clearer standard towards the criteria. The generalization given to the subject on pre structural level was still too weak and there was no indication of understanding towards variability or data context.

3.2. Students' reasoning on pre-IIR level

The subject which has characteristic on pre-IIR level was shown by subject 2 (S2). The subject made data prediction with a vary data on the group who were given book and the group who were not. S2 ordered the evidence based on reason of "I only see, I did not count the mean, at first I saw everything on each pair of data, in my opinion the most stable among the others are this and that (pointing data of class B), even though there is difference but it is not so far". When the subject was asked about the meaning of "stable", the subject answered, "the grade of those who got books and those who did not is not really different".

The students who were in Pre-IIR level did not consider variability aspect in ordering the data pairs. The data prediction they made had small range and they are same on the two data groups. If the data varied, the outlier was still being considered. However, in determining strong evidence in data group comparison, data model with small distribution and which was extremely different was needed. The

students too much used context and it had not made an appropriate conclusion. The students also still used deterministic language without considering the aspect of certainty in the data.

3.3. Students' reasoning on Naïve-IIR level

Subject 3 (S3) which was on Naïve-IIR level predicted data between the students who were given books and who were not by using more context than the data. S3 had understood the variation in the data. However, the explanation was not sufficient because the subject merely looked at the context by stating "It depends on the interest of each student, it depends whether they read the book or not". On part 2 test, the subject only looked on the data of the students who were given books and then they were compared between the classes without considering the data of the students who were not given the books. The result was an inaccurate conclusion.

The other reasoning was given by subject 4 (S4) who predicted data with equal distribution. The data were made with a range of 0 to 100. Only some data which had a different frequency. On second part of the test, S4 looked to consider more the context which made the data pair was almost the same. S4 also used probabilistic language by using the word "maybe" although rare. The strategy which was used by S4 when ordering data pair was to consider a standard which is known as minimum passing criteria score Kriteria Ketuntasan Minimal (KKM) which was 80. S4 compared between the students who got score above minimum passing criteria or exactly the minimum passing criteria with the students who were given book and those who were not. The comparison were then made as standard in ordering strong or weak evidence.

Subject reasoning on Naïve-IIR level involves simple calculation with conclusion which was reasonable enough although not accurate yet. The use of probabilistic language also had been seen although still limited such as the word "maybe" and "almost". The context use was more dominant compared to considering data as the evidence standard.

3.4. Students' reasoning on appropriate –IIR level

The subject's reasoning on this level was adequate in making conclusion about data as shown by subject 5 (S5). When predicting the possible data in the given context, S5 was able to argument by considering aspect of variability and data distribution. When ordering three pairs of data, S5 used strategy of counting gain score on the mean of each group and then ordered them. The reasoning given by S5 indicated the ability to think aggregately. S5 had started to give meaning the data holistically by leaning on gain score standard.

The other reasoning of the subject at appropriate-IIR level also shown by subject 6 (S6). When predicting two groups of data, S6 made data with different range, giving attention to data which did not overlap and considering the normality of data distribution. When ordering the data group, S8 used the minimum passing criteria as standard score to decide the number of students who got a good score.

Subject with criteria at Appropriate-IIR level gave explanation which showed that the subject had enough comprehension of variability and context. Probabilistic language was used accurately. The important thing in comparing some groups of data was the existing overlap data in range and distribution. The subject had this ability at Appropriate-IIR level.

Generally, the existing leveling has developed by Goss [10] was relevant to measure the ability of informal inferential reasoning of junior high school students in Mataram. However, there were some students who were difficult to classify in the levels. One of the example was S7−despite the subject's less understanding about variability−could give an appropriate conclusion about the result of three pairs of data. S7 could not be classified on level 4 appropriate-IIR but the subject's reasoning was too high to be classified into level 3 naïve-IIR.

Subject 7 (S7) predicted data absolutely between the students who were given books and those who were not as seen in figure 1. S7 chose data with small spread with two data which were very dichotomist. Data variability which was seen by S7 was not independent from the subject's understanding towards the context. When ordering the group data (on the second part of task), S7 counted the total of each data then compared between the students who were given books and those who were not in each class.

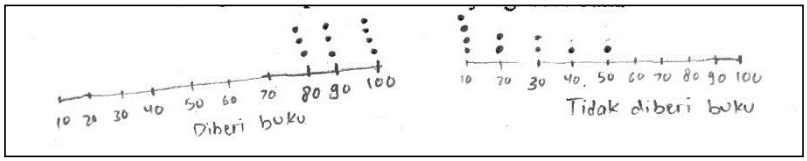

Figure 1. S7's reasoning in predicting the data (left is data of those who were given book, right is data of those who were not given books)

Besides S7, there was S8 (S8) who could not be classified in Naïve-IIR level or Appropriate-IIR level. When ordering the data group which had strong evidence, S8 used the gain score from the total of each data group. S8 gave an explanation such as seen in the interview script below.

Reseacher : How did you order the data in the second part of the question?
S8 : I did addition for all and after that it is subtracted. If I want to find stronger evidence which shows that the students who were given books will have better grade, then the gap should be higher. I mean the gain score of the total.
Reseacher : What is the meaning of the higher gap?
S8 : It means that it has stronger evidence that this is better. I did count the total of each data and the gap is found and then ordered.

The characteristic of reasoning owned by S7 and S8 can be said unique because although in predicting the data they tended to ignore the variability aspect and data distribution, they could show the grades gap which were significant on the two data groups made. In a certain aspect, they did mistake but they were able to draw a final conclusion which was almost accurate.

An important aspect to consider in solving the problem of comparison of several pairs of data is the comparison of intra and inter-group distribution of data. Because the comparison is done not only once, then the understanding of the data both locally and globally should be done by the students. . This kind of understanding is related with local and global view of data. For example, S8 who used global view in looking at the data by considering the gain score of the total value of each data group. Furthermore, S3 only compares student data given the book without considering other data pairs. S3 can be said to have used local view in looking at data.

Local understanding of data involves focusing on the individual value or some data in a data group which is presented in a table or a single point in graph. Global understanding refers to the ability to finding, recognizing, describing, and explaining the general pattern in a group of data which is changing all times (trend) through direct observation towards distribution or statistical technique [14]. Besides that, global data understanding also covers the ability to explain, compare, and predict based on data variability. Looking at graph globally is one of the ways to recognize patterns and general condition which are being analyzed.

Students need local understanding ability when looking the intra group of data distribution. Global understanding then is used to compare among the data groups and decide the order which shows strong evidence that those who were given books had better grades than those who were not. The patterns found in the comparison of data distribution are a gain score of mean, total gain score, and score which passed the minimum passing score criteria.

The switch from thinking locally to thinking globally depends on the attention development of the students. There are 5 ways to give attention to a situation which are (1) attention to global things; (2) attention to difference, differing aspect and differentiated aspect, detailed features, and attributes; (3) attention to relationship on intra parts between parts, the whole parts, between aspects, features, and differentiated attributes; (4) attention to relationship as characteristics where the objects are used by a person to consider and lead to generalization; and (5) attention to characteristic as abstraction of some objects which are formalized and stated independently and produce axiom where deduction can be made [15].

Students who can use data and context appropriately and also draw a right conclusion with probabilistic language may be classified in Appropriate-IIR level. However, if the understanding and the use of the concept of the variability are still less accurate, they cannot be classified at the level. That is why leveling is important to see the students locally and globally view of data.

4. Conclusion

In the school mathematics curriculum in Indonesia, informal inferential reasoning has not been taught. However, it can be done by junior high school students in Mataram. Reasoning abilities of the students mostly at the level 2 Pre-IIR although some are reaching level 4 Appropriate-IIR. But the results show that there are some students who cannot be inserted in any level. The students' reasoning cannot be included in level 3 Naïve-IIR or level 4 Appropriate-IIR. The characteristic reasoning of the subjects are tends to ignore the variability and distribution aspects of the data in making predictions, but they can indicate significant value differences between two groups of data.

Students need a growing attention in looking at the aspects considered in generating conclusions, when solving the problem of comparison of two groups of data, this is closely related to local and global thinking in looking at data. Therefore, this study recommends that the need for further study about the ability to view data locally and globally in the description of the existing levelling of informal inferential reasoning. A more comprehensive description will make it easier to assess students' reasoning. Besides that, this reasoning is also important to be included in the curriculum in Indonesia because this competency will be needed very much in facing the unpredictable world. The results of this study can be used as a reference by teachers to design learning based on the characteristics of students, especially in the learning of statistics.

References

[1] Makar K and Rubin A 2009 A framework for thinking about informal statistical inference *Stat. Educ. Res. J.* **8** (1) pp 82-105
[2] Ben-Zvi D, Bakker A and Makar K, 2015 Learning to reason from samples *Educ. Stud. Math.* **88** (3) pp 291-303
[3] Permendikbud Nomor 24 Tahun 2016 tentang KI dan KD Kurikulum 2013
[4] Martin G 2009 Focus in high school mathematics *Reasoning and sense making* (Reston, VA: National Council of Teachers of Mathematics)
[5] Pratt D and Ainley J 2008 Introducing the special issue on informal inferential reasoning *Stat. Educ. Res. J.* **7** (2) pp 3-4
[6] Garfield J, Le L, Zieffler A and Ben-Zvi D March 2015 Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking *Educ. Stud. Math.* **88** (3) pp 327-342
[7] Zieffler A, Garfield J, Delmas R and Reading C 2008 A framework to support research on informal inferential reasoning. *Stat. Educ. Res. J.* **7** (2) pp 40-58
[8] Ben-Zvi D, Gil E and Apel N 2007 What is hidden beyond the data? Helping young students to reason and argue about some wider universe, in Reasoning About Informal Inferential Statistical Reasoning: *A Collection of Current Reasearch Studies* (UK: University of Warwick)
[9] Jacob B L 2013 *The development of introductory statistics students' informal inferential reasoning and its relationship to formal inferential reasoning* (PhD Thesis, US: Syracuse University)
[10] Goos J M 2014 *A Method for Assessing and Describing the IIR of Middle School Students* (PhD Thesis, US: Western Michigan University)
[11] Biggs J B and Collis K F 1991 Multimodal learning and quality of intelligent behavior, in Intelligence: *Reconceptualization and measurement* (New Jersey: Lawrence Erlbaum Associate Inc)
[12] Pfannkuch M 2006 Comparing box plot distributions: A teacher's reasoning *Stat. Educ. Res. J.* **5**

(2) pp 27-45
[13] Bakker A, Biehler R and Konold C 2004 Should young students learn about box plots *Curric. Dev. Stat. Educ. Int. Assoc. Stat. Educ.* pp 163-173
[14] Ben-Zvi D and Arcavi A 2001 Junior high school students' construction of global views of data and data representations *Educ. Stud. Math.* **45** (1-3) pp 35–65
[15] Mason J and Johnston-WilderS 2004 Fundamental Constructs in Mathematics Education Abingdon (UK: Taylor and Francis)