DEVELOPING 7TH GRADE STUDENTS' INFORMAL INFERENTIAL REASONING

I Gusti Ayu Russasmita Sri Padmi

137785076

STATE UNIVERSITY OF SURABAYA

POSTGRADUATE PROGRAMME

MATHEMATICS EDUCATION

2015

DEVELOPING 7TH GRADE STUDENTS' INFORMAL INFERENTIAL REASONING

MASTER THESIS

A Thesis submitted to

Surabaya State University Postgraduate Program

as a Partial Fulfillment of the Requirement for the Degree of

Master of Science in Mathematics Education Program

I Gusti Ayu Russasmita Sri Padmi

NIM 13775076

SURABAYA STATE UNIVERSITY

POSTGRADUATE PROGRAM

MATHEMATICS EDUCATION PROGRAM STUDY

2014

APPROVAL OF SUPERVISORS

Thesis by I Gusti Ayu Russasmita Sri Padmi, NIM 137785076, with the title *Developing 7th Grade Students' Informal Inferential Reasoning* has been qualified and approved to be tested.

Supervisor I,

Date,

Prof. Dr. Siti M. Amin, M.Pd.

24 Juni 2015

Supervisor II,

Date,

Dr. Agung Lukito, M.S.

23 - 6 - 2015

Acknowledged by,

Head of the Mathematics Education Programme

Dr. Agung Lukito, M.S.

NIP 196201041991031002

APPROVAL

This two-year worth of work is a tribute to all the hoes who did it.

"Someone once told me the definition of hell: it's your last day on earth and the person you become, meet the person you ***<u>could have</u>*** *become."*

-anonymous

ABSTRACT

Sri Padmi, I G. A. R. 2015. *Developing 7th Grade Students' Informal Inferential Reasoning.* Thesis, Mathematics Education Study Program, Postgraduate Program of Surabaya State University. Supervisors: (I) Agung Lukito, M.S., and (II) Prof. Dr. Siti M. Amin, M.Pd.

***Keywords**:* Informal Inferential Reasoning, Realistic Mathematics Education (RME), Pendidikan Matematika Realistik Indonesia (PMRI), Design Research.

Inferential statistics has been proven to be a hard concept for students to understand. On the other hand, as the future citizen of modern society, statistics becomes important for the students to deal with data-laden information they encounter on daily basis. Informal Inferential Reasoning is a current topic in international statistics education proposed to bridge these two problems. The present study is focused on developing 7th grade students' informal inferential reasoning in Indonesia. It is based on Realistic Mathematics Education (RME) also known as Pendidikan Matematika Realistik Indonesia (PMRI). The goal is to contribute to instructional materials in developing IIR as well as accompanying theory on how this material works. The research question of this study is, *how can 7th grade students be supported to develop IIR?* We used design research as the methodology. We conducted two cycles of teaching experiments whose subjects were the 7th grade the students and one teacher in Laboratory Secondary School of State University of Surabaya. The first cycle consisted of 6 students, while the second cycle consisted of 27 students and one teacher. The data were collected through video recordings of the teaching experiment and students' written works.

The data then analyzed by comparing and contrasting the Hypothetical Learning Trajectory to the Actual Learning Trajectory. As the result of the analysis, six lessons on developing IIR were designed; (1) Scout staff; (2) Comparing the Dots; (3) Are girls taller than boys?; and (4) Social Media Addiction.

ABSTRAK

Sri Padmi, I G. A. R. 2015. *Developing 7th Grade Students' Informal Inferential Reasoning*. Thesis, Mathematics Education Study Program, Postgraduate Program of Surabaya State University. Supervisors: (I) Dr. Agung Lukito, M.S., and (II) Prof. Dr. Siti M. Amin, M.Pd.

***Kata Kunci**:* Informal Inferential Reasoning, Realistic Mathematics Education (RME), Pendidikan Matematika Realistik Indonesia (PMRI), Design Research.

Statistik inferensial merupakan konsep yang sulit dipahami oleh siswa. Di sisi lain, sebagai calon anggota masyarakat modern, statistik merupakan alat yang penting bagi siswa untuk menghadapi informasi sarat-data yang mereka temui di kehidupan sehari-hari. *Informal Inferential Reasoning* merupakan topik yang kini dibahas di dunia pendidikan statistik internasional untuk menjembatani dua masalah tersebut. Penelitian ini bertujuan untuk mengembangkan *Informal Inferential Reasoning* siswa kelas 7 di Indonesia. Pendekatan pembelajaran yang digunakan adalah Pendidikan. Pendidikan Matematika Realistik Indonesia (PMRI). Tujuan dari penelitian ini adalah mengkontribusikan materi pengajaran dalam mengembangkan IIR. Pertanyaan penelitian dari penelian ini adalah, *bagamana siswa kelas 7 bisa didukung untuk mengembangkan IIR?* Design research digunakan sebagai pendekatan dalam penelitian. Penelitian ini terdiri dari 2 siklus, dengan subjek siswa kelas 7 dan seorang guru di SMP Lab Universitas Negeri Surabaya. Peserta dari siklus pertama adalah 6 orang siswa, sedangkan peserta siklus kedua adalah 27 siswa dan seorang guru. Data dikumpulkan melalui rekaman video dan hasil kerja siswa. Data tersebut kemudian dianalisis dengan membandingkan Hypothetical Learning Trajectory dan Actual Learning Trajectory. Design research ini menghasilkan 4 pembelajaran dalam mengembangkan IIR: (1) Tongkat Pramuka, (2) Bandingkan Titik-titiknya; (3) Apakah anak perempuan lebih tinggi dari anak laki-laki?; dan (4) Kecanduan Media Sosial.

PREFACE

Om Swastyastu,

My utmost gratitude I deliver to Brahman, the Source of Light and the Magnificent Creator, without whom none of these; the ability to explore the world, the opportunities to meet new people, and the chance to challenge myself academically, would ever happened to my silly little self.

Om Avignam Astu Namo Siddham,
Om Siddhirastu Tat Astu Svaha.

All of us strive for perfection but none of us would ever come close to. I tempted to do best, but I realize there are still so many shortcomings that come along with this thesis. Suggestions and constructive criticism will be very much appreciated.

A lot of people are behind this two year-worth of work. Therefore, my highest appreciation and a bunch of thanks also goes to:

1. Prof. Dr. Siti M. Amin, M.Pd. and Dr. Agung Lukito, M.S; my supervisors in State University of Surabaya.
2. Monica Wijers and Dolly van Eerde, my supervisors in Utrecht University.
3. Dr. Agung Lukito. MS. and Maarten Dolk as the coordinator of the International Master Program on Mathematics Education (IMPoME) in State University of Surabaya and Utrecht University.
4. PMRI Center Board for the opportunity given to me as one of the grantees of the International Master Program on Mathematics Education.

5. All the lecturers and staff of Postgraduate Program, State University of Surabaya.
6. All the lecturers and Staff of Freudhental Institute, Utrecht University. Especially Mark Uwland, without whom we would never survive our first month in the freezing wind of Netherlands.
7. Dikti and Nuffic Neso for the financial support.
8. Ira Dwi Agustina, S.Pd., 7^{th} grade mathematics teacher of Laboratory Secondary School of State University of Surabaya.
9. Aan Hendroanto; the cameraman, the consultant, the observer, and the merciless critic behind this study.
10. All the teachers, the students, and staff of Laboratory Secondary School of State University of Surabaya.
11. My parents and my little sisters.
12. Yosep Dwi Kristanto, my favorite person. It is an honor to be a part of your life.
13. The students of Batch V IMPoME and BIMPoME.

Finally, I pray for the blessings of the Lord upon us, and for this thesis to fulfill its purpose.

Om Santi Santi Santi Om

Surabaya, June 2015

I G. A. Russasmita Sri Padmi

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Research Background

Statistical procedures can be divided into two major categories; inferential and descriptive (Hon, 2010). Descriptive statistics revolves around describing and summarizing the data at hand, while inferential statistics attempts to use said data and what is known of it to infer about a larger population containing it. In most countries, statistics education introduced in primary and secondary school traditionally consist of descriptive statistics, focusing on computation about averages and interpreting graphs (Makar & Rubin, 2009). Inferential statistics is not introduced before tertiary level (Makar & Rubin, 2009; Rossman, 2008).

In recent years, the idea to reform the aim of statistics education has brought a lot of discussion. Major stakeholders in statistics education have consensus about the need of statistical literacy for an individual to function efficiently in modern, information-laden society (Ben-Zvi & Garfield, 2004). Inevitably, the way statistics being taught in school has to be reformed as well, from focusing on computation and procedures to a more holistic approach that goes beyond data analyzing technique (Paparistodemou & Meletiou-Mavrotheris, 2008) and assists students to make sense of the world, which is the essential purpose of statistics.

Amidst this notion of reforming the aim of statistics education, emerged the idea of Informal Inferential Reasoning (IIR). Although no universal consensus has been reached about it, views from various researchers generally agree about it being the fundamental ideas that underpin the understanding of classical statistical inference. It implies that instead of making inference of data using formula, students are encouraged to use informal statistical knowledge that surely is more familiar to them to make inferences about unknown populations based on observed samples (Zieffler, Garfield, Delmas, & Reading, 2008) .

The notion to teach inferential statistics in early stages of statistics education stems from the fact that formal inferential statistics is proven to be a hard concept for students to understand. This difficulty is mainly due to students' lack of experience in reasoning about it, hence the inability to make sense and derive meaning of all the computation and procedures (Zieffler et al., 2008). It might be better if inferential statistics is introduced early in the stage of statistics education by focusing on the informal aspect of it (Zieffler, et al., 2008; Makar & Rubin, 2009), since the mathematics comprising formal inferential statistics is beyond reach of young students.

A lot of studies have showed the succes of developing IIR in young children. Ben-Zvi, Gil and Apple (2007) employed data handling software *TinkerPlots* to support the development of IIR in grade 4-6. Bakker & Gravemeijer (2004) used graphical representation software *Minitools* to support high school students informal reasoning about distribution. Watson & Moritz (1998) introduced IIR through the activity of comparing two data sets for children in grade 3-9, while Pfannkuch (2011) developed similar idea in 15-year-old-students through the investigation of boxplots.

Since IIR encourages students to employ their informal knowledge, problems with meaningful context and activities will support their reasoning. A lot of researches have focused a lot on the use of software, while in most countries this context are probably not suitable due to the lack of facility and infrastructure. In Indonesia, specifically, generally public schools are not equipped with enough computers and sufficient internet connections. In term of curriculums, statistics education for primary and secondary students are traditionally centered around the computation of central tendencies and graph representations.

Although inferential statistics is not specifically mandated to be taught in Indonesia, implementing it will be beneficial in the long run since formal inferential statistics is hard to understand for Indonesian students as well. Promoting inferential reasoning while learning about descriptive statistics can help students bridge the connection between descriptive and inferential statistics. Watson & Moritz (1999) presented the idea that there are many points in the curriculum where the idea of inference can be introduced, together with descriptive statistics.

Some studies in Indonesia have already brought up discussions about more meaningful way in learning statistics; most of them focus on average. Lestariningsih (2012) proposed the use of folklore as a context to support students in learning average, while Assagaf (2014) suggested the use of length measurement as a context. However, as far as the researcher is aware of, there is no study that specifically aims to incorporate IIR in early education in Indonesia.

Based on those issues, it is important to design meaningful contextual activities in order to foster the development of the students' informal inferential reasoning. The study focuses on the 7th grade students in Indonesia.

1.2 Research question

Based on the elaboration above, the research question of this study is formulated as follow, "*How can 7th grade students be supported to develop informal inferential reasoning?"*

1.3 Research Aim

The aim of this study is to design instructional materials to support Indonesian students in developing informal inferential reasoning, as well as to contribute a local instructional theory focused on developing IIR to statistics education in Indonesia.

1.4 Criteria of the study

Based on the research question and aim, we set up a criteria for this study which is elaborated as follows.

1. The students are able to achieve the learning goal of each lesson.
2. The students are able to participate in class discussion and to argue for and against a statement without using any formal statistical terminologies. This is exhibited in the students' worksheet and the transcript of the class discussion.
3. The students are able to make generalization from sample to population, to back up this generalization with data, and to produce this generalization with a degree of uncertainty.
4. The hypothetical learning trajectory (HLT) is realized within the actual learning trajectory (ALT). The HLT and the ALT will be compared and contrasted by means of Dierdorp's matrix analysis.

1.5 Definition of some key terms

It is important to define the key terms involve in this study in order to avoid different or mis-interpretation for the terms. These terms and its definition are elaborated as follows.

1. Inference

   A guess or opionion made by a person based on information that they have.

2. Informal Inferential Reasoning

   Informal inferential reasoning is defined as the way in which students use their informal statistical knowledge to make argument to support inferences about unknown populations based on observed samples. It can also be defined as cognitive activities involved in informally drawing conclusions or making predictions about “some wider universe” from patterns, representations, statistical measures and statistical models of random samples, while attending to the strength and limitations of the sampling and the drawn inference. There are 3 characteristic of IIR: (1) generalization beyond the data, (2) the use of data to back up this generalization, and (3) the employment of probabilistic language (statement of uncertainty) in describing this generalization.

3. Developing

   Developing is defined as progress. In the range of education, developing the students’ understanding means a progress that connects the existing knowledge and the intended learning goal.

4. Supporting

   Supporting is defined as giving assistance or encouragement. Supporting does not inherently mean that the students are given help al the time. Instead, they are expected to form their own knowledge with facilitation and encouragement from the teacher.

1.6 Significance of study

Related to the objectives of the study above, the present study is expected to be able to contribute to the development of a local instructional theory in domain of statistics. These contributions includes the classroom material as well as an ccompanying theory on how it foster the students' IIR.

For the teacher, it is expected to be an innovative way to teach statistics in the classroom as well as to introduce IIR to Indonesian educators. It is also expected to become references for other researchers who will conduct studies on relevant issues.

CHAPTER II

THEORETICAL FRAMEWORK

2.1 Statistical Literacy

Although literacy is more commonly associated with the ability to read and write, the term has been used in various fields to address a basic or minimal set of skills required to function sufficiently in that field. Ben-Zvi and Garfield (2004) proposed a definition of statistical literacy as "basic and important skills that may be used in understanding statistical information or research results" (p. 7). Wallman (1993) defined statistical literacy as "the ability to understand and critically evaluate statistical results that permeate our daily lives–coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions" (p. 1). Gal (2002) identified statistical literacy as a compound of two components namely (a) the ability to critically interpret and evaluate statistical information one encounters in life and (b) the ability to discuss and communicate such interpretations and evaluations.

With the definition of statistical literacy, comes the need to determine what knowledge, skill or competence that has to be possessed and mastered by an adult to be classified as statistically literate. Gal (2002) argued that people's statistical literacy depends on two components namely *knowledge* and *dispositional* components. The knowledge component consists of five interrelated knowledge bases (*literacy*, *statistical*, *mathematical*, *context*, and *critical*), while the dispositional component comprises the accompanying beliefs and attitudes.

Statistical information is an unescapable part of people's professional and daily lives. Modern citizens are bombarded with statistics used by politicians, researchers, and companies in an effort to strengthen their arguments. Moore (as cited in Gal, 2002) stated that data, variation, and chance are universal and ever-present parts of daily life, mentioning the impossibility to find policy question that has no statistical component as an example.

There are two roles that people can assume in regard to statistical information that they may have in everyday life; they can either be the *consumers* of the information, or the *producers* (Gal, 2002). People assume the role of a consumer unconsciously every time they encounter statistical information in everyday life, such as poll result or graphs in a news articles. The producer of legit statistical information, on the other hand, are mostly those working as statisticians or researchers; as in individuals who conduct statistical investigation and procedures on daily basis.

Considering the contexts in which statistical information usually appear, the role of a consumer is definitely much more common than the role of a producer. Nonetheless, people in every profession still need to be able to produce well-grounded arguments based on reliable facts and statistics. To conclude, at some point in their life an adult in modern society will be in a situation where they are demanded to be a consumer or a producer of statistical information, and they need statistical literacy to handle this situation effectively.

Statistics is one of the five domains of mathematics, hence it is traditionally being taught to students in school as part of mathematics curriculum. Despite the effort of statisticians in recent years to establish statistics as a separate and unique discipline from mathematics (del Mas, 2004), it cannot be denied that statistics are strongly linked to mathematics considering the numerical information and the mathematical formulas utilized within it. Gal (2002), in his model of statistical literacy, also argued mathematical knowledge as one of the bases for statistical literacy.

Considering that statistical literacy is a competency that a modern citizen is expected to have, it is one of the expected outcome of statistics education (Gal, 2002). However, teaching students statistics at school does not necessarily mean that its graduates are statistically literate. A lot of researches have brought up interesting findings related to statistical literacy in society, or lack thereof. Mullis et al. (as cited in Gal, 2002) reported that less than half of the students graduated from high school are able to do statistical tasks designed to measure statistical literacy. Kahneman, Slovic, and Tversky (as cited in Garfield, 1995) pointed out misunderstandings by adults about statistical concepts that are prevalent in daily life, such as representativeness of sample and gambling fallacy. Wallman (1993) also pointed out various misunderstandings, misperceptions, mistrusts, and misgivings people have in regard to how statistics are involved in making private and public choices.

Causes behind this tendency to misunderstand statistical information are identified; one of them is the traditional way statistics is being taught in school, which is too much focused on procedures and computations. The focus on procedures and computations is underpinned by an expectation by that the answer should be unique, correct, and numerical, while in reality, statistics is far from that. This contrasting nature of statistics as a knowledge is uncomfortable for students.

Calculating numbers is also tedious work prone to mistakes and providing no mental stimulation. Hence, the students do not only find statistics difficult, they also find learning statistics unpleasant (Garfield & Ben-Zvi, 2004).

The ineffectiveness of the teaching of statistics calls for a reform. Some of the change to be expected are the change of a focus on analysis instead of computation, and a focus on encouraging the students to think and to reason statistically (Garfield & Ben-Zvi, 2004). Therefore, the recent argument in statistic education is how to reform the way statistics being taught in school to produce citizens that are statistically illiterate as the outcome.

2.2 Informal Inferential Reasoning

In statistics, two broad categories can be distinguished: descriptive statistics and inferential statistics (Walpole, 1995). Descriptive statistics are procedures used to find patterns in a set of data as well as to use these patterns to summarize and describe the set of data itself. Inferential statistics relate to the processes of making inference about a population from observing and analyzing a sample taken from that population. That is, descriptive statistics do not go beyond the data at hand, while inferential statistics makes generalization from a sample to the population it's taken from.

There is no doubt that inferential statistics is a profoundly significant knowledge. Inferential statistics is communally agreed as the heart of statistics, since the main reason for statistics to be developed as a science is for people to investigate a phenomenon or to predict a future trend in society (Bakker & Derry, 2011). Those phenomena and future trends are investigated through a series of scientific methods applied to a very limited sample, based on which conclusions and decisions regarding a larger population have to be derived. Generalizing to a larger population from a limited sample is where inferential statistics takes its role.

Looking at the mathematics curriculum, statistics in most countries is introduced in primary school. Up until the end of their secondary education, statistics learnt by students traditionally revolves around descriptive statistics and rarely goes beyond computing average and interpreting graphs (Makar & Rubin, 2009). Inferential statistics, on the other hand, is usually introduced to students in the tertiary level (Makar & Rubin, 2009; Rossman, 2008; Pratt, Johnston-Wilder, Ainley, & Mason, 2008). It revolves around statistics that are important in conducting research, such as hypothesis tests, parameter estimations, or normality tests. Since the inferential statistics being introduced in college statistics classes mainly consists of formal procedures, it is referred to as formal inferential statistics.

Students' difficulty in inferential statistics has been addressed by prominent statistics educators in statistics education journals and conferences (Pratt & Ainley, 2008). The solution proposed to tackle this problem is to introduce statistical inference to the students below tertiary level (secondary and primary). However, since formal inferential statistics is cognitively not suitable for them, it is approached by focusing on informal aspects of it. This is known as *informal inferential reasoning* (IIR).

IIR is a relatively new topic in the field of statistical education. A general consensus about definition and frameworks is yet to be reached. Some researchers, however, have formulated their definitions built around their respective studies. Pfannkuch (2006) described IIR as the drawing of conclusions about a population that is based mainly on looking at, comparing, and reasoning from distributions of data. Rossman (2008) described it as the way students infer beyond the data at hand through reasoned arguments that employs no formal method, technique, or calculation. Another definition is by Zieffler et al. (2008), who formulated IIR as "the way in which students use their informal statistical knowledge to make argument to support inferences about unknown populations based on observed samples" (p. 44), while Ben-Zvi, Gal and Apel (2007) defined informal inferential statistics as "cognitive activities involved in informally drawing conclusions or making predictions about "some wider universe" from patterns, representations, statistical measures and statistical models of random samples, while attending to the strength and limitations of the sampling and the drawn inference" (p. 2).

Makar and Rubin (2009) suggested three key principles that are essential in making statistical inference informally: (a) generalization beyond the data, (b) the use of data to back up this generalization, and (c) the employment of probabilistic language (statement of uncertainty) in describing this generalization.

Some researchers have reported notable progresses in fostering informal inferential reasoning in young students (primary and secondary school). The study by Bakker and Gravemeijer (2004) showed that 7th graders can develop the concept of distribution through informal ways of reasoning. Ben-Zvi et al. (2007) explored the possibilities of developing IIR for 6th grade in technology-rich environment by using data generator software, with similar software also used by Watson (2008) to develop IIR in 7th grade. Both study noted that the students start to produce statements that address the elements of IIR.

Recent study by Makar & McPhee (2014) showed that through a series of predicting activities, grade 1 (5 years old) can be supported to develop IIR. Similar result is also reported by English (2014) on a study done to grade 1 to 3 (6-8 years old). The less strict nature of informal inferential statistics allows it to be taught across various education levels. Even though the main idea is to start with young students, some research also record its implementation for student in late secondary and early tertiary education (Wienber, Wiesner & Pfaff, 2010).

2.3 Example of informal inferential reasoning

A study by Hamlin, Wynn, & Blomm (Rossman, 2008) investigated whether or not infants find an individual appealing because of that individual's reaction toward others. 10-month-old infants where shown a character, called 'climber', that struggle to climb up the hill. Other characters, called 'helper', pushed the climber to the top, while 'hinderer' pushed the climber down the hill. Each infant is shown the characteristics several times, before being presented with the helper and the hinderer. The study recorded that for a sample of 16 infants, 14 chose the helper toy. The question is that, it is because of genuine preference or a mere chance? Do infants really have a preference?

Formal inferential statistics will require formal calculation of an exact p-value or a test statistics. On the other hand, the students without inferential reasoning will probably conclude that the infants do have preference by solely looking on the data.

With informal inferential reasoning, the students will have to make a generalization toward the population of sample in general. Their claim about whether or not the infants have preference will also be based on the data. They will also produce a statement of uncertainty or certainty. Suppose their assumption is that the infants have no preference, then the infant picking the helper can be modeled by flipping a coin and getting the head. It is quite unusual to flip 16 coins and get 14 heads, hence it can be argued that the possibility of 14 infants picking the helper toy without any preference is quite small.

On the other hand, the argument about uncertainty can also come from the quality of the sample. The students can argue for the infants having no preference considering the sample size. Even though almost all the infant chose the helper toy, the size of the sample is too small to represent the whole population.

To conclude, as informal inferential reasoning bridge between descriptive statistic and formal inferential statistics, the students in this stage do not use any formal inferential statistical procedures but do not produce any deterministic statement either.

2.4 The importance of Informal Inferential Reasoning

The reason why Informal Inferential Reasoning is a reasonable topic to be researched is described as follows.

2.4.1 Students’ difficulty in learning formal inferential statistics

Formal inferential statistics are thought to be difficult and students in introductory statistics class often have difficulty understanding the reasoning behind it (Makar & Rubin, 2009; Garfield & Zvi, 2008). Various causes have been identified.

The first one is that descriptive and inferential statistics are taught separately. Students are often unable to link the descriptive statistics they learn in their compulsory education to the inferential statistics they learn in college. Brandom (as cited in Bakker & Derry, 2011) brought up this issue by discussing the distinction between *holism* and *atomism* in statistics education. *Atomism* approach refers to learning individual concepts separately and only about what these concepts refers to, then combining them to form a coherent body of knowledge. On the contrary, *holism* is described as "… one cannot have any concept unless one has many concepts. For the content of each concept is articulated by its inferential relations to other concepts. Concepts, then, must come in packages." (p. 11). *Holism* is preferred over *atomism* because topic-by-topic approach is deemed to be no longer effective.

Learning descriptive and inferential statistics separately also leads to the second case, which is the difficulty to deal with the probabilistic nature of inferential statistics. The traditional approach to statistics in the early grades can encourage an overly deterministic sense about statistics. Upon entering statistics classes in college, the students generally lack experience in working with and reasoning about stochastic events that underpin statistical inference (Zieffler et al., 2008).

The third case is the focus on procedures and algorithms, what is fondly referred to as the "mean-median-mode" syndrome (Bakker & Derry, 2011). It creates a shallow understanding such a way that although students can remember the formula, they often cannot apply it in a slightly alternated context. The concept of mean is one of the examples of this inert knowledge. It is introduced to students as the formula of arithmetic mean when they are as young as grade 6. Students however rarely use it to compare groups, instead only use it descriptively or calculating it out of habit (Bakker & Derry, 2011).

Garfield (as cited in Harradine, Batanero, & Rossman, 2011) stated that the teachers prefer to teach concepts and procedures in the hope that the reasoning will develop by itself. As the result, the students start their formal inference with a reasoning-free background and a detrimental misconception that statistics is only about counting numerical values. Zieffler et al. (2008) also agreed that inferential reasoning is unlikely to develop by itself with maturation or life experience, hence it needs to be intentionally developed. Therefore, it is considered important for the teachers to provide students with ways to attend to other relevant aspects of data sets such as variations, distribution and sample size, in addition to learning definitions and calculations. Without paying attention to its conceptual underpinning, students will continue to find inferential statistics problematic (Pfannkuch, 2006).

Furthermore it has been argued that doing computations should not be the focus anymore, at least not as much as reasoning the behind said computations, as there are fair number of software that can help the students to do the calculations nowadays (Harradine et al., 2011).

2.4.2 Supporting statistical literacy

Inferential reasoning is a central objective of statistical reasoning (Paparistodemou & Meletiou-mavrotheris, 2008). Students are going to encounter and deal with statistical information even if they are not enrolled in statistics class or working in a statistic-related career path. Therefore aside from easing the students in their transition to formal statistical inference (Zieffler et al., 2008), IIR can help students in their future work environment where formal statistical procedures cannot be carried out.

Since the goal of statistics education is for students to achieve statistical literacy in the end of education, it is important to consider the role that statistical inference has in it. Statistical inference has been referred as one of concepts that build statistical literacy. In the framework of statistical literacy proposed by Gal (2002), for example, the statistical knowledge base consists of five parts which includes statistical inference as one of them. The five parts of the knowledge base are: (a) knowing why data are needed and how data can be produced, (b) familiarity with basic terms and ideas related to descriptive statistics, (c) familiarity with basic terms and ideas related to graphical and tabular displays, (d) understanding basic notions of probability, and (e) knowing how statistical conclusions or inferences are reached.

2.4.3 The need for innovation

Being innovative relates to looking beyond the current working system, identifying the ideas of tomorrow and putting them into practice. Innovation in education means trying new ways of doing things in their learning environments. Education system is often viewed as the main army a nation have in an eternal battle against unemployment and constantly being in the center of policy debate, hence in order to be well-equipped in dealing with change, a country need to constantly innovate (OECD, 2010).

Along with the need for statistical literacy, ideas about reforming the current way statistics being taught in school have arisen. Statistic education in elementary and middle grades in particular, is encouraged to shift its focus on techniques of graph construction and calculation of statistical measures to a focus on ideas such as distribution and variability (Leavy, 2010).

Informal Inferential Reasoning is a new topic and therefore is not included in conventional mathematics curriculum. However, it is an interesting new topic that can be the profound innovation to the current statistics education.

2.5 Developing informal inferential reasoning

In order to develop IIR, a series of lessons are to be designed. Therefore, a careful consideration needs to be made regarding what needs to be employed in order to create an effective lesson; a lesson that can bridge students' prior knowledge and the learning goal.

2.5.1 Sample and population

The concept of sample and population has been deemed important by various researchers. Rubin, Hammerman and Konold (as cited in Ben-Zvi et al., 2007) mentioned the idea of sample size and sampling procedure, as well as representativeness, controlling for bias, and tendency as being involved in informal inferential reasoning. Zieffler et al. (2008) also emphasized on sample and population, stating that informal inferential reasoning as a process includes (a) reasoning about possible characteristics of a population based on a sample, (b) reasoning about possible differences between two population, also based on sample, and (c) reasoning whether or not a particular sample and its summary is likely or surprising. In the design research by Bakker and Graveimejer (2004), an activity in which the students take a small subset of data and add more and more data value into it is proposed as a way to support students in developing the idea of distribution. This activity is called *growing sample.*

7th grade students have no conception yet about sample and sampling process since the concept of sample and population is formally introduced in grade 9 (14-15 years old). However since the idea of sample and population is important for IIR, the students need to develop the idea of it even though they do not necessarily have to be exposed to the formal terminologies and definitions. It is also worth noted that in the last activity designed by Assagaf (2014) contains inferential reasoning because the students are expected to use the average value of one group to represent the whole population. This design failed because the students added up the average of all groups and divided it by the number of the groups instead, which was caused because they have not been exposed to the idea of sample and population yet.

Sample and sampling process therefore becomes main idea where the designed lessons should be started with and centered at. Instead of giving them the definition of sample and population, students have to be given the opportunities to experience sampling behavior and to develop the need to sample (Pfannkuch, 2006). It is also important to expose students to the distinction between sample and population (Biehler, 2014) as well as the aspect of sample that determine its representativeness like sample size and randomization. Wroughton (2013) also stated that the need for unbiased sampling method is one of the most important topics taught in introductory statistics courses, thus developing ability to assess quality of sampling method is important in statistical inference.

2.5.2 The use of context

The students will have to look beyond the data at hand, as well as to persuade others and to argue about their inference. To do that, they need to be interested and engaged in the problems. Hence, the choice of context is very important. Real life data, surprising and complex problem situations that contradict their prior beliefs are useful (Ben-Zvi et al., 2007), as well as compelling question that propel students to look beyond the data at hand.

In this study, the context chosen is height measurement (in cm) and the use of social media (in hours). Height is a biological measurement that every individual student is familiar with, while the use of social media is an issue that currently being talked about in society. Both of them are real life data that the students collect themselves.

2.5.3 Dot plots

Activity for statistics lesson for young children are encouraged to be based on *exploratory data analysis* (EDA), which mostly about analyzing and summarizing a set of data mainly with visual methods (Paparistodemou & Meletiou-mavrotheris, 2008). The use of statistical procedures are not compulsory, which is convenient since premature introduction to algorithm is argued to result in inert knowledge (Bakker & Derry, 2011). Therefore the kind of data representation needs to be chosen wisely that encourage students to analyze data visually.

Data visualization that is proposed to be used here is dot plot. Dot plots are one of the simplest statistical chart, initially exist as a hand-drawn graph to depict distribution (Wilkinson, 1999). A dot plot is constructed by plotting data points on a fairly simple scale, which on most cases is horizontal and does not require 0 as starting point. It is used for continuous and quantitative data. They are useful for moderately sized data, as well as to highlight clusters, outliers, range, and other characteristics (Wilkinson, 1999).

Dot plots is fairly easy to construct without computer software, hence it can be supportive in classroom environments that is not technologically equipped. It also supports EDA because it gives way for the students to visually detect the main characteristics of the data.

The use of dot plots has been mentioned in a study done by Pfannkuch (2006) to 15 years old students, where they have to use box plots to reason about inference. The result of the studies suggests that the use of box plots might be too early for students, and that they should be supported to reason with it through keeping the data, in dot plot form, under the boxplots (Bakker, Biehler, and Konold, as cited in Pfannkuch, 2006).

2.5.4 Vocabulary

Language play important roles in mathematics instruction, occupying at least three crucial roles in mathematics classroom which are (a) way of communication, (b) what students build understanding with as they process ideas through languages, and (c) assessment (oral and written communication) (Thompson & Rubenstein, 2000). Students' informal mathematical knowledge are built upon their everyday experiences and cultural background, and in classroom discourse where this kind of knowledge is the subject of attention, language plays central role (Walshaw & Anthony, 2008).

Mathematics as a science offers its own challenge since its vocabulary are subject-specific. Researches have shown how languages and lack of attention to it can result in problem for students, mainly in classrooms with non-native speaking students. Issues related to mathematics vocabularies in mathematics has been the subject of many researches (Thompson & Rubenstein, 2000). Christensen (as cited in Walshaw & Anthony, 2008) conducted a study about mathematical langauges in Maori mathematics classroom that shows how mathematical words can be problematic. Eerde & Hajer (2005) brough forward the language-related problems within multilingual Dutch secondary schools and proposed a design of classroom material to tackle this problem.

Statistics vocabulary are even more exclusively used since it shares very little common vocabulary with other domains of Mathematics itself. The following are issues related to vocabulary commonly found in the field of statistics.

Table 2.1. Issues related to statistical vocabulary

Category of issue	Example of word (English)	Example of word (Indonesian)
Some words are shared by mathematics and everyday life, but they have distinct meanings.	Mode, event, combination.	Modus, kejadian, kombinasi, kelas, tepi atas, tepi bawah.
Some mathematics words are shared with everyday language and have comparable meanings, but the mathematical meaning is more precise.	Average.	Rata-rata, sampel, distribusi, acak.
Some mathematical terms are found only in mathematical context.	Outlier, permutation.	Pencilan, permutasi, varian, standar deviasi..

Some words have more than one mathematical meaning.	Median, range.	Median, range.
Modifiers may change mathematical meanings in important ways.	Number or random number, probability or conditional probability.	Distribusi frekuensi.
Some mathematical phrases must be learned and understood in their entirety.	Stem-and-leaf	Diagram batang daun.
Some words shared with science have different meanings in the two disciplines.	Simulation, experiment.	Percobaan, eksperimen.
Some mathematical terms sound like everyday words.	Leaf, in steam-and-leaf, sounds like leave.	
Some mathematical words are related, but students confuse their distinct meaning.	Dependent events and independent events.	Kejadian saling lepas dan kejadian saling bebas.
Technology may use language in special ways.	LnReg	STDEV
Some words are partially/not translated into Indonesian.		Range, representatif, variabel.

Note. Adapted from "Learning Mathematics Vocabulary: Potential Pitfalls and Instructional Strategies" by Denisse R. Thompson and Rheta N. Rubenstein, 2000, *Mathematics Teacher*, 93(7), p. 569

Just like students cannot understand statistics just by learning formulas and procedures, it is also debatable to assume that they will have sufficient understanding only by being given the definition of these words. Learning a mathematical term is not as simple as getting and learning its definitive meaning(s). It is important to guide the students off everyday language that they are comfortable with and ease them into using technical mathematics language correctly (Leung, 2008).

The solution that is proposed here is exploiting the informal language produced by the students themselves. When teachers and students are engaged in

classroom discourse, informal langauge can play an important facilitative role. Research suggests that informal and formal language can be used in various combinations in class, and that students can use informal language productively to explore concepts represented by technical vocabulary (Leung, 2008). Paying attention to informal language to develop formal language is also proposed by Thompson & Rubenstein (2000), who suggest several interesting strategies such as: 1) building the concept first and then attach the vocabulary to the established ideas, 2) asking students to explain their work orally with the teacher actively correcting, revoicing, and guiding their utterance to be more mathematically precise, and,
3) asking the students to write in the form of a journal where they note down statistical vocabulary that they encounter in class every day and provide them with their own definition.

Several studies have been conducted with statistical vocabulary as main or side attention. Watson & Kelly (2008) conducted a study to students in grades 3, 5, 7, and 9 about their ability to define three fundamental statistical terms: sampe, random, and variation. Pfannkuch (2011) did a study on informal inferential reasoning in which the teacher will introduce the vocabulary in the beginning of the class, and then write it on the board every time it comes up in discussion.

Last but not least, how effective the scaffolding of vocabulary can support statistical literacy also needs to be considered. Statistical literacy can be thought as a combination of statistics and literacy. The statistics part involves understanding of statistical concepts, while the literacy part is about the ability to express statistical concepts in words, as well as to communicate, discuss, and argue about it (Watson & Kelly, 2008). It is evident that subject-specific vocabulary is an inherent part in supporting the ‘literacy’ aspect of statistical literacy. Gal (2002) also mentioned the ability to discuss and communicate such interpretations and evaluations as one of the components that build statistical literacy.

2.6 Realistic Mathematics Education

Realistic Mathematics Education is an approach in Mathematics Education based on the idea of Hans Freudenthal (Gravemeijer, 1994), initially as a way to reform mathematics education in the Netherlands. Freudenthal´s view is that Mathematics is a human activity. As opposed to receiving everything ready-to-use, students must build their own mathematical conceptions based on their own world.

This human activity is centered around the idea of what Freudenthal refers to as *mathematizing,* which he suggested to be the key process in Mathematics education and relates to the process of translating day-to-day problems into equivalent mathematical terms (Gravemeijer, 1994).

To evoke this mathematizing process, RME encourage the use of contextual problem (Gravemejer, 1994). Context-bound problems are easily understood by the students because these are closer to their everyday lives, hence they encourage the students to develop and use their own solutions. Different solutions from different students will lead to a discussion about adequacy, and then turn into the idea of efficiency (Gravemeijer, 1994). From the need to be efficient, a generalization to formal mathematical term arises.

For instructional material with Realistic Mathematics Education approaches, there are three principles which function as heuristic during the designing process. They are 1) *guided reinvention* and *progressive mathematization,* 2) *didactical* *phenomenology*, and 3) *self-developed models* (Gravemeijer, 1994). This elaboration will focus on the first two of the principles, which support the development of Informal Inferential Reasoning.

Guided-reinvention means that the students should be given a chance to experience the process by which the aimed mathematical concepts are invented. Being exposed to the situation that gives rise to those mathematical concepts and being immersed in the related process will encourage students to start developing their own procedures, albeit informally. These informal procedures are later mathematized to formal procedures, which is called progressive mathematization.

Didactical phenomenology states the importance in selecting the problems solved by the students. Historically, mathematics was developed as a way of problem solving, therefore during the design process, we have to select problems that not only offer the students a chance to solve it themselves, but also the possibility to generalize this to other situation.

In reasoning about inferential statistics informally, students use their informal statistical knowledge to make arguments based on observed samples to support inferences about an unknown population. Hence the emphasis is not on statistical procedures and mathematical arguments, but more on reasoning from characteristics of data, conceptual understandings and the provision of explanation.

Informal knowledge, although it seems groundless at glance, is not necessarily purposeless. It is a less formal knowledge resulting from prior formal knowledge (Zieffler et al., 2008). Informal knowledge is important for students as a foundation that they can build on and to reinvent formal concept and representation with. Hence, IIR implies that instead of making inferences of data using statistical formulas, students are encouraged to use informal knowledge that surely is more familiar to them. This informal knowledge will cause them to reinvent the formal statistical concepts, that they can use broadly across many statistics contexts.

In this research, a sequence of instructional materials will be designed to foster the development of IIR. The principle from Realistic Mathematics Education, especially guided reinvention and progressive mathematization, as well as didactical phenomenology, will be employed as the design heuristic during the designing process of the instructional materials.

2.7 The Role of Teacher

The idea of effective learning is where the students participate actively instead of acting as passive receivers. This, however, does not imply that the role of the teacher is irrelevant. The role of the teacher in the classroom will be elaborated as follows.

2.7.1 Decision maker related to the mathematical contents and tasks

A lot of factors have to be taken into consideration during the designing process of a lesson. The formulation of the learning goals is determined by the curriculum in each respective country. The activities are chosen and designed such a way that it can be conducted within the time frame available. The students' prior knowledge and the school's resource are also needs to be considered.

On the other hand, the flow of the lesson needs to be in accordance to the pace of the students' learning. The students need to be given space to explore and to construct their own knowledge. There has to be a balance between predetermined learning goals and the mathematics built by students (Simon, 1995).

The responsibility to be responsive to the students' learning while keeping track with the learning goals leads to the development of *hypothetical learning trajectory* (HLT) . It is an instrument with significant role throughout all phases of design-research. HLT is essentially the same as the learning trajectory that most teacher are familiar of, but designed in such a way so that it can adapt as well as possible to unforeseeable circumstances (van Eerde, 2013). More detailed explanation about HLT will be given in the later chapter.

The teacher holds very important role as a the planner and executor of these three components; the learning goals, the learning activities, and the HLT. In this study, however, the researcher will take over the planner role and leave the executor role to the teacher. This does not mean that the teacher will just stay there waiting for the command from the researcher. Quite the contrary, in fact, because the teacher is the one who are knowledgeable about the classroom community and situation. The teacher's knowledge will be constantly taken into account during the design process and the formulation of HLT.

2.7.2 Setting up the context and classroom environment

As has been explained before, contexts are inherent part in developing IIR. Since the students so far only have experience in working with descriptive statistics, moving beyond the data at hand to an unknown but bigger set of data is a very new idea for them. They need to be engaged to the context to encourage them to look deeper into the data.

Regarding this matter, the role of the teacher is to engage the students by presenting the context as interesting as possible. The teacher can personalize the context by connecting it to their own or the students' experience, or relate it to the current issues happening in society. Questions or prompts by teacher also have important role in eliciting statement of generalization from students.

2.7.3 Establishing social norms and sociomathematical norms

During the course of the lesson, there will be group works and presentations. All the groups will work on different samples and come up with different answers. The students are also used to the deterministic nature of descriptive statistics, in where they can see all the data and there is always an exact and numerical answer. Therefore, general social norms have to be established beforehand to ensure that the lessons run smoothly without being intervened by undesirable forms of interaction between and among classroom community.

Walshwa & Anthony (2008) did a meta-analysis on what sort of pedagogies that through classroom discourse will encourage students' engagement in mathematics. One of the result of this study is that various researches agree on "participating rights and obligations", which essentially means that every students has to engage in classroom discussion, either to explain their thinking or listening to others doing so. Therefore the teacher not only has to establish, but also clarify and enforce discourse participation rules within the classroom community.

During the discussion, it is important for the teacher to lead the discussion and argumentation and facilitate dialogue, not only for the mathematical development but also to ensure important and desirable social outcomes (Walshaw & Anthony, 2008).

Sociomathematical norms, on the other hand, are normative aspects of mathematical discussions specific to students' mathematical activity (Yackel & Cobb, 1996). This includes normative understandings of what counts as mathematically different and mathematically sophisticated or efficient, as well as knowing what kind of mathematical explanation and justification can be count as acceptable.

These norms are not predetermined criteria; they are supposed to be constructed by each classroom community (Yackel & Cobb, 1996). Each class might have different clarification on what can be constituted as mathematically sophisticated or different. In developing informal inferential reasoning, every group will have different inference, method of sampling, or argument for representativeness. Discussion centered around these differences is expected to end with consensus about sophistication or efficiency – what method of sampling is effective and unbiased, or the correct and proper use of statistical vocabulary/terminologies.

2.7.4 Tending the mathematical vocabulary

In this study, the students are encouraged to develop the concept themselves without being pressured to use technical vocabulary at the start. Over the course of the lesson, the students are predicted to use general every day words or phrases in the place of formal terminologies they are yet to be aware of. They will use *bump, scattered,* and *picky* in the place of *typical, spread,* and *bias*, for example.

However, the students are also expected to be able to use technical words correctly, which in terms of sochiomathematical norm can be counted as being mathematically sophisticated. This is where the role of the teacher comes up; to tend to the coping strategy of using generic every day word and help the students shaping it into a mathematically sophisticated argument. Language is also pointed out by Walshaw & Anthony (2008) as an important part of pedagogies that teachers need to pay attention to. They stated that "responsibility for distinguishing between terms and phrases and sensitizing students to the particular nuances weighs heavily with the teacher, who profoundly influences the mathematical meanings made by the students" (p. 533).

What the teachers are advised to do is to emphasize these informal words or phrases by either asking the students to revoice it or revoicing it themselves, or writing them down on a special section of the white/blackboard. In the end of the lesson, the teacher can discuss this word and introduce the equivalent formal words.

2.8 Statistics in Mathematics Curriculum in Indonesia

In Indonesian curriculum, statistics starts to be introduced in primary school. There is clear separation between descriptive and inferential statistics. Descriptive statistics is introduced in a traditional way that focuses on computation and procedures. Inferential statistics, on the other hand, is only started to be introduced at college level statistics.

In primary school, students are introduced to data collection, organization and representation, as well as the measure of central tendency. These concepts developed further in junior high school, together with the concept of probability. IIR is not specifically mandated in the curriculum, however the current curriculum does provide room for it to be developed.

Research about statistics education in Indonesia so far has been focused on developing the idea of measures of central tendency (e.g. Lestariningsih, 2012; Assagaf, 2014). Considering a lot of studies suggest that IIR can be successfully taught to junior high school students, as well as the fact that primary school students are just entering their explorative data analysis phase, a study about the development of IIR is carried out in 7th grade.

Based on the new curriculum 2013, the basic competence and standard competence in statistics for 7th grade are the following.

4.8 Collecting, analyzing, interpreting and presenting data of observation in the form of a table, chart or graph.

4.9 Conducting experiments to find empirical probability of a daily life problem, as well as presenting it in the form of a table and graph.

(Kemendiknas, 2013)

Based on the description above, it can be concluded that:

1. There is a shift in the goal of statistics education, resulting in a more holistic approach that is demanded by the stakeholders of statistics education in the world.
2. Formal inferential statistics is an important yet difficult subjects that creates obstacles for students in learning statistics.
3. Informal Inferential Reasoning is a powerful concept that can help students bridge their informal knowledge to formal inferential statistics, also to assist them in professional environment where formal statistics is needed.

Considering its importance and the lack of discussion on this topic in Indonesia, the aim of this study if to contribute a local constructional theory in 7th grade to develop IIR. The research question is formulated as follows: *How can 7th grade students be supported to develop Informal Inferential Reasoning?*

2.9 The pre-HLT

The aim of this study is to contribute an empirically grounded theory in supporting 7th grade students to develop IIR. To accomplish that, a set of instructional material is designed, consisting of students' worksheet and the accompanying teacher guides, as well as the accompanying *hypothetical learning trajectory* (HLT), which will be explained in detail in Chapter IV.

The instructional material and the HLT is actually the end product of the designing and preparation phase of design research. However, the HLT previously constructed according to the result of literature review, resulting in the pre-HLT. The pre-HLT is further improved by considering the result of the classroom observations and the teacher's interview, resulting in the final version of the HLT which will be implemented in the teaching experiment(s). Therefore, the pre-HLT will be explained as the closing of this chapter.

As explained in the theoretical framework, there are four aspects that are hypothesized to be essential in lessons focused on developing IIR: 1) a focus on sample and population, 2) the use of context, 3) the use of dot plots as the type of data representation, and 4) delaying the use of formal statistical terminology and using its informal equivalent proposed by the students instead.

Schwartz, Sears, and Chang (in Weinberg et al., 2010) suggested that the students' informal knowledge should be developed through activities that provide the foundation for the formal instruction later on. Since the target of this study is 7^{th} grade, they are used to the descriptive idea of statistics and have not yet possessed any understanding about sample and population. The students need to be introduced to sample and population, as well as why sampling is necessary. Aside from that, the students also need to learn about the quality of sampling process. The following are the four ideas that the students need to develop.

1. A part of data can represent the whole (representativeness).
2. It is impractical to take the whole data into account, hence the need for sampling.
3. The bigger the part of data, the more it resembles the population (effect of sample size).
4. All member of population has to have equal chance to be picked as sample (randomization).

Hence, to support the students in developing IIR, the researcher has designed four lessons, built around those four ideas and those four aspects. The lessons, comprising the learning goal and the description of the activity are elaborated as follows.

1. Lesson 1: Scouts out and about

The aim of this lesson is for students to recall the concept of data as well as data collection, representation, and summary that they received in primary school. The context of this lesson is finding out suitable Scout staff length for the students in grade 7.

This lesson contains two activities. The students work in group and each group is given a population box containing pieces of paper with the height of the students written on it. In the first activity, the students take a certain number of data from the population box and represent it in the form of dot plot. They use blue dot to represent male students and pink to represent female. In the second activity, the students work together to create the class dot plot. Each student get a sticker, pink for girl and blue for boys, and they stick it to the plot provided on the whiteboard. The students are then asked to describe their plots and find out the typical value.

Hands-on sampling activity are recommended by researchers (Weinberg et al., 2010) as an effective way for the students to experience the process of sampling by actually taking a part of data from the whole. Moreover, the students are most at ease when comparing data visually and using their hand gesture to explain it (Watson & Moritz, 1999). The use of dot plots and different colors for male and female students encourage this.

The students have been introduced briefly to arithmetic mean in the 6^{th} grade, but the use of it is not encouraged in this lesson. Learning algorithm can be detrimental in the long run (Watson & Moritz, 1999) and those who use the mean as learned algorithm without any appreciation of representativeness were generally unsuccessful in using it in application.

2. Lesson 2 : Compare the dots

The aim of this lesson is to develop the idea of sample representativeness and sample size. In the previous lesson, the dot plot made by the students by taking a certain number of data from the population box is referred to as ***the group chart***, while the whole class dot plot is referred to as the ***class chart***. The context of this lesson is the disappearance of the group chart, so the students have to determine the group chart that is suitable to replace it.

There are three activities contained within this lesson. In the first activity, the students compare and contrast their group chart to the class chart. In the second activity, the students compare and contrast all the group charts to the class chart. The conjecture is that the students pick the group chart that contains similar information to the class chart as the most representative one.

In formal statistical inference, the students usually begin with a single sample and a hypothetical population (Watson & Moritz, 1999). It might be better for younger students if they can compare two actual set instead of one actual and one hypothetical set. The students can actually see how the part of data can contain the same information as the whole.

Since the students take the sample themselves, it gives them some kind of ‘personal ownership’ over their data. This is recommended to evoke productive class’ discussion (Weinberg et. al, 2010), especially when determining which group chart can represent the class chart the best.

In the third activity, the students pick one chart that is considered to be least representative. They are instructed to add more data points to the chart. The conjecture is that the students will see that the more data points they add to the group, the more it resembles the class chart. This activity is called ***growing sample*** (Bakker, 2012).

3. Lesson 3: Are girls taller than boys?

This lesson is designed to develop the need for sampling and the idea of randomization. There are two activities contained within this lesson. In the first activities, the students are asked to find out the typical height of the students in the whole school. The conjecture is that the students will run out of time or space, which will encourage discussion about a more effective way in finding information about a population.

In the second activity, the students are asked to judge a claim that contradicts their previous knowledge. Generally, the male students are taller than the female students. The context for this activities is a health report by the Head of the Students council, who collect his data in an unfair way that is the most convenient for him, i.e. by taking female students mostly from the higher grade. The claim made by the Head of the Students Council is that the female students are taller than the male students. The conjecture is that the students will reject the claim because the data is picked such a way that it produces said statement.

This kind of statement that trigger conflicts with the students' initial belief is encouraged because it encourages the students to look deeper into the problem and evoke meaningful discussion (Bakker & Derry, 2011).

4. Lesson 4: Social media addiction

The aim of this activity is for students to experience the process of inferential statistics, albeit informally. They conduct a mini-research, generalizing their result from sample to population, and use data to back up this statement while taking sample representativeness into account. They are expected to utilize the informal statistical knowledge they already learned during the course of the lesson.

The context for this lesson is social media addiction among grade 7. The students work in group to collect data from grade 7 and present it in a report. Each group then present their work in front of their peers, followed by feedbacks from other groups and class discussion.

CHAPTER III

RESEARCH METHOD

3.1 Research approach

The main goal of this study is to support 7th grade students in developing informal inferential reasoning. This goal also demands for innovation and improvement to be made to the conventional way statistics are being taught in secondary school. To achieve this goal, a series of instructional material intended for that purpose and an accompanying theory on how they work are developed. The instructional material is designed in such fashion so it can foster the emergence of students' IIR. To ensure its effectivity, the instructional materials needs to be implemented in an experiment setting that closely resembles the targeted group, and do so iteratively to adapt to whatever unexpected factors that might reveal themselves along the way.

To fulfill these requirements, a qualitative approach called design research is chosen. It is aimed specifically to develop theories about domain-specific learning as well as its accompanying instructional materials (Bakker & Eerde, 2013; van Eerde, 2013). More importantly, design research emphasizes not only on what educational design that works, but also on why, how, and to what extent it works. It is aimed to "give theoretical insights into how innovative ways of teaching and learning can be promoted" (Eerde, 2013, p. 2), which is in line with the aim of this study.

Design research is conducted in iterative cycles, with each cycle comprises three phases namely preparation and design, teaching experiment, and retrospective analysis (Graveimer & Cobb, 2006). The retrospective analysis is the most crucial phase where the previous phases are reflected. The knowledge of the researcher is the starting point of the cycle, and the result of the retrospective analysis contributes to the researcher's knowledge, that leads to a new cycle.

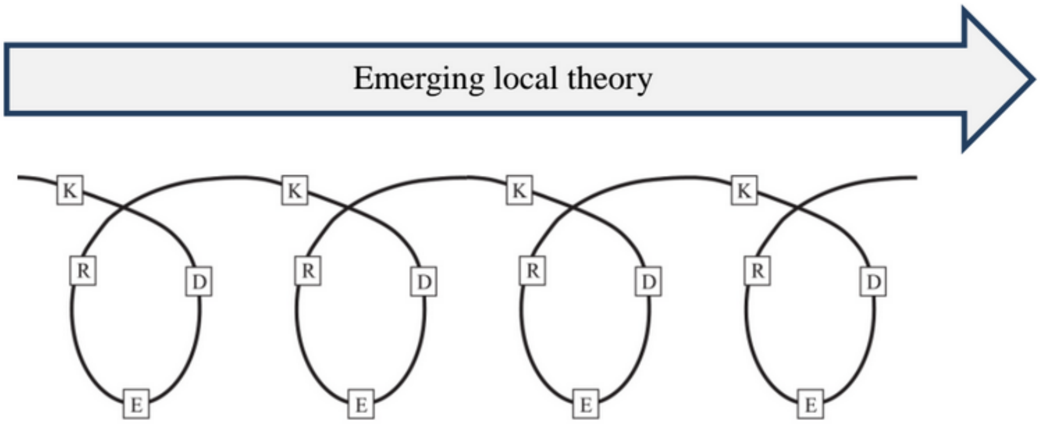


Figure 3.1. Cyclic process of design research (Gravemeijer & Cobb, 2006)

3.1.1 Preparations and design phase

Van Eerde (2013) mentioned three steps taken during this phase: (a) a literature review, in which the researcher studies the present knowledge about the topic; (b) the formulation of research aim and research question; and (c) the development of HLT.

In developing the HLT, in which the researcher start conducting a thorough literature review about the topic of the study, namely how it is portrayed in the curriculum and the textbooks as well as the gap between reality and expectations (Bakker & van Eerde, 2013). In response to the finding it entails, the researcher determines the research aim and questions, as well as developing the HLT (van

Eerde, 2013). This includes establishing learning goals and students' prior knowledge, as well as creates a set of tasks capable to bridge that gap and the accompanying hypothesized learning process.

To ensure this HLT works, aside from the pilot phase that is going to be explained in the next session, the HLT is also discussed with experts. What is expected at the end of the preparation phase is an elaborated HLT and a set of tasks, ready to be implemented in a classroom. The interventionist nature of design research allows the HLT to be modified along the course of the research (Bakker & van eerde, 2013), in order for the instructional materials as the output of the study is the best possible for the intended educational environment.

3.1.2 Teaching experiment

Steffen & Thompson (as cited in Bakker & van Eerde, 2013) stated that the primary purpose of a teaching experiment is to experience students' learning and reasoning at first hand, thus bridging the practice or research and the practice of teaching. In the teaching experiment phase, the HLT produced in a previous phase is tested to see whether or not it works in real life settings as opposed to being hypothetical.

The teaching experiment in this study consists of the pilot teaching experiment and main teaching experiment. The pilot phase is where the researcher tests the instructional materials to a small group of students in order to see how it works in real educational setting. Adjustments to the content and design of the HLT is made according to how the pilot implementation goes. This adjusted HLT is then conducted in the actual teaching and learning process in the main teaching experiment.

Prior to the teaching experiment, the researcher also needs to determine the type and the means of data collection, as well as to discuss the lessons with the teacher (Eerde H. v., 2013). The data collected typically include the students' written works, pre- and post-test, field notes, as well as audio and video registrations.

3.1.3 Retrospective analysis

Since the conjecture in HLT is a hypothetically thought to be how students' learning may occur, it makes sense that the analysis will be about comparing the hypothetical conjecture to the actual learning process. In this phase, the HLT is employed as a guideline to in the process of the analysis (Bakker & van Eerde, 2013). Based on the result of main teaching experiment, further adjustment is made to the HLT which can result in an additional cycle. The cycles stop when the criteria of this study as explained in the previous chapter is achieved.

The approach employed for retrospective analysis in this study is *task-oriented method*, where the HLT and the actual observed learning trajectory is compared by means of Dierdorp's analysis matrix (cited in van Eerde, 2013). This matrix is described as follows.

Table 3.1. Dierdorp's analysis matrix

Hypothetical Learning Trajectory			Actual Learning Trajectory		
Number of problem	Formulation of problem	Conjectures on students'answer	Transcript of video recording	Clarification	(optional) Quantitative expression on how well the HLT match the ALT

Note: Adapted from "Design research: looking into the heart of mathematics education" by H.A.A. van Eerde, 2013, *Proceeding The First South East Asia Design/Developmental Research (SEA-DR) International Conference (pp. 1-10).* Sriwijaya University.

The trustworthiness of a study depends whether or not it is conducted in a valid and reliable way. Each research approach demands specific way to assess its validity and reliability. Bakker & van Eerde (2013) explained in detail about the criteria for validity and reliability in design research setting. Validity refers to the soundness of the reasoning that has led to the conclusions, as in whether or not the data actually measures what we intend to measure. On the other hand, reliability refers to the independency of the study from the researcher.

Each of these criteria can be understood internally or externally. Internal validity refers to the quality of the data and the strength of the reasoning that leads to the final conclusion, while external validity means that the result has to be generalizable. Internal reliability meant that the data collection and analysis have to be independent of the researcher, while external reliability demands the result has to replicable in similar circumstances.

3.2 Data collection

This study employ various means of data collection. The process of data collection in each phase of the study will be described as follows..

3.2.1 Preparation phase

The means of data collection employed here is classroom observation and interview with teacher, and pre-test. The data are collected in the form of video and audio recordings, field notes, and students' written works. The result of this phase is the students' worksheet, teacher's guide, and the accompanying hypothetical learning trajectory.

Table 3.2. describes in more detail this part of the data collection.

Table 3.2. Type of data and method of data collection

Data	Method		
	Interview	Classroom observation	Pre-test
Teachers' belief on mathematics teaching	√	√	
Teaching methods	√	√	
Socio-cultural norms		√	
Classroom management	√	√	
Teacher's familiarity with Realistic Mathematics Education (PMRI)	√		
Teacher's understanding on statistics, inferential statistics, and their familiarity with IIR	√		
Students' prior knowledge			√

3.2.2 The pilot experiment

The first cycle of teaching experiment is conducted as the pilot of this study. The researcher acts as a teacher. The participants are a small group of students with heterogeneous mathematical ability. A recommendation from the teacher will be required in order to gain a sample of students with variation in their mathematical ability.

The aim for the pilot phase is to test the HLT and gain knowledge on how the design works. The data is collected by observation and by collecting the students' worksheet. The type of data collected is video and audio registration of the teaching experiment, field notes, and the students' written works.

3.2.3 The teaching experiment

In the teaching experiment, the HLT being implemented is the adjusted and corrected version of the HLT used in the first cycle. However, the size of the groups is significantly bigger since it is going to be conducted in the actual classroom. Instead of the researcher, the teacher will act as the teacher during the teaching experiments.

The data collection is essentially the same as those in the first cycle; it is done by observing the classroom and collecting the students' worksheets. The data are collected in the form of video and audio recordings, field notes, and written works of the students.

However, it is worth considering not to collect too many data since the data will be too overwhelming to analyze in a limited time (Eerde H. v., 2013).

Consequently, although the researcher will collect data from all the students, he or she will concentrate on a focus group of 4-5 students. The data from this group can be the focus of the analysis while the data from the rest of the students can be an additional information to support the findings related to the focus group. Since the size of the focus group is quite small, it enables the researcher to dig a little deeper for data by interviewing and discussing their work with them.

3.2.4 Pre-test and post-test

Although considered as the best way to approach problems in the field of education, design-based research, or qualitative research approach in general, it still cannot escape its fair share of critique. The critiques mostly concern the researcher's bias and the independence of the result. Barab and Squires (as cited in Terry & Shattuck, 2012, p. 12), for example, stated that, "if a researcher is intimately involved in the conceptualization, design, development, implementation, and re-searching of a pedagogical approach, then ensuring that researchers can make credible and trustworthy assertions is a challenge."

To bridge this problem, a pair of test is conducted in the beginning and the end of the teaching experiment as a means of data collection. In research, this is referred to as pre-test and post-test (see appendix). Both comprise similar problems with different contexts, designed to find out how the students make inference in a hypothetical situation.

The pre-test is conducted during the preparation phase and aimed to investigate the students' prior knowledge. The post-test, on the other hand, is conducted post teaching experiment and aimed to detect whether or not the students develop the intended informal inferential reasoning. From both pre- and post-test, the data are collected in the form of the students' written work.

Even though the qualitative data and analysis is the main focus of design research, the quantitative data and analysis can provide additional information. It can be used to support or test the findings related to the qualitative data and analysis.

3.2.5 Validity and reliability

During the data collection phase, validity and reliability of the study can be accomplished through various ways. Validity can be ensured by conducting a pilot prior the actual teaching experiment and employing multiple means of data collection (triangulation). The pilot phase of teaching experiment as well as the pre- and post-test before and after are also aimed to accomplish validity. The researcher can see whether or not the tests and the HLT actually measure and evoke the intended learning. The ecological validity, on the other hand, can be accomplished by having the actual teacher to teach during the teaching experiment also increase the validity.

The internal reliability can be achieved by using means of data collection that is independent of data collection, like video and audio registration. On the other hand, external reliability can be ensured by explaining the steps of data collection as clear and detailed as possible (traceability).

3.3 Data analysis

3.3.1 Pre-test and post-test

As explained before, pre-test are aimed to measure students' knowledge in regard to informal inferential reasoning, or lack of it. At the start of the teaching experiment, the prior knowledge required to participate in the lesson is also measured.

Pre-test result is analyzed by qualitative approach. The qualitative analysis is done by comparing the students´ answers to the framework proposed by Makar and Rubin (2009) as cited in the previous section of this paper. It will contribute to the adjustment made to the initial HLT, as well as to choose the focus group for the second cycle of learning experiment.

In the end of each learning cycle, the result from pre- and post-test are compared to see if the students develop the notion of informal statistical inference.

3.3.2 The pilot experiment

The data consist of video recording and students' written work. The analyzing process begins with the researcher watching the recording with research questions and HLT as guidance to ensure that the researchers will not miss any important fragments. Notes and transcript are made based on the interesting fragment observed. Then, the excerpts of the fragment are analyzed by means of task-oriented methods. As an additional data, discussion and interview with the focus group are also analyzed.

To create ideal assumptions about students' learning, researcher needs to look both for instances that confirm student's learning and instance that confirm the absence of it. This is an iterative process, meaning that whatever conclusion is generated by the researcher based on a certain fragment, it is to be tested at other fragments to see whether the conclusion pertains or needs to be modified. The result of the analysis is used to make necessary adjustment and modification the HLT.

3.3.3 The second teaching experiment

The analysis conducted for the data gathered in second cycle teaching experiment is essentially similar to those conducted for the data from the first cycle. The result of the analysis is used to answer research questions, generate conclusions, modify HLT and determine whether or not additional cycle of teaching experiment is needed.

3.3.4 Validity and reliability

Validity and reliability of the study is also affected by the data analysis. It can be elaborated externally and internally. Internal validity is accomplished by having more than one type of data (triangulation) and comparing the transcribed lesson episodes to the HLT. External validity, which concerns the generalizability of the study, is achieved by the use of contextual problems, which is hoped to enlighten how this theory might work if applied to similar contexts.

Reliability can also be elaborated into external and internal reliability. External reliability means that the study is replicable, and this is achieved by the ensuring the trackability of the research. To ensure trackability, the research process is described as detail as possible, so the readers can comprehend the steps of the research. Internal reliability, on the other hand, is accomplished by means of intersubjectivity, which means that the study is peer-reviewed by colleagues of the researcher and consulted to the expert.

The teaching experiment is conducted in SMP Lab Surabaya (Laboratory Secondary School of State University of Surabaya). After thorough discussion with the teacher, Mrs. Ira, it is decided that the first cycle is conducted in class VIIB while the second cycle in class VIIC. Based on their grades, these two classes have no significance difference in term of mathematical ability. The teacher, Mrs. Ira, also supported this by stating that students in SMP Lab is not categorized into classes based on their mathematical ability.

Six students, out of 27, participate in the first cycle, while 27 students participate in the second one. The complete research timeline is described in the following table.

Table 3.3. Timeline of research

Data Collection	Date	Participants
Classroom observation and the teacher's interview`	16 February 2015	The actual teacher Students of class VIIB
Pretest and students' interview	18 February 2015	6 students from class VIIB
Cycle 1		
Lesson 1: Scout Staff	18 February 2015	6 students from class VIIB
Lesson 2: Compare the Dots	25 February 2015	
Lesson 3: Are Girls taller than Boys?	4 March 2015	
Lesson 4: Social Media Addiction	11 March 2015	

Post-test and the students' interview	12 March 2015	
Pretest	31 March 2015	Students of class VIIC
Focus group students' interview	1 March 2015	4 students from class VIIC
Cycle 2		
Lesson 1: Scout Staff	1 April 2015	Students of class VIIC
Lesson 2: Compare the Dots	3 April 2015	
Lesson 3: Are Girls taller than Boys?	6 April 2015	
Lesson 4: Social Media Addiction	7 April 2015	
Post-test	8 April 2015	Students of class VIIC
Focus group students' interview		4 students from class VIIC

CHAPTER IV

HYPOTHETICAL LEARNING TRAJECTORY

As has been explained before, the students' thinking and learning process is as important as the learning goals that need to be accomplished at the end of the lesson. Simon (1995) stated that "A teacher may pose a task. However it is what the students make of that ask and their experience with it that determines the potential for learning." (p. 145). Fosnot and Dolk (2001) also stated that the students can go off in many directions as they explore, struggle to understand, and make sense of the world hypothetically.

Simon also emphasized the importance of a reflexive relationship between the designing process of the learning activities and the consideration of the thinking and learning that the students might engage in. The need to consider the learning goals, the activities, and the students' thinking gives rise to what known as thinking gives rise to what known as *hypothetical learning trajectory* (HLT).

Simon (1995) proposed the notion of HLT as the model of decision making the teacher must execute in the classroom. In addition, Simon described HLT as made up of three components: the learning goals that determine the direction, the learning activities, and the *hypothetical* learning process. The term hypothetical refers to the fact that it is only a prediction of how the students' understanding will evolve during the learning activities.

According to Simon (1995), the notion of HLT goes against the idea of having rigid learning goals or linear trajectory. Over the course of the lessons, the teacher's interactions with the students will bring new knowledge on the students' learning and thinking. This new knowledge results in the modification of the HLT, which changes the design of the learning activities. On the other hand, the change in the design of activities also determines prediction of the students' learning. The symbiotic relationship between the hypothetical learning process and the design of the learning activities is an essential part of HLT.

The aim of this study is to contribute an empirically grounded theory in supporting 7th grade students to develop IIR. To accomplish that, a set of instructional material is designed, consisting of students' worksheet and the accompanying teacher guides, as well as the accompanying HLT. There are 4 lessons which will be conducted over the course of 4 meetings. The instructional material will be tried out during the teaching experiment phase in Indonesian 7th grade classroom.

The instructional material and the HLT is the end product of the designing and preparation phase of design research. The HLT is previously constructed according to the result of literature review, and further improved by considering the result of the classroom observations and the teacher's interview. Therefore, overview of the classroom observations and the teacher's interview will be elaborated before the HLT. The learning trajectory, on the other hand, is depicted in Figure 4.1.

4.1 The overview of the classroom observations and the teacher's interview

Prior to the teaching experiment, a classroom observation followed by an interview with the teacher. The data is the teaching and learning process in the class where the second cycle or main teaching experiment would be implemented.

The topic happened to be taught during the observation is coordinate system. Since this topic is included in the curriculum but not included in the text book used by the school, the teacher had to provide the material herself. The observation gave light to the following information. The teacher started the lesson by asking the students to read the current topic quietly for a few minutes, followed by an introduction to coordinate system. Some students did the problems provided by the teacher as an example, in front of the class.

After the introduction, the teacher distributed the worksheet among the students followed by instruction. There was no group work and the students were instructed to work individually. The teacher left the students alone yet they worked quite well by themselves. The teacher explained that due to work reason, she had to leave the class and let the students do the investigation themselves.

After a period passed by, the teacher was back and invited the students to correct their work together. For each problem, the teacher asked one of the students to do it in front of the class and then listed the correct answers on the blackboard. The teacher mainly led the discussion from the front of the class over the course of the lesson. The discussion was quite straightforward and the teacher did not challenge the students for more sophisticated answer or different ways to get there.

In general, the teacher's way of teaching was quite conventional. She spoke loud and clear hence the students were engaged with the lesson. She was quite responsive with the students' response and curiosity in learning, by improving or making the assignments more difficult if the students can solve said assignments easily. She was also not strict when it comes to following text-book.

Another interesting insight is the classroom's social norm. To ensure the students working quietly and did not disturb the classroom condition, the teacher established a rule where the students had to pay an amount of money whenever they leave their table.

As for the teacher themselves, her name is Ira Dwi Agustina, S.Pd., known by her students as Mrs. Ira. She is graduated in mathematics education. She has been teaching mathematics for 7th grade students for approximately 5 years.

The interview confirmed the fact that Mrs. Ira was overwhelmed with the demand of the latest curriculum; the curriculum 2013. She expressed concern about the dense formation of curriculum 2013, especially the fact that Statistics in this curriculum is introduced earlier than in the previous one. Aside from collecting data, the learning goal for the students is also being able to represent data in various charts and graphs.

The demanding curriculum also cause the limited time allocated to be an obstacle. Even though she was personally in favor toward group work and project-based lesson, there is practically no room for student-centered activities in the classroom.

As for *Pendidikan Matematika Realistik Indonesia*, which is Realistic Mathematics Education in Indonesian context, Mrs. Ira claimed to be only familiar with it as an approach in elementary school mathematics. She has never been educated in it nor applied it in her classroom. She also stated that she thinks of

Statistics as a subject that is mainly about number and formulas and does not need much exploration.

By considering the result of the classroom observations and the teacher's interview, the HLT and the teacher's guide are adjusted as follows.

Prior to the teaching experiment, a classroom observation followed by an interview with the teacher. The data is the teaching and learning process in the class where the second cycle or main teaching experiment would be implemented.

The topic happened to be taught during the observation is coordinate system. Since this topic is included in the curriculum but not included in the text book used by the school, the teacher had to provide the material herself. The observation gave light to the following information. The teacher started the lesson by asking the students to read the current topic quietly for a few minutes, followed by an introduction to coordinate system. Some students did the problems provided by the teacher as an example, in front of the class.

After the introduction, the teacher distributed the worksheet among the students followed by instruction. There was no group work and the students were instructed to work individually. The teacher left the students alone yet they worked quite well by themselves. The teacher explained that due to work reason, she had to leave the class and let the students do the investigation themselves.

After a period passed by, the teacher was back and invited the students to correct their work together. For each problem, the teacher asked one of the students to do it in front of the class and then listed the correct answers on the blackboard. The teacher mainly led the discussion from the front of the class over the course of the lesson. The discussion was quite straightforward and the teacher did not challenge the students for more sophisticated answer or different ways to get there.

In general, the teacher's way of teaching was quite conventional. She spoke loud and clear hence the students were engaged with the lesson. She was quite responsive with the students' response and curiosity in learning, by improving or making the assignments more difficult if the students can solve said assignments easily. She was also not strict when it comes to following text-book.

Another interesting insight is the classroom's social norm. To ensure the students working quietly and did not disturb the classroom condition, the teacher established a rule where the students had to pay an amount of money whenever they leave their table.

As for the teacher themselves, her name is Ira Dwi Agustina, S.Pd., known by her students as Mrs. Ira. She is graduated in mathematics education. She has been teaching mathematics for 7^{th} grade students for approximately 5 years.

The interview confirmed the fact that Mrs. Ira was overwhelmed with the demand of the latest curriculum; the curriculum 2013. She expressed concern about the dense formation of curriculum 2013, especially the fact that Statistics in this curriculum is introduced earlier than in the previous one. Even though she was personally in favor toward group work and project-based lesson, there is practically no room for student-centered activities in the classroom.

As for *Pendidikan Matematika Realistik Indonesia,* which is Realistic Mathematics Education in Indonesian context, Mrs. Ira claimed to be only familiar with it as an approach in elementary school mathematics. She has never been educated in it nor applied it in her classroom. She also stated that she thinks of

Statistics as a subject that is mainly about number and formulas and does not need much exploration.

4.2 The Hypothetical Learning Trajectory

Based on the observation and interview, some revisions are done on the pre-HLT, resulting in the HLT implemented in the first cycle of teaching experiment.

In the first activity in the first lesson, the students only have to make one type of graph, which is dot plots. However, due to the teacher's advice to make the students' worksheet more in line with the curriculum, more type of graphs were added. Hence, the students are not only expected to present the data in dot plots, but also in a frequency distribution table, line graph, and bar chart.

In Lesson 2, the students are given quite an open instruction in comparing all the group charts. Basically they were only asked which group chart can represent the class chart the best. Due to the nature of the class discussion, it might be better if the instruction is not left completely open to give the students a sense of guidance. Therefore, the students are given a table to assist them in comparing the all the charts. This table provide characteristics based on which the students have to compare the group charts and find out which one is more suitable to represent the class chart.

	THE CLASS CHART	GROUP CHART 1	GROUP CHART 2	GROUP CHART 3
What is the height of the shorthest student?				
What is the height of the tallest student?				
The typical height of the studen is ... cm				
What is the average height of the students?				
Which one is taller, girls or boys?				
Is there anything more that you notice from the charts?				

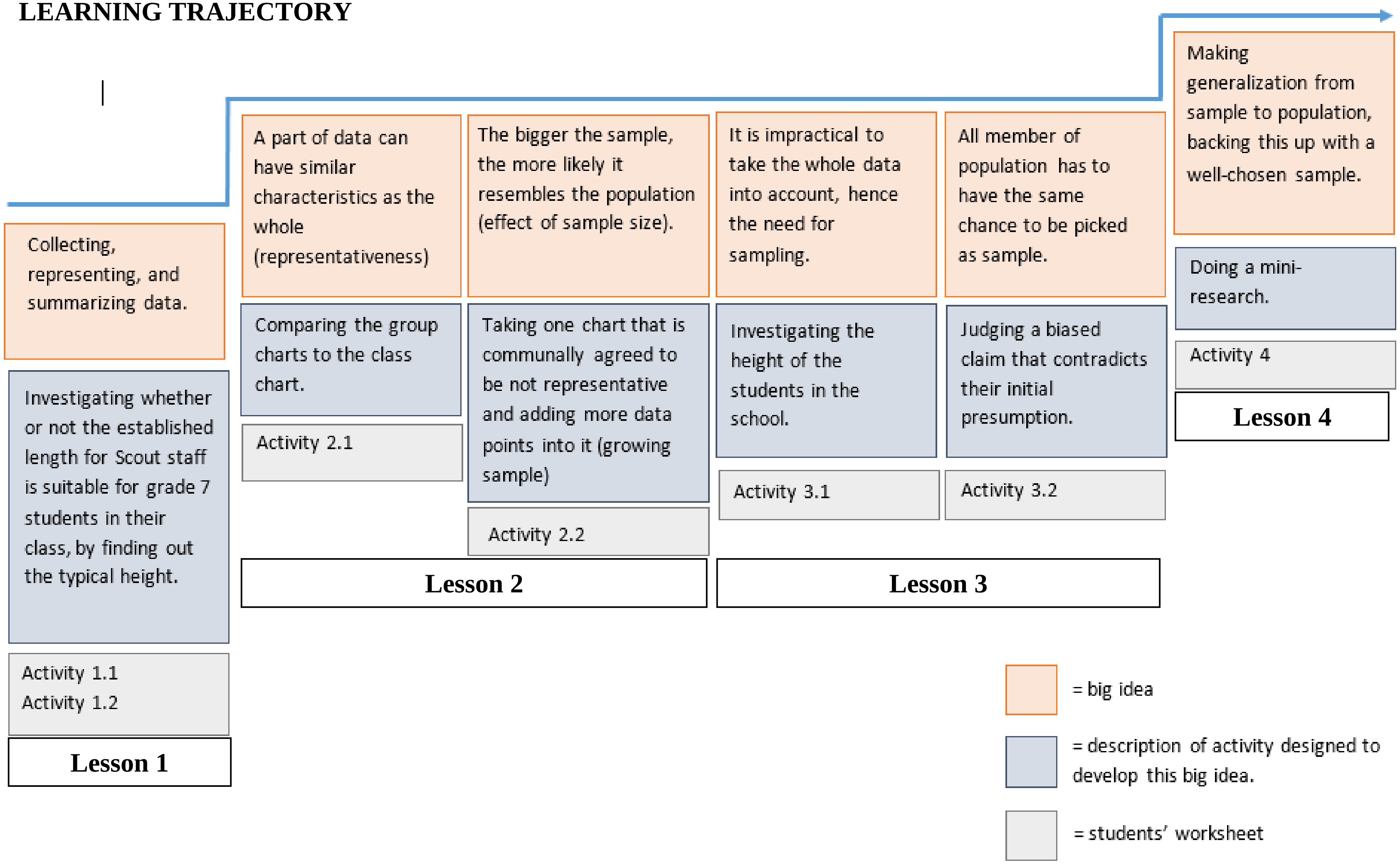
Figure 4.1 The planned table for the instruction in activity 2.1

There are also 3 activities in Lesson 2: activity 2.1 where the students compare the class chart to their own group chart; activity 2.2 where they compare all the group charts to the class chart; and activity 2.3 where they do the 'growing sample' activity. Due to the limited time and considering the fact that both activities are almost similar, activity 2.1 is removed.

Generally, since there are only 6 participants of the first cycle, hence they cannot collect data from their own class. Therefore we provide data from a virtual classroom. All charts are no longer put in the worksheet and put on separate sheets instead. All problems become less open-ended and the students are given more guidance.

The detail of the complete HLT is elaborated as follows.

LEARNING TRAJECTORY



4.2.1 Lesson 1: Scouts out and about

4.2.1.1 *Starting points*

Students have been introduced to statistics in 6th grade, especially data collection and arithmetic mean. Students have not been introduced to the concept of sample and population. Before participating in the lesson, it is assumed that students are able to

- add, subtract, multiply and divide rational numbers in all representativeness;
- perform arithmetic calculation to solve problems;
- collect, organize and represent data;
- find the measure of central tendency of a group of data;
- calculate arithmetic mean of a group of data; and
- measure length using metric measurement

4.2.1.2 *Learning goals*

The aim of this lesson is for students to recall the concept of data as well as data collection, representation, and summary that they received in primary school.

After completing this lesson the students will be able to

- collect data and put them in an inventory (list or table);
- represent data in the form of dot plot; and
- summarize data, either by using measures of central tendency, measures of dispersion, or relation between variable.

4.2.1.3 *Description of activity*

The students' background knowledge in statistics consist merely of descriptive statistics, so they are going to utilize this to find the typical value from a set of data. Hence, they need a context where they will be inclined to find the typical value. The context chosen for this lesson is finding whether or not the length of Scout staff is suitable for 7th grade students.

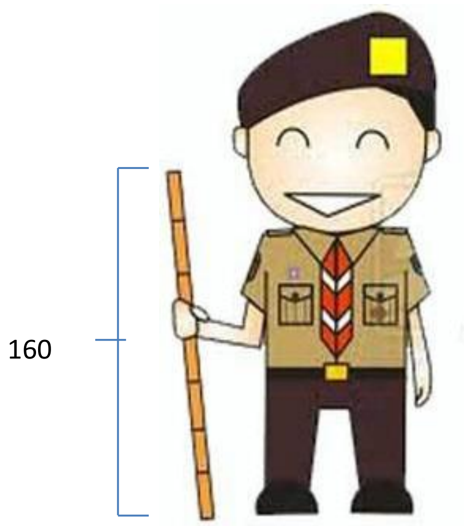


Figure 4.1. Illustration of the staff of Indonesian Scout

In Indonesia, the Scout staff has an official length that has been established by the Board of Indonesian Scout, which is 160 cm. To find out whether or not this length is suitable for 7th grade students, firstly they will need to know the typical height of the students and then compare it to the length of the Scout staff. They are going to start with 7th grade students in their class.

The lesson consists of two activities. In the first activity, the students collect data from a small group of students in the class, represent it in the form of dot plot, and summarize it. The second activity is similar to the first, only with a bigger set of data. The students collect data from all students in the class, represent it in the form of dot plot, and summarize it as well. The idea of the first and second activity is to give the students concrete idea in dealing with sample and population. Real life data is used here to make the problem more meaningful for students.

Even though the data collected will be from the students in the class, due to the limited time they are not going to literally do measurement during the lesson. It is best for the students to be prepared with the information of their own height. Hence one day or two before, the teacher will have to ask the students to measure their height at home.

Throughout these two activities, the students will be working in a group of 3 or 4 students. This arrangement will last until the end of the lesson sequence; a rule the teacher should inform the students about once the groups have been established.

During the pilot phase, the researcher will be working with a small group students. The number of the students will be insufficient to make a concrete distinction between sample and population. Therefore, the students simulate the process of data collection by taking data from *population bag*. A population bag is a bag containing small pieces of paper with the height and the sex of each student in the class written on it.

First activity

The teacher starts the lesson by setting up the context about Scout staffs. The teacher can ask whether or not the students know the length of their own staff, and that all Scout staffs in Indonesia have the same measurement (160 cm) because the Board of Indonesian Scout establishes it that way, which hopefully will lead to the need to find out the typical height of 7th grade students. To simplify it, the students are going to investigate the height of 7th grade students in their own class first.

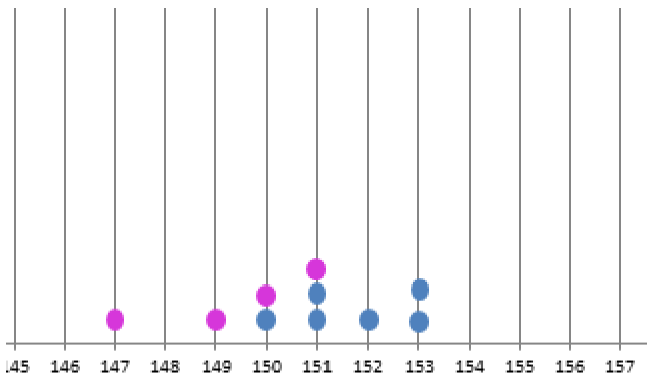


Figure 4.2. Example of group chart.

In this activity, the students work in groups. The teacher distributes worksheet Activity 1.1, a pink marker, a blue marker, and an A3 paper with a dot plot scale on it to each group. Even though dot plot is quite simple and its construction is quite straightforward, the teacher should set aside time to explain how to construct a dot plot. The teacher also needs to establish that pink dots represent girls, while blue dots represent boys.

Each group will gather data from different number of students; for example, group 1 may collect data from 12 students while group 2 collect data from 16 students. The instruction about the number of data to be collected will be provided in the worksheet. The students write down the data in the table and present them in the form of dot plots. From this point onward, this will be referred to as the ***group chart***. Then the students discuss and summarize their dot plot. Moving beyond descriptive statistics, the students are also asked to predict what the graph of the whole class will look like.

Afterward, the teacher asks 1 group to present their dot plot and data summary in front of the class. The other groups compare them with their own findings. In the end of the activity, the teacher asks the students whether or not they agree with the presenting group's prediction of the chart of the whole class.

Second activity

The second activity is quite similar to the first one, except that in this one, the students work together to create a dot plot using data from the whole class. The students will keep working in groups.

Prior to the lesson, the teacher prepares an A3 paper with a horizontal scale on it and hang it on the board. The teacher starts the activity by distributing blue and pink stickers among the students; blue stickers for boys, pink stickers for girls. Then each student sticks the sticker to the scale, depends on their own height. In the end they will have a dot plot that represent the height of students of the whole class. From this point onward, this will be referred to as the ***class chart.***

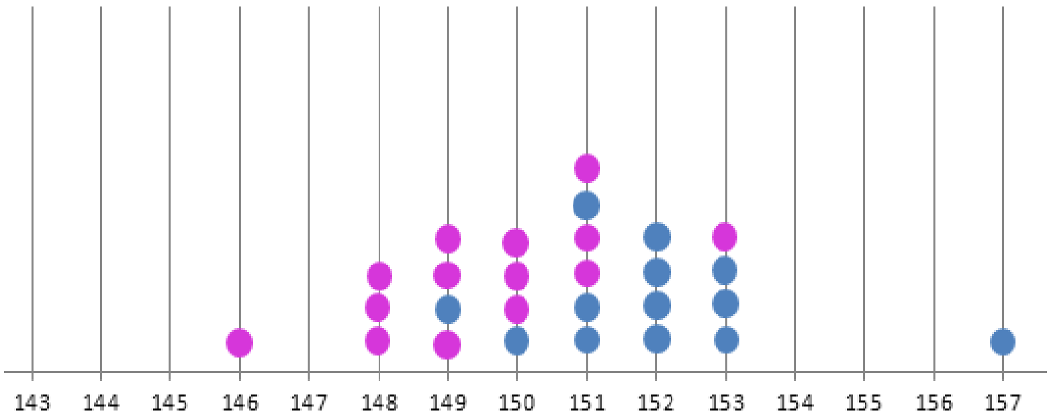


Figure 4.3. Example of the class chart

Afterward, the teacher distributes the worksheet for activity 1.2 and the A3 paper for class chart to each group. The students work in groups to copy the class chart to the scale on the A3 paper. They discuss the class chart and summarize it, similar to the activity 1.1. In the end, they go back to the problem in the beginning of the class and discuss whether or not the length of 160 cm for a Scout staff will be a suitable measure for 7^{th} grade students.

There is no presentation. Instead, the teacher conducts class discussion where the students summarize the class chart together. In the end of this lesson, the students will produce two dot plots; the group chart and the class chart. The purpose is to give them the concrete idea of sample and population.

4.2.1.4 *Conjecture of the students' reaction*

- The students summarize data visually. They notice the dot at both ends of the dot plot and mention the tallest and the shortest students. Their attention can also go to where the majority of the dots are stacked (modal value). They might also notice the difference between blue and pink dots and connect it to the difference between male and female students' height. Example of students' statements are:

- *Most of the dots are stacked at 148 cm, so most students are probably of that height.*
- *The shortest student is 140 cm, while the highest is 158 cm.*
- *This dot is stand alone on here, the person might be really short.*
- *There are more pink dots than blue dots on the left, so girls are taller than boys.*

- In finding the typical, the students might analyze the chart visually and use modal value as typical height. They might get confused if the distribution of data is not a perfect bell shape, which means that there are outliers or more than one value with dominant frequency. Some of the students might analyze the chart numerically and use arithmetic mean. Some might unable to find the typical. Example of students' statement are:
    - *148 cm, because most of the dots are at that height.*
    - *Ummm its between 146 and 150 cm so I guess that´s what the typical height is.*
    - *I can't do it, the dots are too scattered.*
- The students might get confused in predicting the chart for the whole class. They might use the characteristics of their group chart and use that to predict the class chart. Hence they might use the same maximum and minimum or typical value, and fill up the rest of information on their own.
- In concluding whether or not the height of the Scout staff is suitable to 7th grade students, the students probably use the typical height and compare it to the length

of the staff. Some students might also consider the maximum and minimum value (or outliers, if present).

4.2.1.5 *Discussion*

It is important to note that the students may not have time to do actual measurement in the class, therefore the teacher is advised to ask the students the previous day to measure their height at home. The teacher should also prepare a measuring tape in case someone forgets to measure their height. The teacher also has to set up a dot plot axis for the students to stick their sticker to.

Aside from the practical matters, there are other important roles for the teacher during this lesson. In order to make data analysis meaningful, the students need to be engaged to the context. Hence, the role of the teacher is to engage them when she or he sets up the context. It is advisable to bring up real life matter or ask students 'experience. Some of the statement or question the teacher can use is:

- *When I was a Scout I always think that it isn't fair that all Scout staffs are of the same length. What do you think?*
- *Can you use your Scout staff comfortably? Do you think it's too long?*
- *Do you think the class next door also finds the Scout staff uncomfortable to use?*
- *What is the length that you recommend?*

Even though the students are already introduced to the algorithm of arithmetic mean in primary school, they are not pressured to use it in this lesson. The teacher's role is to take the focus away from the formula and back to its role as a measure of central tendency. The point is for the student to be able to summarize their data, and not to calculate the mean.

Since backing up statement with data is one of the goal of IIR, the teacher has to emphasis the use of data to back up every of students' statement. Therefore for every answer provided by the students, the teacher has to ask them to back it up with their data, if they have not already.

The teacher also has to pay attention to the informal words the students might come up with that stand in the place of formal statistical concept, like *bump, stacked, shortest/highest student*, etc.. It is important for the teacher to revoice these words every time they come up, and write them down on a special section on the board.

Lastly, teacher's most important role is during the discussion. In activity 2.1, since every group is working with different subset of data, there are a lot of different answers and opinions. The teacher has to bridge these differences by emphasizing that there is no one right answer; each answer can be correct as long as its backed up by data. In activity 2.2, when the students are summarizing the class chart, compare it with their prediction in the previous activity and ask them about why the prediction is correct and why not.

4.2.2 Lesson 2: Compare the dots

4.2.2.1 *Starting points*

In this lesson, the students start to move beyond descriptive statistics to inferential statistics. Even though they have no conception yet about what sample and population are, they will be exposed to a concrete idea of it by investigating a set of data, a subset from that set of data, and comparing the two. Prior to this lesson, it is assumed that the students are able to

- collect data and put it in an inventory (list or table); and
- summarize data visually or numerically, either using measures of central tendency (mean, median, modus), measures of dispersion (minimum and maximum value, or spread), and relation between variables.

4.2.2.2 *Learning goals*

The aim of this lesson is to develop the idea of sample representativeness and sample size. After completing this lesson students will be able to

- compare two sets of data in the form of dot plot using the summary of the data;
- identify the similarities and the differences between the summary of two dot plots;
- recognize that a subset of the data (sample) can have the same characteristics as the whole data (population);
- recognize that the bigger the sample, the more likely it resembles the population; and
- explain the effect of sample size.

4.2.2.3 *Description of activity*

This lesson consists of 2 activities. In the first activity, the students compare and contrast all the group charts and determine which one represents the class chart the best. In the second activity, the students single out a group chart that is communally agreed to be not representative and growing it.

First activity

The teacher starts the lesson by setting up the context. Still in the context of height measurement, the teacher tells an imaginary situation in which the class chart is gone and only the group charts remain. The teacher then leads the discussion to whether or not the students will be able to get the information about the height of the students of the whole class, only from their group chart.

The teacher asks the students to investigate what information about the height of the students in the class that they can retrieve from the group chart. The teacher then announces that they are going to do a competition between group charts to see which one can represent the class chart the best. The students get a copy of all the group charts and work in group to compare and compare them to the class chart. In the end, all the groups vote which group chart represents the class chart the best. The group chart who wins presents their chart in front of the class, followed by a class discussion.

Second activity

In the previous activity, students may or may not notice that the group chart with bigger size of data tend to easily follow the characteristics of the class chart. The third activity, is designed to foster this idea. The students are given one group chart that is communally agreed to “lose” in the competition held in the previous activity, add more data to this group chart, and formulate their conclusion about what happens when the size of the data is getting bigger and bigger.

Since it will be impractical to collect the data all over again, each group is given a population bag containing the data of the whole class. The students take data values form this population bag, put them into the chart, and examine the group chart as its grow bigger and bigger.

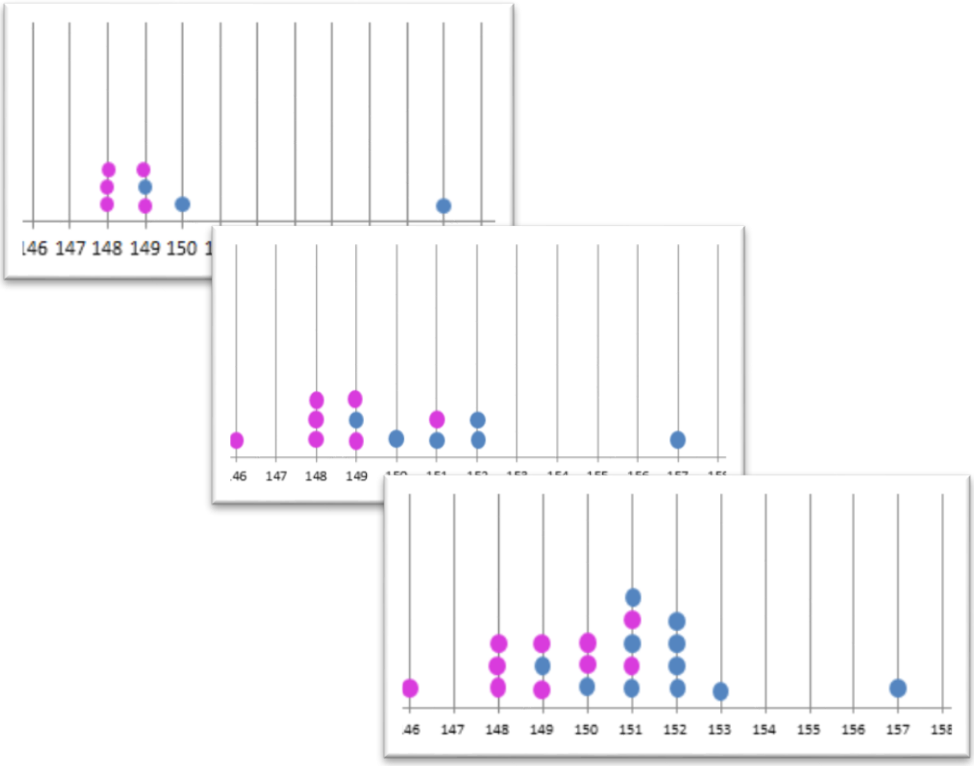


Figure 4.2.1 Example of growing sample activity.

4.2.2.4 *Conjecture of the students' reaction*

In the first activity, the students might analyze the data visually. They already summarize group chart and class chart before. Therefore in this activity, they probably use these summaries and find out which summaries are similar or which are different between the two charts. It also possible that they simply use the shape to compare the two charts. Some students might have difficulty in comparing the class and group chart because the two have different number of dots. Other might choose to use arithmetic means.

In the second activity, the students might stick to visual analysis. It is very possible that they will not use arithmetic mean since it will be complicated to calculate every group’s means and compare it to the class’ mean. Some students might find that more than one group chart can represent the class chart, and they insist to have more than one winner in the competition. Before they already use data summary to compare their own group chart to the class chart, but now that there are more than one group chart of various size of data, it is very possible that the students notice this and consider sample size as one of the factor that effect sample representativeness.

The third activity is the least complicated of the three. Students can see that the more data they insert into the chart, the more the group chart will resemble the class chart. It is very likely that they can arrive into the conclusion that bigger samples are more representative.

4.2.2.5 *Discussion*

As in the previous lesson, the teacher also has to pay attention to the informal words the students might come up. Since the concept of representativeness is very new to the students, the teacher needs to emphasis the fact that the small data “represents” the big data. Statistics is one of the mathematics domain that has quite an extensive specific vocabulary and the students need to be introduced to them. Words like ‘represent’ is one of the informal words that the teacher needs to emphasis and probably to write down on the board.

Teacher also has to emphasis the use of data to back up every of students' statement. Therefore for every answer provided by the students, the teacher has to ask them to back it up with their data, if they have not already.

When the student are comparing group chart and class chart and they find it difficult because the two have different number of dots, give a guiding question such as: *What information can you get from the class chart? Now can you find it in the group chart?*

The teacher's most important role is during the discussion In activity 2.2, all groups probably have different answer about which sample represents the class chart better. The teacher needs to bridge these differences by emphasizing that there is no one right answer and that what matters is how they can support their argument by basing it on their data.

4.2.3 Lesson 3: Are girls taller than boys

4.2.3.1 *Starting points*

In this lesson, students move deeper into inferential statistics. In the previous lesson, they already explore the notion of sample representativeness and sample size. Therefore prior the beginning of this lesson, it is assumed that the students are able to

- summarize dot plots, either using measures of central tendency (mean, median, modus), measures of dispersion (minimum and maximum value, or spread), and relation between variables;
- compare two sets of data in the form of dot plot using their data summary;
- know that a subset of data can represent the whole data; and

- know that the bigger this subset, the more likely it resembles the whole data.

4.2.3.2 *Learning goals*

This lesson is designed to develop the need for sampling and the idea of randomization. After completing the lesson, the students will be able to

- making a conclusion about a big set of data by taking only a subset of it;
- explain why sampling is needed;
- detect bias in data collection; and
- explain the idea of randomization; that every member of population has to have the same chance to be picked as sample.

4.2.3.3 *Description of activity*

This lesson consists of 2 activities. The first activity is designed for the first goal, where the students are tasked to find out about the typical height of the students of the whole school. The second activity is designed for the second goal, where the students are tasked to investigate a claim about the height of the students of the whole school.

First activity

The first activity is still in the context of height measurement, only this time the students deal with a much bigger set of data. The students work in group to investigate the typical height of the students in the whole school. Each group is given a population bag, which, as has been explained before, is a bag containing small pieces of paper with the height and the sex of each student in the school written on it. Therefore, each piece of paper represents one student and the whole bag represent the population of the school.

Figure 4.2.2. The population bag

These data are fictional, and there is no distinction between grades. The only variables are gender and height. The students present this data in the form of dot plots, summarize it, and attempt to find the typical value of the students in the school. In the end, they produce what will be referred to as ***the school chart.*** This dot plot is significantly different from the other two dot plots because (a) the data source is significantly bigger, and (b) not all data value is portrayed in the dot plot.

Second activity

The second activity is still in the context of height measurement. But instead of the students making a conclusion about the typical height of the students in the school, they are provided a claim made by a fictional character about it.

The students are also provided about the process of data collection that leads to this claim, which is a map with the area of data collection shaded. The area of data collection is specifically made to be not fair; the fictional character only gather his data from 1st grade students and the member of girls' volley club. His claim is that "*girls are taller than boys.*"

The students are then work in group to investigate whether or not this claim is acceptable.

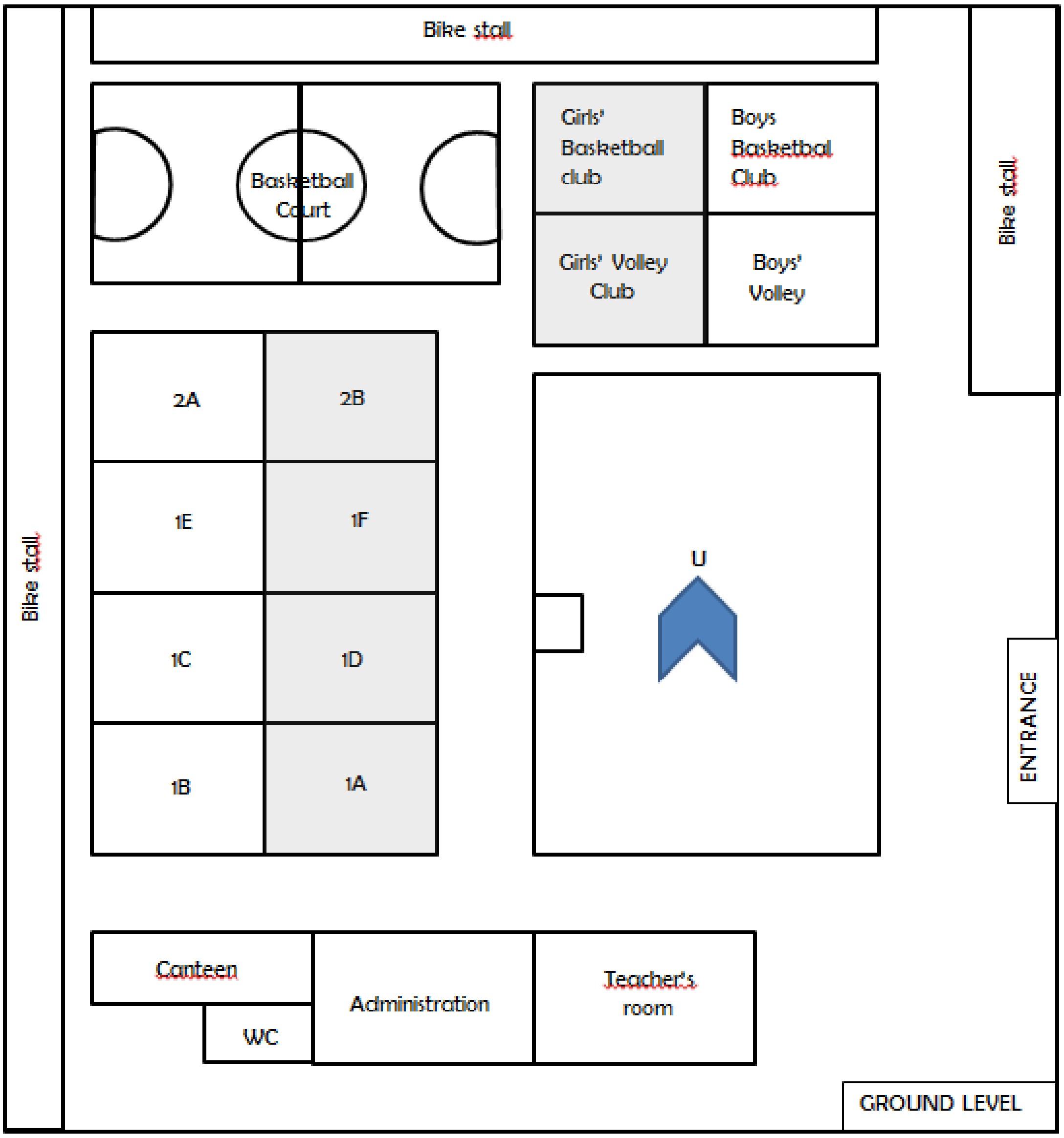


Figure 4.2.3. The map of the biased data collection.

4.2.3.4 *Conjecture of the students' reaction*

During the first activity, the first possibility is that the students try to put all the data into dot plot until they run out of space or realize that they are wasting time.

Prior entering this lesson, the students already learn the idea that the bigger the sample, the better. Therefore some students might employ that notion and already plan to take only a subset of data from the population bag. They, however, are only concerned about the size and try to take as many data as possible that they can fit in the dot plot.

It is also possible that the students pick the data. They probably take the data one by one randomly first, and then realize than the data that they take does not seem like a representative data; it probably contains too many girls or the students are mostly tall. Hence, they decide to be selective about what value they insert into the dot plot by putting back the value that they do not want and actively digging the population bag to find the value that they want. They have this idea in their head about what the population will look like, and they try to fit into that idea.

Some students might take some data randomly, by employing the idea that a subset of data can have the same characteristic as the population. However it is very possible that they do not have any reasoning about this random selection.

The second activity is designed to contradict the students presumption. Most students probably assume that the boys are taller than girls. Given a claim that *"girls are taller than boys*" and the stark pickiness in the way of the fictional character collect his data, it is very possible that the students will come up with the idea of bias or random. Although they probably do not say these words yet and use informal equivalent words instead. The example of students' statements are the following.

*He is really **picky**, it is only grade 6 students and the girl from volley club.*

- *It is not **fair**, of course the girls are going to be taller! I mean, they are from volley club.*
- *I bet his claim will be different if he takes the data from the boys in volley club too.*

4.2.3.5 *Discussion*

As in the previous lesson, the teacher also has to pay attention to the informal words the students might come up. The teacher needs to revoice these words and write it down on a special statement on the board. The teacher also has to emphasis the use of data to back up every of students' statement. Therefore for every answer provided by the students, the teacher has to ask them to back it up with their data, if they have not already.

The most important role for the teacher is leading the discussion. All the possible students' reactions for the first activity offer their own point of discussion that can open doors to an interesting debate. If the students try to put all the data from the population bag into the chart, remind them about the previous lesson, about how they can get information about a set of data from a subset of it. Questions the teacher can ask are, for example *putting in all the data seems like a lot of work, isn't it? Do you remember about what we did last meeting?*

If they try to pick their data, ask what their reasoning to do so. The students might explain their idea of an ideal chart, that they believe will represent the population the best. Challenge this idea with the fact that they have no idea at all about all the data in the population bag, unlike in the previous lessons. Questions the teacher can ask are, for example, *Oh, so you think that the data you take so far seems weird. How do you know that?*

Meanwhile, if they try to take a small subset of data randomly, ask their reasoning to do so and how they can be confident that the data they take will represent the population. Questions the teacher can ask are, for example; *How about the data that you left out? Look at the tallest students in the data that you picked already; what if there are a much taller student you left in the population bag?*

It is important to note that until this point, the students still have not been introduced formally to the words "sample" and "population". Therefore it is advised for the teacher to not use these words and stick with the informal words the students use.

In the end of this lesson, the teacher set the context and assign the task for Lesson 4. The last activity is a mini research where the students everything they have learned so far to conduct an investigation involving data from the whole school. Since this activity is quite big, the allocated time for one meeting is considered to be not suitable. The teacher assign the project in the end of Lesson 3 instead, so the students can do it in their free time. In Lesson 4, the students will present their result to their peers.

4.2.4 Lesson 4: Social media addiction

4.2.4.1 *Starting points*

Prior to this lesson, it is assumed that the students are able to

- summarize dot plots, either using measures of central tendency (mean, median, modus), measures of dispersion (minimum and maximum value, or spread), and relation between variables.
- know that a subset of data can represent the whole data.
- know that the bigger this subset, the more likely it resembles the whole.
- know why sampling is needed.
- know that every member of the population has to have the same chance to be picked as sample.

4.2.4.2 *Learning goals*

The aim of this activity is for students to experience the process of inferential statistics, albeit informally. They conduct a mini-research, generalizing their result from sample to population, and use data to back up this statement while taking sample representativeness into account. They are expected to utilize the informal statistical knowledge they already learned during the course of the lesson. Therefore, after completing this lesson the students will be able to

- produce a statement that indicates a generalization from sample to population;
- use data to back up this statement, taking into account the representativeness of the sample; and
- state it with a degree of uncertainty.

4.2.4.3 *Description of activity*

This lesson consists of 1 activity, which is the mini-research. Since this activity is quite big and requires quite some time to finish, the teacher assigns them in the end of lesson 3 so the students can work on it in their free time.

The context used in this activity is social media addiction. The students are presented with a statistics about social media addiction among teenagers in Indonesia, quoted from a local newspaper. They then attempt to investigate whether or not this claim applies to the students in the school. They work in group to do a survey with the students of the whole school as the source of data, as well as collect, organize and analyze their data. In the end, they present the result of their research in the form of a project report.

This activity is very open ended. The students are not provided with data or population bag like in the previous lesson. The students collect the data themselves, present them (still in the form of dot plots), and report their findings. The only thing that is not very open-ended is that the students are given a ready-made report form that they only have to fill in, so they do not have to write the report themselves.

The students conduct the research during their free time between Lesson 3 and Lesson 4. In lesson 4, they only have to present their findings. Each group gets an opportunity to present their finding, while the other groups give feedbacks.

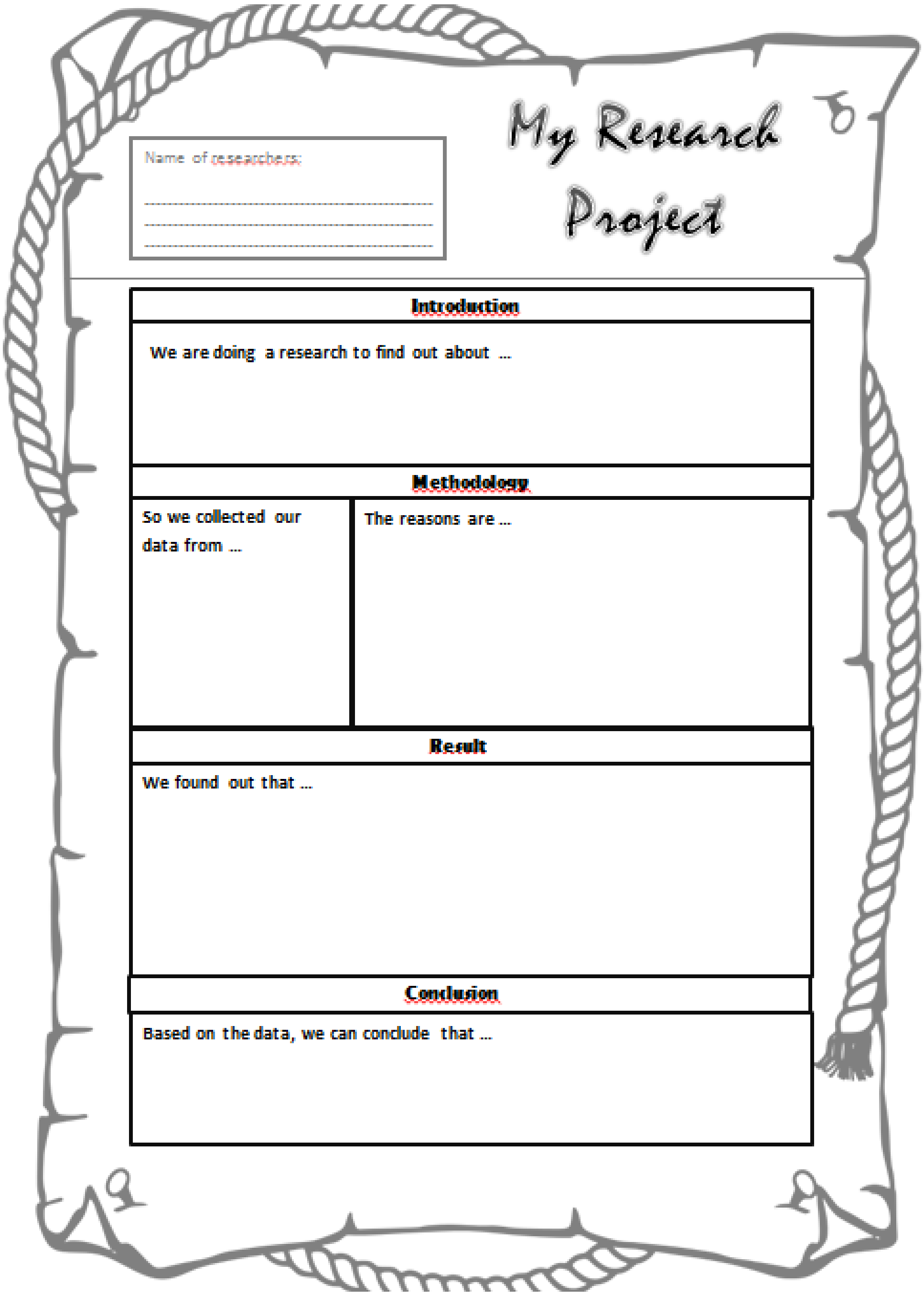


Figure 4.2.4 The layout of research project report

4.2.4.4 *Conjecture of the students' answer*

Students are predicted to not collect data from the whole school and instead choose a sample instead. Students' report, especially their methodology and answer may varies, depends on the sample that they choose. Since the students will use real life data, the result of the research will depends on the data from the school.

4.2.4.5 *Discussion*

This activity is designed to give an idea about scientific research and how statistical conclusion is reached. As in the previous lesson, the teacher also has to pay attention to the informal words the students might come up. The teacher needs to revoice these words and write it down on a special statement on the board. The teacher also has to emphasis the use of data to back up every of students' statement. Therefore for every answer provided by the students, the teacher has to ask them to back it up with their data, if they have not already.

Also it has to be noted that the students still have not been introduced formally to the words "sample" and "population". Therefore it is advised for the teacher to not use these words and stick with the informal words the students use. In the end of this lesson, which is the end of the whole sequence, the teacher can gather all the informal words and concepts that have been gathered during the course of the lesson, discuss it with the students, and introduce the formal equivalent of these words.

During the discussion, it is very important for the teacher to lead the discussion and bridge the differences between students. This is a very open ended task and all the groups work with different sample, therefore they will came up with different answers. The teacher has to emphasize that this is okay; that this is the problem that real statisticians also has to face in the real world. The most important thing is how strong their argument is to defend their conclusion, which demands the utilization of all the informal statistical concepts they already learn in the previous lesson. Therefore the discussion should be centered around

Each group can come up with different inference, and what make these inferences different from each other is how strong their argument in defending them. Center the discussion around

- their methodology, as in step by step explanation in how they are doing their research;
- the informal statistical ideas that they learn from the previous lesson (representativeness, effect of sample size, and randomization) and how they use it in their research.
- the result as in the dot plot and data summary; and
- the final conclusion.

Questions that the teacher can ask, for example;

- *Why do you think the other group get different conclusion?*
- *Did you collect your data in a fair way? How? Remember, we talked the importance about not being picky, right?*
- *The other group collect their data from a much bigger number of person. Do you think it is good? What if it is big but it isn't fair?*
- *How sure are you that your conclusion represent all the students in the whole school?*

CHAPTER V

RETROSPECTIVE ANALYSIS

To strengthen the predicting power of the HLT and ensure that the instructional material evokes the desired reasoning (IIR), the material and the accompanying HLT need to be tested in an episode of teaching experiment and then revised accordingly. This brings us to the next part of design research cycle; teaching experiment and retrospective analysis, which is the main discussion of this chapter.

The retrospective analysis is defined by Graveimeijer and Cobb (as cited in Akker, et al., 2010) as examining the data set collected during the teaching experiment which aims at supporting the researcher to revise and improve the local instruction theories as well as answering the research question. It gives light to the result of the research. It is considered as an underlying principle that explain how and why the designed instructional material works. This chapter will be divided into two parts, one for each cycle. In each part, the comparison will be made between students' prior knowledge, meaning their knowledge before the teaching and learning process, and students' acquired knowledge, meaning the knowledge that they acquire during the learning process. Throughout this chapter, triangulation between students' answer during pre- and post-test, video registration of the classroom experiment, and the students' written works during the learning process will be made.

5.1 First cycle of teaching experiment

As has been explained before, the first cycle is conducted with 6 out of 27 students of class VII B. These students are Diana, Indira, Gaby, Anggi, Indra, and Yosep (pseudo-names). The analysis of the first cycle will be started with the result of the pretest and the students' interview, followed by a description of the teaching experiment and the analysis, and finally the refinement made to the Hypothetical Learning Trajectory.

5.1.1 Pre-assessment

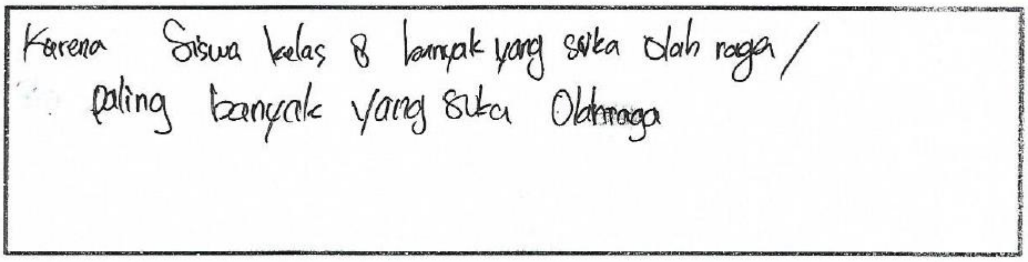
The pre-test consists of one problems with 4 sub-questions. The data used were virtual and presented in the form of bar chart, depicting interest in sports among 96 middle school students, which is hypothetically randomly chosen from 500 students. The chart shows that the students who like sports are mostly grade 8, followed by grade 7 and 9.

The explanation about the result of the pretest will be organized in two parts; descriptive statistics and inferential statistics. The first and second question are focused on descriptive statistics. There are some questions about the graph that the students have to answer, and then they are asked to mention 1 thing that they find most noticeable about the graph. All students can answer the first sub-question correctly, showing that they understand the information presented in the bar chart.

The answer for the second question, however, are quite varied. Only 2 students, Diana and Yosep, are able to spot one thing that they find most interesting about the graph, which is the grade that likes sports the most and the least, respectively. The other students chose to describe the whole chart, and one went totally irrelevant by describing the purpose of the chart instead.

The third and fourth question start to move into the range of inferential statistics. The students are asked to determine from which grade a student is if they say they like sports. Students are then asked to provide reasoning for their answer, referring to the chart.

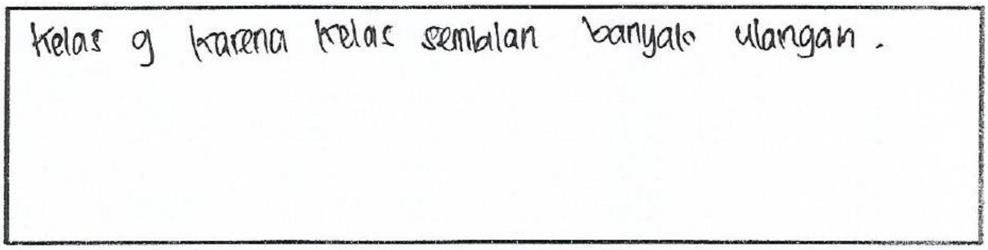
Three students (Indira, Indra, Yosep) chose grade 8, referring to the bar chart without considering the fact that it is only 96 of 500 students. These three students still think of the data presented in the chart descriptively.


Karena Siswa kelas 8 banyak yang suka olah raga /
paling banyak yang suka Olahraga

Because there are a lot of students from grade 8 who likes sport/most number of students who like sport.

Figure 5.4. Yosep's answer for question 4

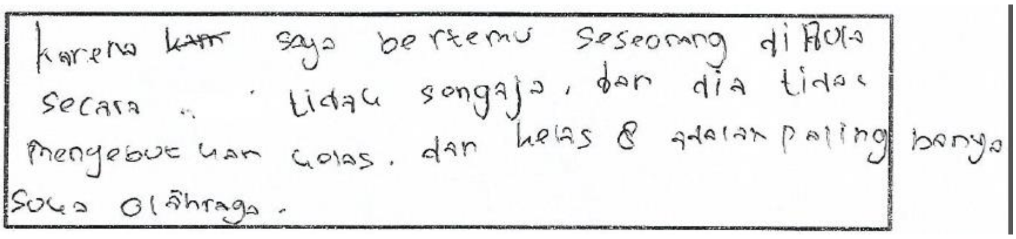
Two students fail to grasp the questions. Diana's answer is grade 9 and her argument is because the number of grade 9 is the largest, hence it was the most possible to meet grade 9 randomly in the hallway. Gaby mistakenly read that the student they met in the hallway dislikes sports, hence her answer is grade 9. According to the interview conducted after the pre-test, she confirmed a very personal argument, which is because she thinks grade 9 is too busy with exam to do sport.


kelas 9 karena kelas sembilan banyak ulangan .

It's students from grade 9, because they have many exams.

Figure 5.2 Gaby's answer for question 4.

The closest answer to inferential statistics came from Anggi, who chose grade 8 as her answer. Her argument is that they met the student in the hallway randomly and the student does not mention their grade, and since in the graph grade 8 is the one that likes sports the most, it is very possible that the students is a grade 8.


karena ~~kam~~ saya bertemu seseorang di Aula secara tidak sengaja, dan dia tidak menyebutkan kelas. dan kelas 8 adalah paling banyak suka olahraga.

Because I just happen to meet someone in the hallway, and they didn't say what grade, and [in the chart] grade

Figure 5.3 Anggi's answer for question 4

In general, the students were still thinking of the data descriptively and did not exhibit any sign of probabilistic language. . They also did not find it problematic that the chart only depicted 96 out of 500 data. They did derived conclusion about something that might not be part of the information offered by the chart.

Table 5.3 provide alignment issue between the result of the pre-test and the element of IIR.

Table 5.3 Alignment issue between the result of pre-test and the element of IIR

Framework	Comments
Generalization beyond the data.	The students produce a statement about a larger population by considering information from its sample.
Data as evidence	Most of the students, when asked to argue about their answer, refer to the chart provided. Only one students use their personal knowledge.
The use of expression of uncertainty.	No sign of probabilistic language. The statement that the students made was deterministic.

5.1.2 Teaching experiment

This section focuses on the first cycle of the teaching experiment. It consisted of 4 lesson. The explanation will be started on brief description of the tasks and the learning goals, followed by an elaboration on how the lesson went and the analysis of the result.

5.1.2.1 Lesson 1: Scout Staff

The first lesson is aimed for students to recall the statistical knowledge they already learned in previous grade, which is collecting and representing data. It consisted of two similar activities, only that the data the students work with in the first activity is of smaller size than in the second activity. The context is finding out suitable size for Scout staff in grade 7. The students have to collect data about students' height and represent them in various form, only that the data they gathered in the first activity will represent sample, while the data from the second activity represent population. In both activity, the discussion was centered on the dot plots. The students were going to analyze the dot plots to answer two main questions: 1) what they can find out about the height of the students from the chart, and 2) what the typical height of grade 7 students is.

Activity 1.1

To start the first activity, the teacher asked the students about Scout extracurricular and how many of them participating in it. The teacher then proceed to inform the students about the regulation by the Board of Indonesian Scout regarding the fixed length of the Scout staff for Indonesian scout. The discussion then moved to whether or not this size is suitable for all Scout in Indonesia. Some students argue that it is not fair because it is probably more suitable for male students, while female students are generally shorter. In these arguments, it can be seen the students' conception on what the typical height for male and female grade 7 students is.

The teacher instructed the students to collect the data by taking pieces of paper out of the class population box and list them on the table provided in the worksheet. The teacher then showed all the charts that the students had to put their data into, with specific instruction on putting in the data into dot plots form using the blue and red markers. The students then proceed to work.

The students had no significant challenge during the data collection and data representing part of this activity. Dot plot, although was a new way to represent data for them, proven to be an interesting chart to construct. Generally, the students were able to analyze the data visually and not pressured to do any calculation. When the students are asked to mention one thing that they notice from the chart, most of them focus on the difference of height between boys and girls by noticing that blue dots are mostly on the left side of the chart while pink dots are on the right side. Similarly, when answering question about typical height, most of the students focused on the modal values.

The students' first struggle came up during predicting the class chart. After some guidance, the students resorted to assuming that the class chart must also possesses the group chart's characteristics.

The compatibility of the HLT and the ALT is presented in the Dierdorp matrix analysis below.

Table 5.4 Compatibility between HLT and ALT of activity 1.1

No.	Task	HLT	ALT	Quantitative expression
3.	Look at your chart. What can you say about the height of the students whose data you collected?	Mentioning the tallest and shortest students.	The students focused on the fact that the blue dots are distributed more on the left side of the chart, while the pink side are on the right, and conclude that the male students are generally taller than the female students.	+
		Mentioning the majority by referring to the modal value.		
		The difference between male and female students' height.		
4.	If someone ask you, what is the *typical* height of the students whose data you collected, what is your answer to that?	Finding the typical visually through the modal value.	The students analyze the data visually and focuses on the part of the chart where the dots are mostly stacked (modal values)	+
		Finding the typical numerically through arithmetic mean.		
5.	Predict what the chart will look like if you collect data from the students of the whole class!	Using characteristics of the group charts to predict the class chart.	The students assumed that must also possesses the characteristics of those of the group chart.	

The only obstacle is the time limitation. Four types of chart were proven to be too much to work on in one lesson-hour. Even though the students already possessed prior knowledge about frequency table, calculating the percentage of the frequency of each data point was stressful for them and hence, error-prone. To tackle this, the teacher decided to leave the table half way and instructed the students to move on to the other graph. This difficulty turned out to be an interesting point of discussion about which type of data representation is the most attractive and efficient.

Activity 1.2

To start the second activity, the teacher explained that the students are going to construct a dot plots from one population box together. The teacher showed the stickers and the A3 paper on which the students must construct the start. All six students were then sat together around a table and input the data together.

Since they were working together, the discussion which is initially intended to be within each group turned out to be between and among groups. The students discussed the questions in the worksheet together and then wrote down the agreed upon answer on the worksheet. Since the questions are similar to the ones they got for Activity 1.1, the discussion went one a little bit more smoothly since the students already established the ground for the answer. In general, their strategy were more or less similar to those they use in the previous activity. However, due to the time restriction, the discussion about which group's prediction match the actual class graph cannot take place.

The compatibility of the HLT and the ALT is presented in the Dierdorp matrix analysis below.

Table 5.5 Compatibility between HLT and ALT of activity 1.2

No.	Task	HLT	ALT	Quantitative expression
2.	Look at your chart. What can you say about the height of the students whose data you collected?	Mentioning the tallest and shortest students.	The students focused on the fact that the blue dots are distributed more on the left side of the chart, while the pink side are on the right, and conclude that the male students are generally taller than the female students.	+
		Mentioning the majority by referring to the modal value.		
		The difference between male and female students' height.		
3.	If someone ask you, what is the *typical* height of the students whose data you collected, what is your answer to that?	Finding the typical visually through the modal value.	The students analyze the data visually and focuses on the part of the chart where the dots are mostly stacked (modal values)	+
		Finding the typical numerically through arithmetic mean.		
4.	So if you want to design a Scout staff especially for the students in this class, what size do you think is suitable?	Using the typical height to determine the suitable height of the scout staff.	All groups consider the typical height of the students and design the Scout staff with height in accordance to that.	+

5.1.2.2 Lesson 2: Compare the Dots

The second lesson is focused on developing 2 big ideas; 1) that a part of data can represent the whole (representativeness), and 2) the bigger this part of data, the more likely it represents the population (effect of sample size), both are the topic of activity 2.1 and activity 2.2, respectively. The context used here is the disappearance of the class chart, so the students have to determine which group chart can represent the class chart the best in term of the information it depicts. In activity 2.1, the students choose one from three group charts to represent the class chart the best. In activity 2.2, the students adding more and more data value to the least representative to chart, also known as *growing sample* activity.

Activity 2.1

The teacher started the lesson by asking the students about their predictions of the class graph and which one of their predictions actually matched it. The teacher collected all of the group charts and stick them to the whiteboard. The researched then proceed to distribute the worksheet, the charts, and the table to the students. The students discussed within their groups for a few minutes about the difference and the similarities of each group chart to the class chart.

Afterward, the teacher asked the students to vote which group chart is the most and the least representative one.

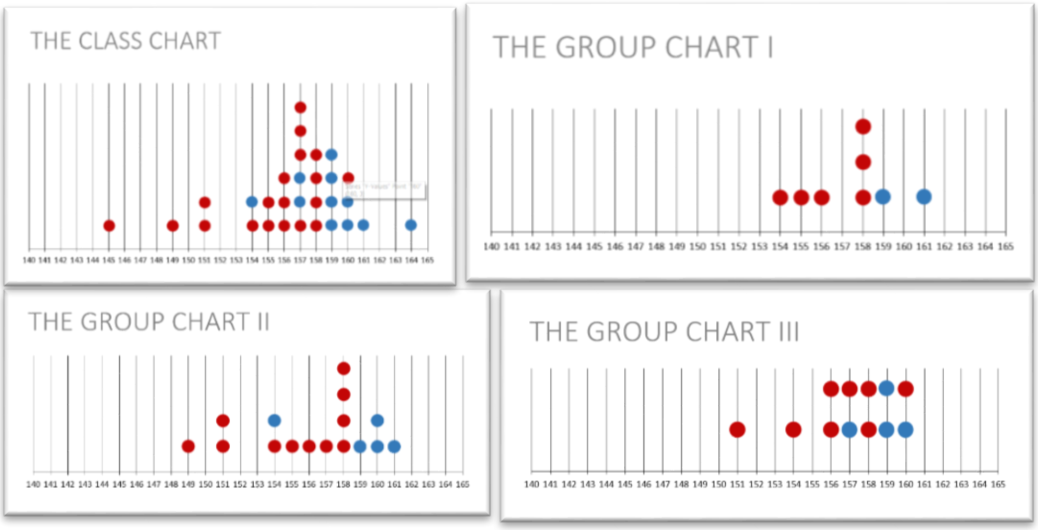


Figure 5.4 The class chart and the group charts of cycle 1

As predicted, the students analyze the dot plot visually to find the characteristics of the set of data. The students use the same strategy when asked about the typical height (in Bahasa Indonesia: *tinggi kebanyakan siswa*) and average height (in Bahasa Indonesia: *tinggi rata-rata*), which is to find the modal clump (modal values) in the dot plot. Gaby, for example, was unable to find the typical value in her chart (Group 3) because for her, the chart is too “flat”.

As found in the previous meeting, unless asked specifically to use the formula of arithmetic mean, the students think of *average* as the typical value or mode. The preferred choice of using modal value as measure of central tendency is definitely encouraged by the use of dots as data representation, meaning when using different representation, the students will probably use different measure. We can interpret that the dot plot did helped the students to construct the notion of central tendency, so the notion of average no longer means 'add-and-divide.' None of the students, however, detected the difference between charts from the shape of the charts.

A case with statistical vocabulary arose. Yosep claimed that graph 1 is the least representative because the 'size' is different. Apparently, he thought of the size here as the width of the dot plots, not the number of data points. Unfortunately, the teacher did not follow up this case.

All groups picked group charts 2 as the most representative and group chart 3 as the least representative. Group 2 had to go to the front of the class and explained it. It was ended by a conclusion that group chart 2 can represent the class chart because it gives similar information.

The following Dierdorp matrix analysis depicts the compatibility issue between the HLT and ALT for activity 2.1.

Table 5.5 Compatibility between HLT and ALT of activity 2.1

No.	Task	HLT	ALT	Quantitative expression
1.	Which one of the group chart is the most suitable to represent the class' chart?	Analyze the charts visually then picking the chart that has similar modal clump, maximum or minimum value, or distribution of blue and pink dots.	All students detected the characteristics of the group charts as well as compared and contrasted the group charts based on those characteristics.	+
2.	Which one of the group chart is the least suitable to represent the class chart? What	Analyze the charts visually then picking the chart that has different modal clump, maximum or minimum value, or	One group compares the shape, while another group compares and constrasts the characteristics of the chart.	+

	can you do to make it more suitable?	distribution of blue and pink dots.		

Activity 2.2

The teacher started activity 2.2 by bringing group chart 3 into attention. In group chart 3, when the teacher emphasized the difficulty in finding the typicality because they have to use a very wide range of modal values, the students enthusiastically agreed. This inability to find typicality was communally agreed to be the reason group chart 3 is the least representative. The teacher then asked the students about what they could do to make group chart 3 more representative. One of the student, Gaby, made a good discussion point by answering, “Adding (more data) to it.”

The teacher then proceed to ask each group to focus on their own copy of group chart 3. The teacher took 1 population bag, took out additional piece of paper by herself in number as instructed in the worksheet, and read it aloud for the students to put the data in the group chart 3.

The conjecture of this activity is quite straightforward. As the number of data gets closer and closer to the total number of students the class, students can see that the group chart starts to look more like the class chart. Transcript 1 shows how the discussion goes when the students were adding the data value, while figure 6 shows the transformation of group chart 3.

Fragment 1

Teacher: Are you guys done? Okay. The second one is Ita, she is a girl, 157 cm.
Gaby: It’s rising!
Indira: It’s got promoted.
Diana: Ummm, 158 is the most *unknown* now …
Yosep: It’s just the same!
Teacher: the third one is Rudi, he’s a boy, 158 cm.
Teacher: Okay, the data added up 5 more points, what can you notice?
Diana: 157 is the tallest one, but …
Teacher: Is the shape better than before?
Anggi: Yes
The others: It’s better!
Indira: Much better, miss, it rose.
Teacher: What aspect is being corrected? Can you find the typical value now?
Anggi: 157 ..
Teacher: 157-158, right? It’s getting closer (to the class chart), right?
Teacher: Okay, so the number is adding up and getting closer to the number of students in the class. What can you notice?
Indira: The heights, is starting to be more visible, it’s starting to look like *the class chart*

Based on the transcript, it is quite evident that the students were able to see that the chart of group 3, which started out to be quite ‘flat’, was starting to look more and more like the class chart. The students used word like ‘rise’ and ‘promoted’ to address the construction of the modal clump. Because now they could determine the typical value, as opposed as the initial shape of group chart 3, they also stated that the shape was much better than before. Therefore, it can be interpreted that the students are starting to conceive the notion that the bigger the part of data, the more likely it resembles the whole.

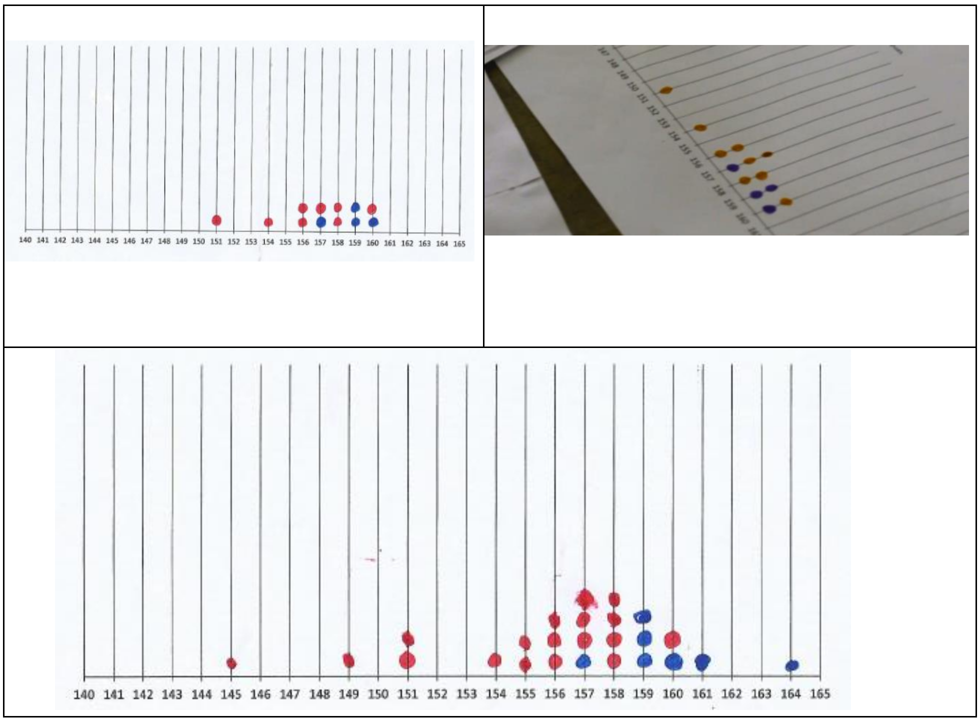


Figure 5.5. The "growth" of group chart 3.

The following Dierdorp matrix analysis shows compatibility issue between the HLT and ALT for activity 2.2.

Table 5.6 Compatibility between HLT and ALT of activity 2.2

No.	Task	HLT	ALT	Quantitative expression
3.	As the number of data that you take is getting closer to the total number of students in the class, what is your conclusion?	Stating that the more data value added to the group chart, the more it resembles the population.	All groups comment that the graph is getting 'higher' and that the typical value is getting close to that of the class chart. They also claimed that the bigger the number of data inserted into the group chart, the more it resembles the class chart.	+

5.1.2.3 Lesson 3: Are Boys taller than Girls?

The third lesson is designed to develop two big ideas, namely 1) with a really big set of data, it is impractical to take the whole data into account; and 2) all member of population have to have equal chance to be picked as sample. The context is finding information about the height of the students in the whole school. This lesson consists of two activities, activity 3.1 and activity 3.2, designed to develop big ideas 1) and 2) respectively. In activity 3.1, the students are tasked to create a dot plots depicting the height of the students of the whole school. In activity 3.2, the students have to judge whether or not the report of the students' height made by the Head of the Students Council is acceptable.

Activity 3.1

Prior starting the activity, the teacher reminded the students' of the previous meeting. A short discussion commenced about which group chart wins and why. The teacher then distributed the worksheet ad population box, all the while announcing that the students were going to construct a dot plot of the whole school. This chart was later referred to as the *school chart.* Afterward, one of the group presented their work to the front of the class.

During the construction of the school chart, the students had no indication that they realize this task is much more demanding than the previous one. They seemed to take pieces of paper randomly out of the population box and plot them on the provided dot plot scale, just like they did in the previous lesson. Only after the teacher reminded them of the limited time, that the students realized that trying to put all the data into the chart is futile.

For the second problem, the students are tasked to determine the difficulty they encounter in constructing the school chart. All students mentioned time and the number of data as the obstacles in constructing graph, just as conjectured. The students claimed that it is impractical to take the whole data into account, but none of them make the connection to the previous meeting, where they can gain information about the whole data by analyzing only a part of it. Only after the teacher made them recall the previous meeting that the students came into a conclusion that taking the whole data is impractical and they can consider only a part of it instead.

In the third problem, the students are tasked to describe what information that they can gain about the height of the students in the school, based on the school chart. Just as conjectured, the students analyzed the school chart visually.

The chart they discussed is shown in Figure 3.

Fragment 2
Yosep: In this chart, the boys are taller, the height is 160 something.
Teacher: How about the typical height?
Yosep: The typical height .. 160.
Indra: 161!
Yosep: No, its 160. 158 to 160.
Indra: Oh, right.
Teacher: Really? How come you arrive at that conclusion?
Yosep: (points at the chart, where the dots cluster the most)
Teacher: Oh, I see, that’s where all the dots are, right? You guys see that, the dots? They cluster over there, don’t they?
Diana: It is around 150 Miss.
Teacher: Well where is the hill then? From 158 ...
Indra: (point to the clump) ... until 160.

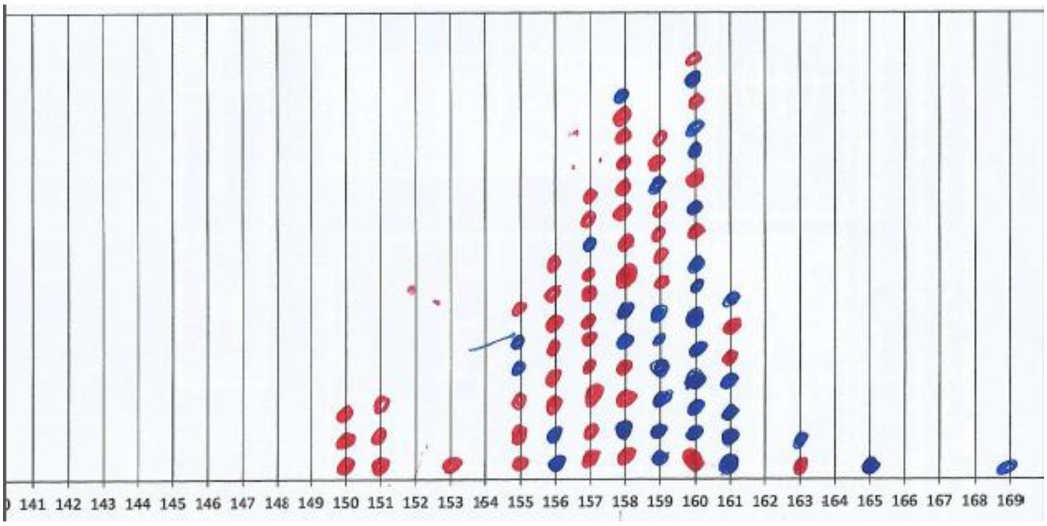


Figure 5.6. The school chart of group 1

This discussion even moved further into expression of uncertainty. The school charts by all three groups apparently have the same typical value, which was ranging between 158 and 160. It was also interesting that they used range (modal clump) instead of only using one modal value. Although when considering only individual chart the students were in doubt whether or not the whole data says the same information, they were quite sure after three charts say the same thing, as shown in the following fragment (Fragment 3).

Fragment 3
Teacher: What do you think? Do you guys agree? Group 2?
Diana: Our grup is also 158 to 160 Miss.
Teacher: Oh, it's the same? You guys? (points to group 3)
Gaby: Its 158 to 159.
Teacher: Okay, if three small data have said the same thing, are you guys sure that is the information given by the big data?
Indra and Yosep: We are sure, Miss.
Teacher: Is it certainly the same?
Diana: Well, not exactly Miss.
Teacher: What is the possibility then?
Diana: Not certain, but very possible.
Indra: (The possibility will be) .. big.
Teacher: I know, right? All the charts say the same thing.

In general, the activity of constructing charts of the whole school is able to help the students to be aware of the need for sampling. Even though they have to be guided to connect this activity to the previous big idea, which is a part of data can represent the whole, the students were able to grasp the idea of why sampling is needed by exposing them to the problem that arise when working with a really big set of data.

The following Dierdorp matrix analysis depicts the compatibility issue between the HLT and ALT of activity 3.1.

Table 5.7 Compatibility between HLT and ALT of activity 3.1

No.	Task	HLT	ALT	Quantitative expression
1.	Making dot plot of the height of the students in SMP Santa Maria.	Trying to put all the data into the dot plot before realizing it is too impractical	All groups tried to put all the data values into the chart and stopped half way because the time is not enough.	+
		Not trying to input all the data value and instead already taking a subset of it during the first try, but they try to get as much data value as possible into the dot plot.		
		Being picky about the data value that they put in the dot plot by selecting the paper and putting back the ones that they consider unfit for their idea of what the school chart might look like.		
		Showing the sign for randomization,		

		either by shaking the bag or stirring the pieces of paper instead, and then taking a subset of it to be put in the dot plot.		
2.	Describe the way you create the dot plot. Is there any difficulty that you did not find when constructing the previous dot plots?	Focusing on the limited time or the number of data values as the obstacles in inserting the data into the chart.	All groups claim that even though constructing dot plots is easy, the number of data to be put into the chart is too much while the time is limited.	+
3.	What can you say about the height of students in grade 7? Which one is taller, girls or boys?	Focusing on the modal clump, the shortest or the tallest heights, or the difference between the height of girls and boys.	Group 3 mentioned that girls are shorter than boys. Group 1 mentioned the difference in number between boys and girls, and the typical height. Group 2 mentioned the typical height. All groups mentioned the tallest and shortest students.	+

Activity 3.2

The teacher started the activity by introducing the context, which is a report about students' height in a fictional school made by its Head of the Students' Council. The Head of the Students' Council only takes data from some students in grade 1, grade 2, and the girl basketball and volleyball club, as shaded in the given map, because it is closer to his office. The number of girls is much larger than the boys, and the girls are athletes who are generally taller than their peers, hence the dot plot included in the report by the Head of the Students' Council shows that the girls are generally taller than the boys. The problem is designed so that it contradicts the students' presumption in order for them to look deeper into the data. After quite a long discussion, the students wrote down the result of the discussion on their worksheet, followed by a presentation by Group 3.

The students analyzed the graph visually and conclude that the girls are taller than the boys. The context did contradict the students' presumption because they compared the report to their school chart previously, emphasizing the difference in result.

The students, however, did not immediately see anything wrong with the way the Head of the Students' Council choose their data. They claimed because it is what they see on the plots, and because the report state what is shown on the chart, the report must be legitimate. The students had to struggle to conceive the idea that every member of the population has to have equal chance to be picked as sample, or else the sample will not be representative. The following fragment shows the struggle the students experienced during the whole class discussion.

Fragment 4
Teacher: Is the conclusion correct?
All: Yes ...
Diana: Because I see it (the plots) with my own eyes.
Teacher: But he took-he took the data only from this part of the school (points to the map). Please consider the way he collects the data. He only collect it from grade 2, grade 1, girl basketball club and girl volleyball club.
Teacher: What are the students in the volley and basketball club like?
Yosep: They are ... all girls?
Teacher: Okay ... is that all?
Yosep: Tall?
Teacher: Yes, they are tall. So is the Head's conclusion correct?
Indra: (inaudible) for now ..
Teacher: He stated for the whole school that the girls are generally taller than the boys, however the data is not from the whole school but ...
Indira: but its only from part of it ...

Teacher: If it is like that do you think the conclusion correct?
Indra: It can't be like that ...
Teacher: Why not?
Yosep: Because it is not-it's like, he did not consider the whole school. So it's not all.
Teacher: Okay, it is not all. But we have concluded that it is impossible to consider all the data, right?
Yosep: (inaudible) it has to be **evened** ...
Teacher: Evened! What do you mean by 'evened'?
Yosep: (inaudible mumble)
Teacher: So, what is the correct way to collect data? Can we do it like this?
Indira: It's not all students ..
Teacher: Its not all ... alright, but we agreed that it is not practical to take the whole data right?
Yosep: It has to be even Miss.
Teacher: Even? Can you please explain that?
Yosep: (inaudible mumble)
Indra: Speak clearly!
Teacher: Okay, it is ***fair*** then?
Yosep: Nooooo ...
Teacher: Why not?
Indira: its better if he takes it **randomly** ...
Teacher: He wasn't fair because he only collects data within the shaded area whle the conclusion include the whole school right?
All: Yes ..
Teacher: So what if you guys are the Head of the Students' Council? What would you do?
Yosep: If I want to be fair, Miss ...
Techer: Right, how?
Yosep: I'll take, like, the students from girl and boys basketball club, so there would be like, both boys and girls ..

The bolded words are informal word that the students use in place of formal statistical terminologies. None of the students took the fact that most of the girls are from sports club as a problem. Instead, they focused more on the imbalance between genders.

From the discussion, it was evident that the students were able to see why the data collection is biased and why it is important to avoid doing so. However, although they had started to conceive idea of randomization, they were still having difficulty to come up with a randomized way of collecting data. The following

Dierdorp matrix analysis depicts the compatibility issue between HLT and ALT of activity 3.2.

Table 5.86 Compatibility between HLT and ALT of activity 3.2

No.	Task	HLT	ALT	Quantitative expression
1.	Based on the chart, what can you say about the heights of students in the school?	Focusing on the modal clump, the shortest or the tallest heights, or the difference between the height of girls and boys.	In general all groups noticed that female students are taller than male students. One group pointed out the difference between the class chart and the school chart, because in the class chart, the male students are taller.	+
2.	The Head of the students' council looked at the chart, made his calculation, and said to himself, "***Hmm, apparently girls are taller than boys.***" Do you think his result about the height difference between male and female students is reasonable? State your reason!	Claiming that it is reasonable because the dot plot says so.	Two groups claimed that the decision of the Head of the Student Council is unacceptable because the number of data collected from female students is far larger than those from male students.	+
		Claiming that it is unreasonable because from previous results, boys are taller than girls.		
		Claiming that it is unreasonable because the size of the data collected is not big enough.		
		Claiming that it is unreasonable because the head of students' council is *picky* or *unfair* in collecting the data.		

3.	What should Adi do to get better data?	Stating that the head of the student's council should be fair in the data collection, for example by going to class to class and taking equal number of data from every class.	The students initially cannot see the problem with the way the Head of the Student Council collect his data. After some guidance, they agreed that the head of the student council should collect equal number of data from boys and girls. One group insisted that the size of the data collected should be increased.	+

5.1.2.4 Lesson 4: Social Media Addiction

The fourth and last lesson is a mini research project, designed to be a ground where the students are expected to employ all the informal statistical ideas they have learned in the previous lesson. The context is finding out information about social media addiction among grade 7 students in Laboratory Secondary School of State University of Surabaya.

In the end of the Lesson 3, the teacher already introduced and discussed the context with the students. Each group was given a report form with a table and a dot plot scale included. There was a one-week gap between Lesson 3 and Lesson 4 where the students were expected to work on this report. The reports were then presented at the Lesson 4.

From the class discussion, it seemed that all groups already possessed a sense of sample and population because they only collect data from some students, yet they acknowledged that the information revealed by their conclusion is about all students in their school. They also referred to their plots for every claim and statements that they make. In general, we can interpret that they were able to make generalization from sample to population and use data to back up their generalization.

The following fragments show the discussion that happened when Group 1, Indra and Yosep, presented their work in front of their class.

Fragment 5

Yosep: … the next is (data collection) method. Explain the following things. What kind of data do you need? How long the students usually spend their time in social media? Who do you take your data from? Lab School students.
Teacher: From lab school students! What grade?
Yosep: 7, 8, 9.
Teacher: And then?
Yosep: Where do you collect the data? At the school. When? Last Monday. How do you collect it? By looking for people and ask them. What do you do with the data? Finding out the average.
Teacher: Do you collect your data from all students in grade 7, 8 and 9?
Yosep: Yes.
Teacher: All of them.
Yosep: No, just one by one …

However, even though that the students already possessed a sense of which one the sample and which one the population, the way they collected their sample still suggested that they were still having difficulty in applying the idea of a representative sample in the process of data collection. In term of size, they only collected data as much as the given table allowed them to. They did not consider how many sample size will be representative for a population of 200 students.

Meanwhile, in term of randomization, the students did attempted at giving equal chance to every member of the population, based on their grade and gender, even though in the end they still collected data from their own class the most. This attempt at randomization became an interesting point of discussion their data being not ***even*** and ***fair*** (in Bahasa, ***adil*** dan ***rata***) enough, as shown in fragment 6. In the next fragment, fragment 7, it can be seen an expression of uncertainty that the students make in deriving conclusion.

Fragment 6

Teacher: Exactly, we already learned that it will be impractical to consider all data right? Part of data can represent the whole, right? We also have learned that it should be ***fair*** and ***even*** right? How do you collect your data so it will be fair and even?
Yosep: Ummm, we have 3 students from 7A, 3 students from 7B, 9 students from 7C, 5 from 8A, 5 from 8B, we have none from 9A, and 1 from 9B.
Teacher: Okay, so they (Indra and Yosep) collected data from every classes, or at least there are students from every class. Do you agree with the way they collect their data?
All: Yes.
Teacher: Is it fair? Is it even?
Anggi: No, there are way more 7th grade students than other grades.
Teacher: So, from the perspective of data collection, it is ..
Yosep: Yes, it is not that fair.

Fragment 7

Teacher: Now describe your graph.
Yosep: In this graph, the students typically (in Bahasa, ***rata-rata***) use social media sites for roughly 2 or 3 hours each day.
Teacher: Are you sure with your conclusion?
Yosep: Yes.
Teacher: So you think all students play for 2 or 3 hours?
Yosep: Ummm (hesitates) well hopefully yes.
Teacher: So? Are you sure? Or not?
Yosep: No …
Teacher: Why?
Yosep: Because there is still other answers.

What Yosep referred to as other answers is the data that they did not collect. It can be interpreted that Yosep expressed uncertainty about his conclusion, and it comes from sample variability. Meaning that he is aware that there are a lot of possibility on the sample he collects from the population. On the other hand, none of the students use the sampling process as the source of certainty (or uncertainty) of their conclusion. In general, even though the students could assess their own sampling process in term of sample size and randomization, they were unable to link the representativeness of their sample to the certainty of their conclusion.

During the presentation, the students have to be guided a lot. Instead of presenting their reports in a story-telling way, the students treated it like question and answer. The presentation, class discussion as well as their reports also suggested that the students are still used to the question-and-answer type of mathematics problem. They were also uncomfortable in expressing opinions either spoken or written.

Table 5.9 depicts a Dierdorp matrix analysis elaborating compatibility issue of the HLT and ALT (in line with elements of Informal Inferential Reasoning) of Lesson 4.

Table 5.8 Compatibility issue between HLT and ALT of Lesson 4

Task	Elements of IIR	HLT	ALT
A newspaper article have revealed worrying trend among the teenagers; social media addiction. Aside from robbing them from the opportunity to actually experience genuine human interaction, this also increases cyber-crime rate. The study this article refers to claim that ***Indonesian teenagers spend up to 12 hours a day on the social media sites.*** Now your assignment is to become a researcher and invesigate whether or not this claim is correct for the students of Lab Secondary School. Use the resul of your research to answer the following question: ***How many hours in a day the students of Lab Secondary Schhol usually spend their time on social media sites?***	Making generalization from sample to population	Considering all data in deriving conclusion about a population.	The students collect data from some parts of the school and making conclusion that apply for all students in the school.
		Considering only part of data in deriving conclusion about a population.	
	Use data to back up this generalization.	Personal experience overrule data and context.	When deriving conclusion, the students refer to their dot plots without considering the context.
		Over-reliance on context over data or vice-versa.	
		Consider both data and context together when drawing informal inferences.	
	The use of expression of uncertainty.	No expression of uncertainty; the students derive deterministic conclusion.	The students produced expression of uncertainty which is based on sampling variability. Although they can detect and describe the flaws and the strength in their sampling process, they did not use it to justify their conclusion.
		Presence of expression of uncertainty, attempt at justifications which can be based on context, data, or personal experience.	
		Presence of expression of uncertainty, with appropriate justification based on data (method of sampling, the method of data collection, etc.)	

5.1.3 Post-assessment

The post-test was conducted in the end of cycle 1. The test consisted of similar problem as those in the pre-test, although the context was altered a bit. In this test, the data used was virtual and presented in the form of bar chart, depicting pet rabbit ownership among citizen of a certain city in Indonesia. The sample size is 96 and taken from three ethnicities; Balinese, Sundanese, and Maduranese. The chart shows that the highest pet rabbit ownership goes to the Sundanese, followed by Balinese and Maduranese.

The explanation about the result of the pretest will be organized in two parts; descriptive statistics and inferential statistics. The first and second question are focused on descriptive statistics. There are some questions about the graph that the students have to answer, and then they are asked to mention 1 thing that they find most noticeable about the graph. All students can answer the first sub-question correctly, showing that they understand the information presented in the bar chart.

For the second question, there is a subtle improvement from the pre-test. Most of the students could detect aspects that they found most interesting about the charts. They varied from the unfairness in the way the sample was taken, to the fact that there are a lot of Maduranese who do not own a pet rabbit. Two students went totally irrelevant and describe the purpose of the chart, instead.

The third and fourth question start to move into the range of inferential statistics. The students are asked to determine whether or not someone owns a pet rabbit if they say they are from Madura. Students are then asked to provide reasoning for their answer, referring to the chart.

There is a stark contrast from the answer given by the students in the pre-test. The students seemed to be more cautious and uncertain in drawing inference. Most of them stated that they cannot be certain whether or not they can decide about the pet ownership of this person, although only one dedicated this uncertainty to the sampling variance. Only two students derived deterministic conclusion based solely on what they can see from the chart.

In general, the students were still thinking of the data descriptively. Table 5.3 provide alignment issue between the result of the pre-test and the element of IIR.

Table 5.10 Compatibility between the result of post-test and the element of IIR

Framework	Comments
Generalization beyond the data.	The students made claim about a population based on what they can notice from sample.
Data as evidence	Most of the students, when asked to argue about their answer, refer to the chart provided. Only one students use their personal knowledge.
The use of expression of uncertainty.	The students are much more cautious and uncertain in drawing their conclusion.

5.1.4 The refinement of HLT from cycle 1 to cycle 2

Analysis of the teaching and learning of Lesson 1 to 4 reveal that some refinement needs to be applied to the course. Table 5.10 depicts the revision made to the cycle 1 HLT, resulting in the HLT implemented in the second cycle of teaching experiment. The complete initial HLT is elaborated in Chapter 4.

Table 5.11 Refinement of the HLT from Cycle 1 to Cycle 2

<table>
<tr><th>HLT from cycle 1</th><th>Refinement for the cycle 2</th></tr>
<tr><td>
Lesson 1
Activity 1.1

Due to the teacher’s advice to make the students’ worksheet more in line with the curriculum, more type of graphs were added on the task 2, activity 1.1. Hence, the students are not only expected to present the data in dot plots, but also in a frequency distribution table, line graph, and bar chart.
The students are going to collect data by asking their friends directly
The students describe what they predict the class chart is going to look like.

</td><td>
Lesson 1
Activity 1.1

Time restriction makes it difficult to create multiple graph. After consulting with the actual teacher, the students only have to create dot plots.
To save time, the students collect their data from population box.
Since the students had some difficulty in describing the chart, they are provided a scale where they can draw their imaginary class chart instead.

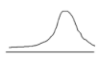
</td></tr>
<tr><td>
Lesson 2
Activity 2.1

The students given a table to assist them in comparing the all the charts. This table provide characteristics based on which the students have to compare the group charts and find out which one is more suitable to represent the class chart.

<table>
<tr><th></th><th>THE CLASS CHART</th><th>GROUP CHART 1</th><th>GROUP CHART 2</th><th>GROUP CHART 3</th></tr>
<tr><td>What is the height of the shorthest student?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>What is the height of the tallest student?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>The typical height of the studen is ... cm</td><td></td><td></td><td></td><td></td></tr>
<tr><td>What is the average height of the students?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Which one is taller, girls or boys?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Is there anything more that you notice from the charts?</td><td></td><td></td><td></td><td></td></tr>
</table>
<ol start="2">
The charts of each group are exhibited in the front of the class and the students work in group to determine the similarities and difference of each charts by examining them together.

</td><td>
Lesson 2
Activity 2.1

Due to the actual learning trajectory where none of the students are given additional category where they are expected to detect the shape of the graph.

<table>
<tr><th></th><th>CLASS CHART</th><th>GROUP CHART 1</th><th>GROUP CHART 2</th><th>GROUP CHART 3</th><th>GROUP CHART 4</th><th>GROUP CHART 5</th><th>GROUP CHART 6</th><th>GROUP CHART 7</th></tr>
<tr><td>What is the height of the shortest student?</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>What is the height of the tallest student?</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>The typical height of the student is ... cm.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Which one is taller, girl or boys?</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sketch the chart!
Example
</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>
<ol start="2">
Due to the chaos that ensued when all students are examining the charts together, each group is given a copy of each of the group charts.

</td></tr>
</table>

Activity 2.2	Activity 2
3. During the growing sample activity, the students describe their charts every time they add data points to it. Example: Take the heights from 5 more students and add them to the plot. What is happening to the chart? Describe it!	3. To evoke the students' struggle in examining the graph, they are instructed to predict what the graph is going to look like before they add data points to it. Example: What happen to the chart if you add 5 more data points to it? Predict it! Take the heights from 5 more students and add them to the plot.
Lesson 3 **Activity 3.2** 1. The context used in the problem emphasizes on the unfairness in the way the Head of the Students' Council collect his data, which happens to contain a lot of girls from sports club. This collection is reflected in the conclusion about girls being taller than boys. 2. The students describe the information contained in the chart constructed by the Head of the Students' Council, followed by judging whether or not his claim is acceptable.	**Lesson 3** **Activity 3.2** 1. A change is made to the context. It is made to be much more simple. It still emphasizes on the unfairness in the way the Head of the Students' Council collects his data, but now the bias is solely comes from the imbalance of number between genders of student. 2. The students did not have to describe the information contained by the chart anymore. Instead, the focus is now solely on judging whether or not the claim made is suitable.

5.2 Second cycle of teaching experiment

The second cycle was conducted in class VIIC of SMP Lab Surabaya. This time around, the actual teacher who in this occasion is Mrs. Ira, conducted the teaching experiment. The HLT implemented here was the refined HLT of cycle 1. Two observers recorded the teaching and learning process.

There were 27 students as participants for the second cycle. They worked in group of four, with the arrangement of the students for each group was made by the teacher to ensure the mixed mathematical ability. There were 7 groups in total. For a detailed result, the researcher and the teacher selected one group to focus on, which is group 4.

The analysis of the first cycle will be started with the result of the pretest and the students' interview, followed by a description of the teaching experiment and the analysis, and finally the refinement made to the Hypothetical Learning Trajectory.

5.2.1 Pre-assessment

As with the first cycle, the second cycle was also started with a pre-test to assess the students' prior knowledge. The test used was similar as those used in the first cycle. The pre-test consists of one problems with 4 sub-questions. The data used were virtual and presented in the form of bar chart, depicting interest in sports among 96 middle school students, which is hypothetically randomly chosen from 500 students. The chart shows that the students who like sports are mostly grade 8, followed by grade 7 and 9.

The explanation about the result of the pretest will be organized in two parts; descriptive statistics and inferential statistics. The first and second question are focused on descriptive statistics. In the first one, the students are asked about some information related to the chart. The purpose is to assess whether or not the students understand the information presented in the chart. All students answered this question correctly, meaning that they know how to read bar chart and understand the information conceived within it.

As for the second question, the students are asked about the things that they found most noticeable and interesting about the charts. As can be seen from the result of pre- and post-test in the first cycle, the culture of expressing opinion strongly influence the students' answer for this question. Some students decided to play safe by either describing the whole chart or explaining the purpose of the chart.

In this cycle, however, only one student was able to detect one aspect of the chart that they found most interesting, which is the class that likes sport the most. The other students, on the other hand, decided to describe the whole chart and mentioning one by one the information contained within it, or went totally irrelevant by explaining what they thought the purpose of the chart is.

The third and fourth question start to move into the range of inferential statistics. The students are asked to determine from which grade a student is if they say they like sports. Students are then asked to provide reasoning for their answer, referring to the chart. There was no variance in the students' answer; all of them straightforwardly answered 8^{th} grade and refer to the chart as the reason behind it. This shows descriptive nature on their part.

In general, the students were still thinking of the data descriptively and did not exhibit any sign of probabilistic language. They also did not find it problematic that the chart only depicted 96 out of 500 data. They did derived conclusion about something that might not be part of the information offered by the chart.

Table 5.12 provide compatibility issue between the result of the pre-test and the element of IIR.

Table 5.12 Alignment issue between the result of pre-test and the element of IIR

Framework	Comments
Generalization beyond the data.	The students made claim about a set of data based on what they can notice from a part of the data.
Data as evidence	The students referred to the chart when asked to argue about their answer.
The use of expression of ucnertainty	No sign of probabilistic language. The statement that the students made was deterministic.

5.2.2 Teaching Experiment

This section focuses on the second cycle of the teaching experiment. As in the first cycle, it consisted of 4 lesson. The explanation will be started on brief description of the tasks and the learning goals, followed by an elaboration on how the lesson went and the analysis of the result. The explanation will be centered around the class discussion and written work of the focus group, unless mentioned otherwise.

5.2.2.1 Lesson 1

The aim of the first lesson is for students to recall the statistical knowledge they already learned in previous grade, which is collecting and representing data. The context is finding out suitable size for Scout staff in grade 7. There are two activities in this lesson where the students have to collect data about students' height and represent them in various form. In both activity, the discussion was centered on the dot plots to answer two main questions: 1) what they can find out about the height of the students from the chart, and 2) what the typical height of grade 7 students is.

Although the two activities are similar, the size of data gathered by the students are different. In the first activity, the dot plot was constructed from the data of a certain number of student. On the other hand, the dot plot in the second activity was constructed from the data of the whole class. From this point onward, the data they gathered in the first activity will be used as sample, while the data from the second activity represent the population.

Activity 1.1

The beginning of the first cycle was started by the teacher setting the context about Scout staff. All students are involved in Scout (known as Pramuka in Indonesia) as their extracurricular activity. The teacher asked one student to read the problem out loud, followed by a discussion about whether or not the length of 160 cm is suitable for all Scout in Indonesia. In this part of the lesson, the teacher made a good point about none of the different answers being wrong as long as the students can argue about it. They went on to discuss whether or not the length will be suitable for 7th grade students. Unfortunately, the discussion was very brief and teacher did not emphasize the matter of typical height of grade 7.

The teacher instructed the students to work on Activity 1.1, followed by a specific instruction on putting in the data into dot plots form using the blue and red markers. The students then proceed to work. Each group was given a different instruction on what number of data they have to collect. The focus group, in this case, was instructed to collect data from 10 students.

In general, the students had no significant challenge during the data collection and data representing part of the activity. Dot plot, although was a new way to represent data for them, proven to be an interesting chart to construct. As it turned out, the chart constructed by the focus group was quite flat.

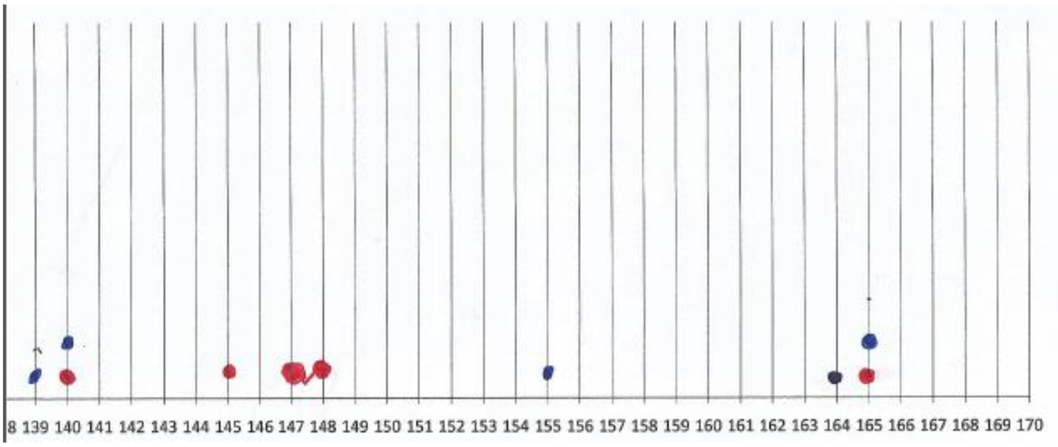
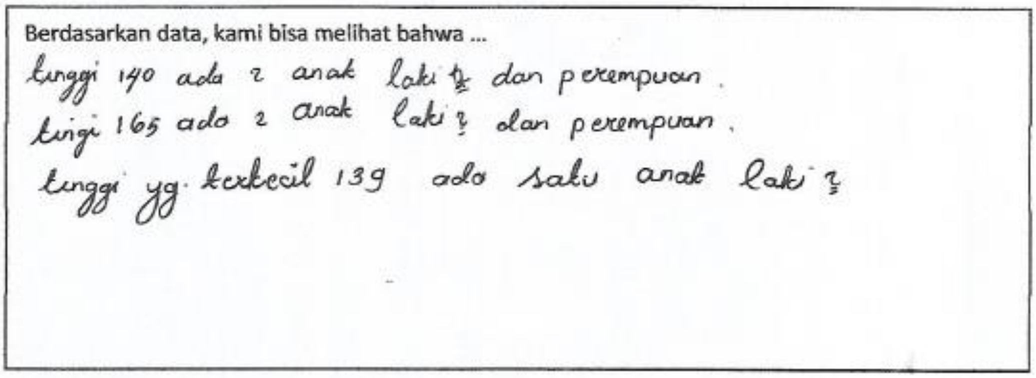


Figure 5.7. The group chart of the focus group.

There were 3 main problems which were the subject of the students' discussion during this activity. In the first one, the students were asked to describe their chart. The students mentioned the data values with the highest frequency, as well as the shortest student.



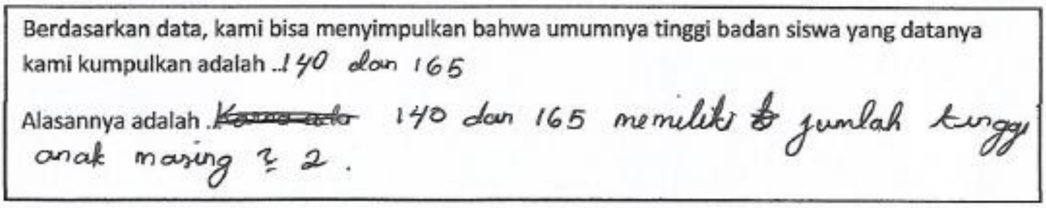
Berdasarkan data, kami bisa melihat bahwa ...

tinggi 140 ada 2 anak laki² dan perempuan.
tingi 165 ada 2 anak laki² dan perempuan.
tinggi yg terkecil 139 ada satu anak laki²

Figure 5.8 The answer of the focus group for the first problem

The flatness of the chart seems to give the students difficulty in determining the characteristics of the chart. Both modal values were not of that extreme frequency and the students settled with mentioning both of them. The shortest students also took their attention. Based on the first cycle, it is predicted that this contradict the students' initial presumption about male students being taller than their female peers. In general, it was evident that the students went directly to explore the data visually to determine its characteristics.

The second problem asked the students to determine the typical height of the students in grade 7. The word used here is *rata-rata,* which in English is closer to average than typical. As an answer to this problem, the students mentioned both the modal values.


Berdasarkan data, kami bisa menyimpulkan bahwa umumnya tinggi badan siswa yang datanya kami kumpulkan adalah ..140 dan 165

Alasannya adalah . 140 dan 165 memiliki jumlah tinggi anak masing² 2.

Figure 5.9 The answer of the focus group for the second problem

It can be interpreted that without specific instruction to calculate the arithmetic mean, the students are more familiar with average as typical value than arithmetic mean.

In the third problem, the students were asked to predict what the chart will look suppose they collect data from the whole class. The chart produced by the students were even more flat than before. They kept the modal values in the original frequency and added dots around it.

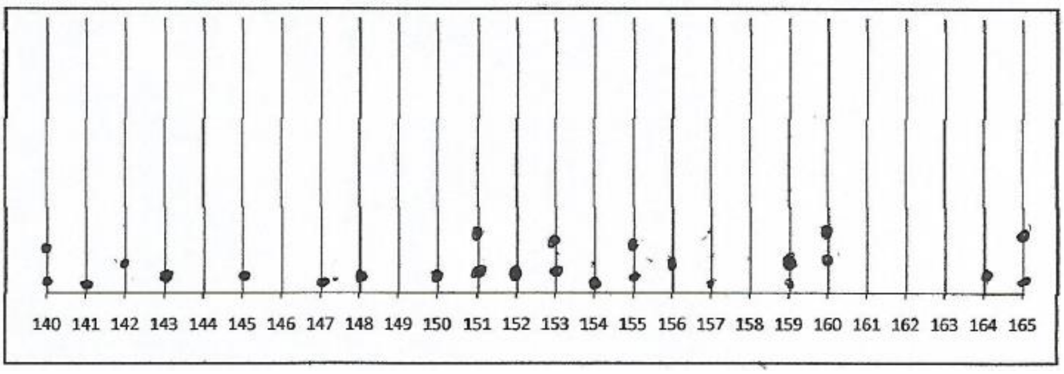


Figure 5.10 The answer of the focus group for the third problem

We interpret that the students did not consider possibility that the characteristics of the smaller data can also be present in the bigger one. Therefore, they did not use the characteristic of the group chart to predict the characteristic of the class chart.

The following Dierdorp analysis matrix shows the compatibility issue between HLT and ALT of activity 1.1 in Cycle 2.

Table 5.13 Compatibility between HLT and ALT of activity 1.1

No.	Task	HLT	ALT	Quantitative Expression
3.	Look at your chart. What can you say about the height of the students whose data you collected?	Mentioning the tallest and shortest students.	The students' attention went to the most stark characteristic of the fairly flat chart, which are the modal values. They also mentioned the shortest student.	+
		Mentioning the majority by referring to the modal value.		

		The difference between male and female students' height.		
4.	If someone ask you, what is the *typical* height of the students whose data you collected, what is your answer to that?	Finding the typical visually through the modal value.	The students determine the typical height of the students as the modal value.	+
		Finding the typical numerically through arithmetic mean.		
5.	Predict what the chart will look like if you collect data from the students of the whole class!	Using characteristics of the group charts to predict the class chart.	The students kept characteristic of the group chart and added dots randomly around it.	+

Unfortunately, the teacher did not instruct the students to construct the predicted chart with red and blue markers, hence it was not evident what they think about the distribution of male and female students. The limited time was also an obstacle because the students did not have a chance to work on the bar and line graphs.

Activity 1.2

In the second activity, the students worked together to construct a class chart. Different from the first cycle, this time each student was given a sticker and they stuck it to the scale provided in the front of the class.

Figure 5.7 The students construct the class chart together

The students were tasked to copy the class chart to a separate scale provided in their worksheet. The aim is to familiarize them with the chart and its feature. However, due to the time limitation, the students did not have any chance to do this activity. Instead, all the groups used the class chart in front of the class together to do their worksheet.

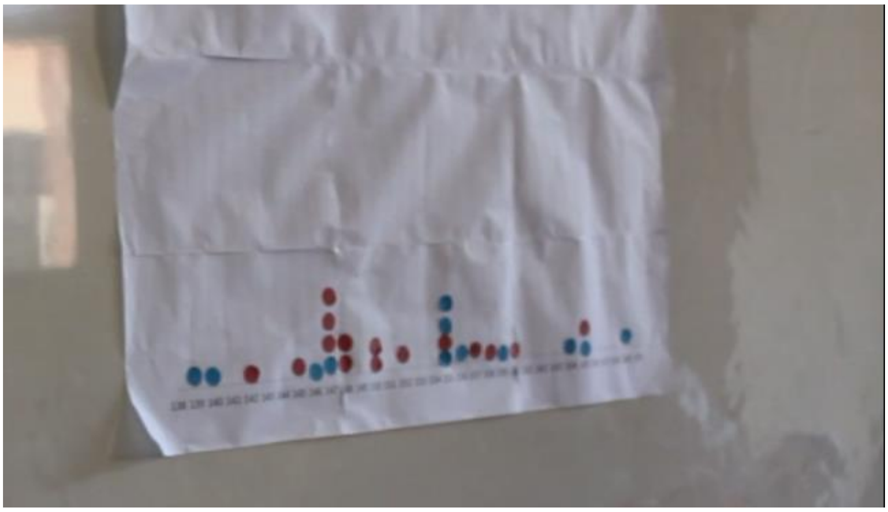


Figure 5.12 The finished class chart

There were also 3 problems the students have to work on during this lesson. The first and the second problem is similar to those in the activity 1.1, which were mentioning the characteristics of the chart and finding the typical height. With the class chart, the students employ the same strategy they used with the group chart, which is exploring the chart visually. In the third problem, the students were asked to determine a Scout staff of what length they would design for the students in their class. The students settled with 169 cm because it is the height of the tallest student in the class. They considered it the best size because it can accommodate anybody.

During the class discussion, all groups proposed different length for their version of ideal Scout staff. The teacher emphasized that there is no right or wrong in this case, as long as everyone can give a good argument on their solution.

The following Dierdorp analysis matrix shows the compatibility issue of HLT and ALT of activity 1.2 in Cycle 2.

Table 5.13 Compatibility between HLT and ALT of activity 1.2

No.	Task	HLT	ALT	Quantitative expression
2.	Look at your chart. What can you say about the height of the students whose data you collected?	Mentioning the tallest and shortest students.	The students' attention went to the the modal values.	+
		Mentioning the majority by referring to the modal value.		
		The difference between male and female students' height.		
3.	If someone ask you, what is the	Finding the typical visually	The students determine the typical	+

	typical height of the students whose data you collected, what is your answer to that?	through the modal value. Finding the typical numerically through arithmetic mean.	height of the students as the modal value.	
4.	So if you want to design a Scout staff especially for the students in this class, what size do you think is suitable?	Using the typical height to determine the suitable height of the scout staff.	The students choosed the height of the tallest student as the size of a suitable staff for their class.	+

The limited time once again became obstacle during this lesson because the students did not have a chance to reconstruct the class chart. Using height measurement as a context apparently was also a problem because it is not the size the students were familiar with. Consequently, some of them forgot what their height is and just put their sticker randomly on the chart.

5.2.2.2 Lesson 2

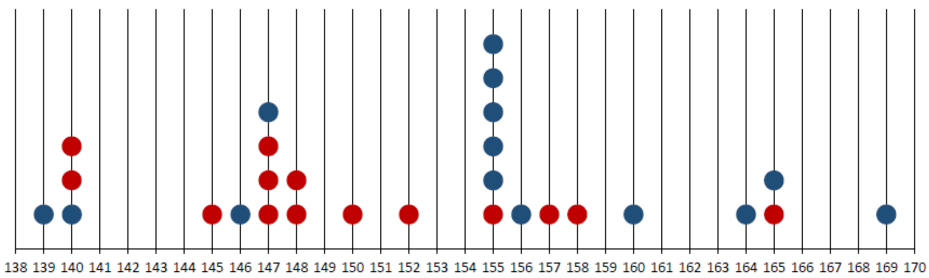
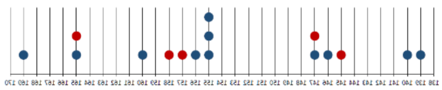
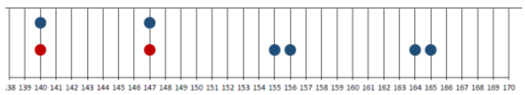
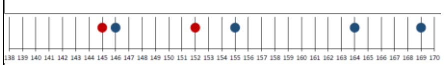
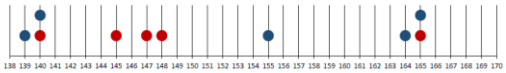
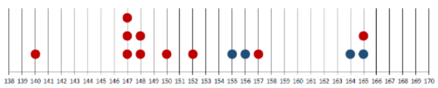
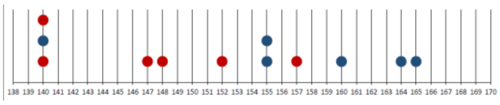
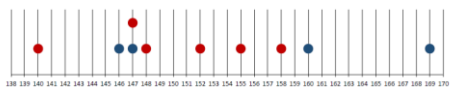
The aim of the second lesson is to develop 2 big ideas; 1) that a part of data can represent the whole (representativeness), and 2) the bigger this part of data, the more likely it represents the population (effect of sample size). There are two activities in this lesson, activity 2.1 and activity 2.2, aimed to develop big ideas 1) and 2), respectively.

Activity 2.1

The context used here is the disappearance of the class chart and the students are tasked to choose one group chart as the replacement. There is only one problem the students has to solve in this activity, which is to find the group chart that represents the class chart the best and the worst. Each group was given a copy of all the group charts as well as the class chart, and a table where they can fill out all the chart's features. The conjecture is that the students will compare the charts, analyze them visually, and use the similar or different features of the charts to determine which one is the most and least representative.

Table 5.14 shows all the group charts and the class chart.

Table 5.15 The class chart and all the group charts

<table>
<tr><td colspan="2">The class chart</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>The chart of Group 1</td><td>The chart of Group 2</td></tr>
<tr><td></td><td></td></tr>
<tr><td>The chart of Group 3</td><td>The chart of Group 4</td></tr>
<tr><td></td><td></td></tr>
<tr><td>The chart of Group 5</td><td>The chart of Group 6</td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">The chart of Group 7</td></tr>
<tr><td colspan="2"></td></tr>
</table>

As predicted, the students analyze the charts visually to determine its characteristics and filled out the characteristics in the table provided, as shown in figure 5.9. The students' answer for the most representative group chart is group chart 1. Their argument is simply that "*it is almost similar to the class chart.*" Their answer for the last representative group chart is group chart 3, and in similar fashion, their argument is that "*it really does not resemble the class chart.*"

	GRAFIK KELAS	GRAFIK 1	GRAFIK 2	GRAFIK 3	GRAFIK 4	GRAFIK 5	GRAFIK 6	GRAFIK 7
Berapakah tinggi badan siswa terpendek?	139	139	140	145	139	140	140	140
Berapakah tinggi badan siswa tertinggi?	169	169	165	169	165	165	165	169
Umumnya siswa memiliki tinggi badan .... cm	155	155	140, 147	Beda ;	140, 165	147.	140	147
Siapa yang lebih tinggi, siswa laki-laki atau perempuan?	laki-laki	laki-laki	laki-laki dan perempuan	laki-laki	laki-laki dan perempuan	laki-laki dan perempuan	laki-laki	laki-laki
Buat sketsa bentuk grafiknya! Contoh:								

Figure 5.13. The table with all the chart' characteristics

It seems that the students determined the most representative group chart by singling out the one with the most number of similar characteristic to the class chart; same goes with the least representative group chart. Therefore, it can be interpreted that the students' idea of a part of data being representative is if it contains the same information as the whole.

Since Group 1 won the chart competition, the teacher invited them to the front of the class to present why their chart win. The class discussion revealed something interesting about the students' discussion, as depicted in the following fragment.

Fragment 6

Teacher: So, apparently the chart of group 1 is the most suitable to represent the class chart, right?
Students: Yes.
Teacher: I want group 1 to come to the front of the class.
Teacher: Explain - Ravi, behave. Explain to the others why- the others, please be quiet – explain why the chart of group 1 is the one that fits the class chart the best.
Retno : Because they are both of equal ***size*** or have similar shape.
Teacher: Why do you think- Ravi, get to the front of the class – why do you think your graph chart can represent the class chart so well?
Retno : The heights of the students in the group and class chart are similar. And then … the tallest heights are the same. Then, the typical heights are also the same. And then … who is taller, boys or girls? Boys, for both charts.
Teacher: A lot of things are similar, right? While for other charts, we can't say the same thing.
Teacher: Anyone can give additional explanation on why it's the chart of group 1, not the others? Nobody?
Teacher: Why not the chart of group 3, for example?
Retno : Because it's just only…
Teacher: Only what?
Retno : It's just a straight line *makes gesture*
Teacher: Okay, why is it like that, then?
Retno : Because the ***height*** is just one (data point)
Teacher: Just one. Okay. Anything else? No? Okay, you can go back to your seat.

As evident from the fragment, group 1 also chose their own group chart as being most representative because they contain the same information as the class chart. Interestingly, their argument about the chart of group 1 being the least representative came from their shape, which can be interpreted of a budding understanding of data as aggregate.

This reasoning by shape also has some dangerous turnabout. Even though it enables the students to see data as an aggregate, they seem to treat the chart like they would a geometry figure. Some cases related to language gives light to this finding. Retno's statement about ***size*** did not refer to the number of data points in the set, like its formal definition in statistics, because it is clear the number of data points are different. The ***size*** she meant here was closer to its meaning in daily life, which was how much space the chart takes on the dot plot scale. Her statement about the ***height*** of the chart also rooted on the same problem. Unfortunately the teacher did not pay attention to these cases and discuss it further.

The following table depicts the Dierdorp analysis matrix elaborating the compatibility issue of the HLT and ALT of activity 2.1 in cycle 2.

Table 5.14 Compatibility between HLT and ALT of activity 2.1

No.	Task	HLT	ALT	Quantitative expression
1.	Which one of the group chart is the most suitable to represent the class' chart?	Analyze the charts visually then picking the chart that has similar modal clump, maximum or minimum value, or distribution of blue and pink dots.	All students detected the characteristics of the group charts. They compared and contrasted the group charts based on those characteristics.	+
2.	Which one of the group chart is the least suitable to represent the class chart? What can you do to make it more suitable?	Analyze the charts visually then picking the chart that has different modal clump, maximum or minimum value, or distribution of blue and pink dots.		+

Activity 2.2

The aim of this activity is to develop the idea of the effect of sample size. This activity is centered on one task, which is to make the least representative chart more representative. In the previous activity, group chart 1 as the most representative chart was constructed from data with the largest size. On the other hand, the least representative chart, which was group chart 1, is the data with the smallest size. This result becomes a good foundation for the current activity to be set on.

The worksheet given to the students instructed them to predict what will happen if they added a certain number of data points to the plot, and then to actually add them. The conjecture of this activity is quite straightforward. As the number of data gets closer and closer to the total number of students the class, students can see that the group chart starts to look more like the class chart.

During the implementation, however, the teacher instructed the students to add data points first. The students, as a result, described what the dot plots turned out to be, instead of predicting it.

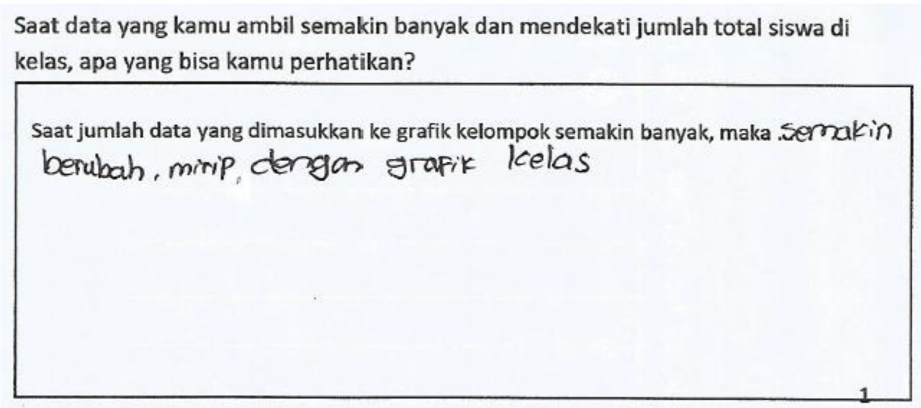
The following fragment depicts the struggle that the students experienced during the discussion.

Fragment 7

Researcher : Now explain what happened. How was it?
Tio : [inaudible]
Researcher : Try to explain. Before you added 5 more data points, chart of group 3 is really flat, right? Now after you added 5, what change did you see happening to the chart?
Tio : This one, umm …
Ayomi : 155 (pointing to the chart)
Researcher : What is with 155?
Tio : Two more people adding up …
Researcher : Go on, explain it. What else can you see? Its no longer flat, right? We have this 155 here …
Tio : There are two (more) people now …
Researcher : Now (that the chart is no longer flat) can you answer what the typical height of the students might be?
Tio : 155!
(they proceeded to add up more data points to the plot)
Hana : Boy, 165.

Ayomi : (inserting data points into the chart while comparing it to the class chart)
Tio : We're done, Miss.
Researcher : So, if you want to determine which group chart can represent the class chart the best, what would you do?
Ayomi : Adding (the data points) up …
Researcher : So do you think the size of the data has anything to do with that?
Ayomi & Tio : Yes.
Researcher : How?
Tio : It's adding up, Miss.
Researcher : So if the size of the data is small, is it possible for it to look like the class chart? Like group 3, they only have 6 points, while group 1 has 16. So what does it affect?
Ayomi : More data that you have to put into it Miss …
Researcher : Is that so? Group 3 only has 6, so it looked flat and doesn't look like the class chart while group 1 has 16 points. So?
Tio : We can see a difference here …
Researcher : So, the more data points you put into it …
Ayomi : The more it looks like the class chart!
Researcher : So do you think the size of the data is important?
Tio : Yes, it is ..

From the fragment, it can be seen that even though the students struggled to make sense of what they are doing, in the end they were able to grasp the idea that the larger the size of the sample, the more it resembles the population.


Saat data yang kamu ambil semakin banyak dan mendekati jumlah total siswa di kelas, apa yang bisa kamu perhatikan?

Saat jumlah data yang dimasukkan ke grafik kelompok semakin banyak, maka semakin berubah, mirip, dengan grafik kelas

1

As the data that you took is getting closer to the total number of students in the class, what did you notice?
If the number of data points is getting larger and larger, then [the chart] will change, it resembles the class chart.

Figure 5.14 The students' conclusion in the end of activity 2.2

The compatibility issue between HLT and ALT of activity 2.2 is shown in the following Dierdorp analysis matrix.

Table 5.17 Compatibility issue between HLT and ALT of activity 2.2

No	Task	HLT	ALT	Quantitative expression
3.	As the number of data that you take is getting closer to the total number of students in the class, what is your conclusion?	Stating that the more data value added to the group chart, the more it resembles the population.	All groups comment that the graph is getting ‘higher’ and that the typical value is getting close to that of the class chart. They also claimed that the bigger the number of data inserted into the group chart, the more it resembles the class chart.	+

5.2.2.3 Lesson 3

The aim of the third lesson is to develop two big ideas, namely 1) with a really big set of data, it is impractical to take the whole data into account; and 2) all member of population have to have equal chance to be picked as sample. The context is finding information about the height of the students in the whole school. This lesson consists of two activities, activity 3.1 and activity 3.2. In activity 3.1, the students are tasked to create a dot plots depicting the height of the students of the whole school. In activity 3.2, the students have to judge whether or not the report of the students’ height made by the Head of the Students Council is acceptable.

Activity 3.1

The aim of this activity is to build the purpose of sampling. The tasks are quite similar from before, only that this time the students are working with population box depicting a school population of 250. The conjecture is that when given a very large set of data and a very limited time to analyze them, they will see that taking all data into consideration is impractical.

The teacher asked one of the students to read the problem, then distributed the worksheet ad population box, all the while announcing that the students were going to construct a dot plot of the whole school. This chart was later referred to as the *school chart.*

At the beginning, the teacher asked whether or not the students will have difficulty. The students initially claimed that they would not find any difficulty. The following fragment shows the discussion that happened in the beginning of the lesson, when the students was constructing the chart.

Fragments 8

Researcher: Is it necessary to take all the data?
Tio : Excuse me?
Researcher: Do you have to take all the data?
Tio : Nope.
Researcher: No.
Tio : Just as needed.

It was unclear what they mean by a needed number, but it was evident that even prior the lesson, the students already know that they did not have to take all the data into account. There was also no sign of picking, but Agung did shake the box while Hana picked them randomly. It can be interpreted as a sign of randomization, although the teacher did not follow up this further.

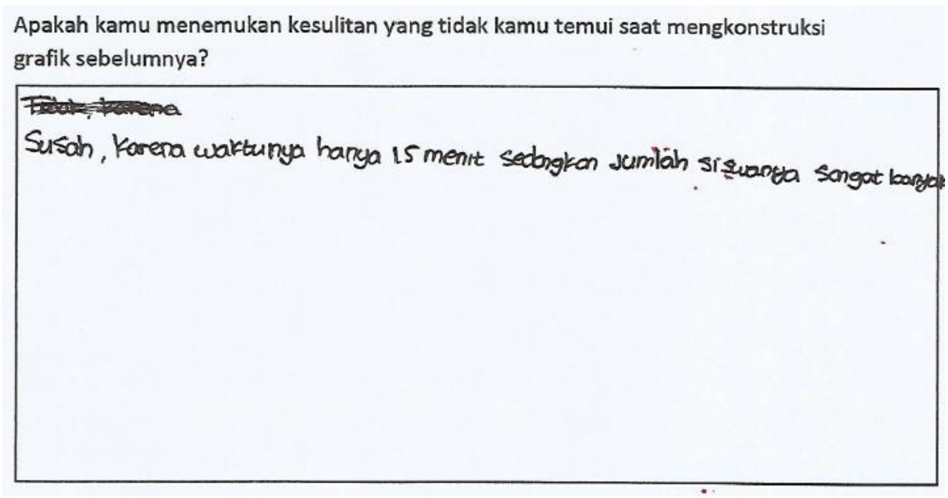
The following fragment describe the discussion that happened when the students were working on their worksheet.

Fragment 8

Teacher : Did you find any difficulty (in constructing the chart)? Did you try to insert all the data to the plot?

Students : No …
Teacher : Suppose you have to take all 250 students, and you have to insert all of them into the plot in 15 minutes, is that difficult?
Tio : Yes, Miss.
Teacher : How so?
Tio : (the data) too many, Miss.

The students did not struggle at first as conjectured because they already possess this knowledge that only part of data can represent the whole is it depicts the important information contained. In the end, they were able to grasp the other purpose of sampling, which is to save time and resource.



Did you find any difficulty that you didn't find when you were constructing the previous charts?
It was difficult, because the time is only 15 minutes while there are a lot of students.

Figure 5.15 The students' conclusion at the end of Activity 3.1

The downside of this lesson happened when the teacher reminded the students of the previous lesson, about group 1 who did not take the whole data yet their chart still depicts all important information anyway. Instead of letting al students experience difficulty, the teacher emphasized quite strongly that they don't have to take the whole data.

The following Dierdorp matrix aalysis depicts compatibility issue between HLT and ALT of Activity 3.1.

Table 5.18 Compatibility issue between HLT and ALT of activity 3.1

No.	Task	HLT	ALT	Quantitative expression
1.	Making dot plot of the height of the students in SMP Santa Maria.	Trying to put all the data into the dot plot before realizing it is too impractical	The students did not try to put in all the data, but only taking a subset of it. No consideration seem to be made in term of number of sample, they also did not seem to pick their sample either.	+
		Not trying to input all the data value and instead already taking a subset of it during the first try, but they try to get as much data value as possible into the dot plot.		
		Being picky about the data value that they put in the dot plot by selecting the paper and putting back the ones that they consider unfit for their idea of what the school chart might look like.		
		Showing the sign for randomization, either by shaking the bag or stirring the pieces of paper instead, and then taking a subset of it to be put in the dot plot.		
2.	Describe the way you create the dot	Focusing on the limited time or the number of	The students claimed that the number of data to be	+

	plot. Is there any difficulty that you did not find when constructing the previous dot plots?	data values as the obstacles in inserting the data into the chart.	put into the chart is too much while the time is limited.	
3.	What can you say about the height of students in grade 7? Which one is taller, girls or boys?	Focusing on the modal clump, the shortest or the tallest heights, or the difference between the height of girls and boys.	The students mentioned the typical heigth as well as the gender of the tallest and shortest students.	+

Activity 3.2

The aim of this activity is to foster the idea of randomization, which can roughly understood as equal chance for every member of population to be picked as sample.

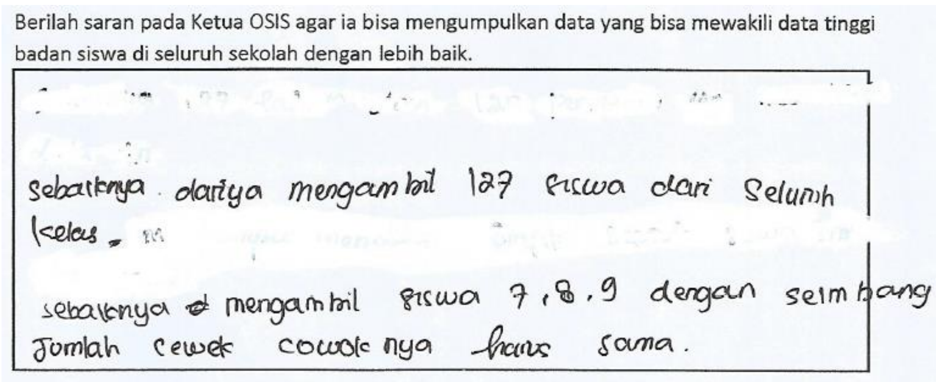
To start with, the teacher asked one student to read the worksheet. The context used here is a story about Daria, the Head of the Students' Council of SMP Labschool. Daria writes a students' health report that includes information about the students' height. However Daria, who is a 9th grade female students, decide to only took her data mostly from female students in her grade. The reason is because she is too shy to talk to the boys, as well as for the sake of convenience. The claim made by Daria is that the typical height for the students in the school is 157-158 cm and the girls are taller than the boys.

The students were tasked to judge whether or not this claim is suitable and if not, to offer a better way in collecting data for Daria. The following fragment shows the discussion that happen during the lesson.

Fragment 9

Teacher : can you read the conclusion out loud for me please?
Tio : The typical height is 157-158 cm and the girls are taller than the boys.
Teacher : It's that a correct conclusion?
Ayomi : No
Teacher : Why?
Ayomi : It's because – it's because she was shy to interview the boys ..
Teacher : Hmmm. What do you think?
Tio : Yes?
Teacher : Why?
Tio : Umm ..
Teacher : Okay, can you please look at the chart. The chart did say that the girls are taller than the boys, and the typical height was so-so. But like Ayumi said, is this correct?
Tio : Yes.
Teacher : For all students in the school?
Tio : Oh. Well, wrong then.
Teacher : Why is it wrong?
Tio : [tried to answer but failed]
Teacher : You answered just now. It's because she only took data from …
Ayomi : only from the girls …

For an ideal way to collect the data, the students emphasized on the same number of students for both genders as well as all the grades. The following conclusion that they give in the end of activity 3.2 gives light to this.


Berilah saran pada Ketua OSIS agar ia bisa mengumpulkan data yang bisa mewakili data tinggi badan siswa di seluruh sekolah dengan lebih baik.

sebaiknya dariya mengambil 127 siswa dari seluruh kelas.

sebaiknya mengambil siswa 7, 8, 9 dengan seimbang Jumlah cewek cowok nya harus sama.

Give an advice to the Head of the Students Council about a better way to collect data.
It's better if Daria took 127 students from all class.
*It's better to take a **balanced** number of students from grades 7, 8, 9. The number f boys and girls have to be equal.*

The number 127, which more or less a half, can be interpreted as the number of sample deemed by the students as large enough without having to take the whole data. They also came up with the word "*seimbang*" (in English, ***balanced***) to describe their ideal way of collecting data. Comparing with the result of Cycle 1, this can be seen as close to the word "rata" (in Englis, ***even***) by the students in Cycle 1.

Although the students did not come up with any informal words that can stand in the place of ***random*** or ***bias***, the fragment shows that they could conceive an initial understanding about the importance of all member of population having equal chance to be picked as sample. However, even though they can detect bias in the data collection, they still have a hard time coming up with a randomized way of collecting data.

The following Dierdorp analysis matrix depicts compatibility issue between HLT and ALT of activity 3.2.

Table 5.17 Compatibility between HLT and ALT of activity 3.2

No.	Task	HLT	ALT	Quantitative expression
1.	The Head of the students' council derive the following conclusion, "***In SMP Lab School, te students' typical height is 157-158 cm and girls are taller than boys***." Do you agree with this conclusion?	Claiming that it is reasonable because the dot plot says so.	The students did not agree with this conclusion because the head of the students' council chose more girls than boys.	+
		Claiming that it is unreasonable because from previous results, boys are taller than girls.		
		Claiming that it is unreasonable because the size		

		of the data collected is not big enough.		
		Claiming that it is unreasonable because the head of students' council is *picky* or *unfair* in collecting the data.		
2.	Give an advice to Daria so she can collec data in a better way!	Stating that the head of the student's council should be fair in the data collection, for example by going to class to class and taking equal number of data from every class.	The students advised what they think a suitable size of sample as well as taking equal number of students for both genders and all grades.	+

5.2.2.4 Lesson 4

The sequence of lessons was ended with a mini research project. Here, the students are expected to employ the informal statistical ideas they have learned in the previous lessons. The context is finding out information about social media addiction among grade 7 students in Laboratory Secondary School of State University of Surabaya.

To save time, the teacher introduced and discussed the context with the students in the end of Lesson 3. Each group was given a report form with a table and a dot plot scale included. There was a few days gap between Lesson 3 and Lesson 4 where the students were expected to work on this report. All groups then presented the result of their mini-research in the Lesson 4.

In general, the presentation by the students in the Cycle 2 was much better than those in the Cycle 2. The students talked more and were able to tell their result like a story, instead of question-and-answer problems. The teacher also emphasized a lot on the correct way to open, carry out, and close a presentation. The report of the students did not reveal the students' thinking significantly because of their reluctance to write and to express opinion.

The class discussion revealed that the students possessed an understanding about sample and population because they were able to make a statement about all students in the school by only collecting, representing, and analyzing data taken only from a part of the population. The students referred to their charts for every claim that they make. Therefore, we can interpret that they were able to make generalization from sample to population and use data to back up this generalization.

However, even though that the students already possessed a sense of which one the sample and which one the population, the way they collected their sample still suggested that they were still having difficulty in applying the idea of a representative sample in the process of data collection. In term of size, they only collected data as much as the given table allowed them to. They also did not consider how many sample size will be representative for a population of 200 students. Meanwhile, in term of randomization, the students did attempted at giving equal chance to every member of the population, based on their grade and gender, even though in the end they still collected data from their own class the most.

The students acknowledged this weaknesses and related it to the ability of their sample to represent the population. The following fragment, a presentation by Group 3, depicts this.

Fragment 10

Rafli : We are from the Group 3 and we are going to present-pre[had a hard time pronouncing presentation] how many hours the students of SMP Lab spend their time in social media sites everyday. We collected data from the students of SMP Lab [inaudible] from some students in the grade 9, 7, and 8. We collected this data at Monday, Tuesday, and Wednesday.
Teacher : Okay, what's the result then?
Rafli : The longest is 10 jam. The shortest is 1 hours. At average is 3 hours.
Teacher : The averae is 3 hours. Can you show us your chart?
Rafli : [shows his chart]
Teacher : [addresses the other students in the class] So guys, what do you think? Can their data represent all students in SMP Lab?
Students : No.
Teacher : Why? Gina, can you explain why?
Gina : because it is only some students.
Teacher : Well, do you have to ask every students?
Gina : No, Miss. But [in our data] is mostly from grade 7. And it's mostly girls.

There was also an interesting point of discussion centered on the students' attempt at randomization. The students discussed about their data being not ***balanced*** (in Bahasa, ***seimbang***) enough, a word that has come up previously in the Lesson 3 where the students were introduced to randomization.

In the next fragment, it can be seen an expression of uncertainty that the students make in deriving conclusion. It was a presentation by Group 4, albeit incomplete because one of the member, Ayumi, isn't present.

Fragment 11

Teacher : Now I want to ask you- well the others can answer too- are you sure about the conclusion that you make about the students of SMP Lab typically spend 2-3 hours on social media sites?
Tio : I think so.

The students: Not yet.
Teacher : Why not? Tio?
Tio : Because it is not balanced.
Teacher : Not balanced? Not balanced how?
Tio : [inaudible mumble]
Teacher : Speak clearly, Tio.
Tio : The students from grade 8 and 9 are very little.
Teacher : So next time you are doing a research, what do you have to do? Taking all students from grade 7, 8 and 9?
Tio : Well doesn't have to be like that, just a part of it.
Teacher : A part of it? Like how?
Tio : It has to be ***equal.*** Like, ***balanced.***

The word ***balanced*** has come up before, but Tio interestingly expressed his thought in other word, which is ***equal.*** We interpret this as his attempt at explaining what a good sample is like. In general, the students used the sampling process as the source of certainty (or uncertainty) of their conclusion, even though there has to be a lot of guidance by the teacher for them to do so.

Table 5.18 depicts compatibility issue between the HLT and ALT (in line with elements of Informal Inferential Reasoning) of Lesson 4.

Table 5.20 Compatibility issue between HLT and ALT of Lesson 4 in Cycle 2

Task	Elements of IIR	HLT	ALT
A newspaper article have revealed worrying trend among the teenagers; social media addiction. Aside from robbing them from the opportunity to actually experience genuine human interaction, this also increases cyber-crime rate. The study this article refers to claim that ***Indonesian teenagers spend up to 12 hours a day on the social media sites.*** Now your assignment is to become a researcher and invesigate whether or not this claim is correct for the students of Lab Secondary School. Use the resul of your research to answer the following question: ***How many hours in a day the students of Lab Secondary Schhol usually spend their time on social media sites?***	Making generalization from sample to population	Considering all data in deriving conclusion about a population.	The students collect data from some parts of the school and making conclusion that apply for all students in the school.
		Considering only part of data in deriving conclusion about a population.	
	Use data to back up this generalization.	Personal experience overrule data and context.	When deriving conclusion, the students refer to their dot plots without considering the context.
		Over-reliance on context over data or vice-versa.	
		Consider both data and context together when drawing informal inferences; may start to integrate appropriate statistics.	
	The use of expression of uncertainty.	No expression of uncertainty; the students derive deterministic conclusion.	They can detect and describe the flaws and the strength in their sampling process. They expressed uncertainty which was contributed to the quality of their sample.
		Presence of expression of uncertainty, attempt at justifications which can be based on context, data, or personal experience.	
		Presence of expression of uncertainty, with appropriate justification based on data (method of sampling, the method of data collection, etc.)	

5.2.3 Post-Assessment

The post-test was conducted in the end of cycle 2. The test consisted of similar problem as those in the pre-test, although the context was altered a bit. In this test, the data used was virtual and presented in the form of bar chart, depicting pet rabbit ownership among citizen of a certain city in Indonesia. The sample size is 96 and taken from three ethnicities; Balinese, Sundanese, and Maduranese. The chart shows that the highest pet rabbit ownership goes to the Sundanese, followed by Balinese and Maduranese.

The explanation about the result of the pretest will be organized in two parts; descriptive statistics and inferential statistics. The first and second question are focused on descriptive statistics. There are some questions about the graph that the students have to answer, and then they are asked to mention 1 thing that they find most noticeable about the graph. All students can answer the first sub-question correctly, showing that they understand the information presented in the bar chart.

For the second question, there is a subtle improvement from the pre-test. Most of the students could detect aspects that they found most interesting about the charts. A lot of students are focused on the sampling process that is unfair. Some students went totally irrelevant, by describing the purpose of the chart, for example.

The third and fourth question start to move into the range of inferential statistics. The students are asked to determine whether or not someone owns a pet rabbit if they say they are from Madura. Students are then asked to provide reasoning for their answer, referring to the chart.

The difference from the cycle 1 is that in this cycle, the students derived deterministic conclusion based solely on what they can see from the chart. In general, the students were still thinking of the data descriptively. Table 5.19 provide alignment issue between the result of the pre-test and the element of IIR.

Table 5.21 Alignment issue between the result of post-test and the element of IIR

Framework	Comments
Generalization beyond the data.	The students made claim about a population based on what they can notice from sample.
Data as evidence	All students when asked to argue about their answer, refer to the chart provided.
The use of probabilistic language	No expression of uncertainty.

5.2.4 Discussion

The result of the second cycle reveals various interesting findings. Making generalized statement about a population seems to depend on teacher's prompt, meaning the students did produce this kind of statement naturally without claiming that they do not know. Most students argued with data, instead of with their personal knowledge. In regard to statement of uncertainty, instead of contributing their certainty or uncertainty to the quality of their sampling, the students chose to contribute to sampling variability, which seemed to come naturally to the students. There were also cases related to mixing up daily and statistical words that came up.

The result of post-test in both cycle shows no significant difference from the pre-test. However, this is predictable since informal inferential reasoning is not something that can be fully evoked from the students in a very short period of time.

The students' written work and the fragment of class discussion already shows what can be interpreted as a budding understanding of informal inferential reasoning.

Due to the limited time and availability of the school, this study was conducted in very short time. This serves as a major drawbacks during the study. As a reflection, the researcher would like to propose several revision to the study. The most major change is related to the fact that this study basically aimed at developing too much in too short of time. Therefore, in general, longer timespan for each cycle is needed so there is time for the students to truly conceive each big ideas. In fact, each big idea in this lesson could be a topic of research of their own. It will help if there are a few lessons dedicated solely for the students to exercise making prediction about a population from a sample.

A more detailed refinement includes the use of more familiar measurement for the students. This study uses height measurement, which in fact the students are not too familiar with. As a result, the students often forgot their own height which sometimes delays the lesson. Other measurement like shoe size will be much more familiar for the students.

The researcher also would like to suggest to pay more attention to the statistical language. Statistical language is actually designed to be a part of this study, but due to the limited time and the density of the topic, the teacher did not pay sufficient attention to it.

CHAPTER VI

CONCLUSION

6.1 Conclusion

The aim of this study is to contributing a local instruction theory on statistics education, especially informal inferential reasoning. Based on the finding of this study as well as the retrospective analysis, we can answer the research question: "*How can 7th grade students be supported to develop Informal Inferential Reasoning?*" as follows.

1. Exposing a concrete form of sample and population enable the students to actually conceive the idea that a part of data can represent the whole. The students were able to see which characteristics of population represented in the sample, hence ease them when they finally work with and make inference about an unknown population.
2. The growing sample activity enables the students to conceive the idea about the effect of sample size. The students were able to physically see that the bigger the size of the sample, the more it resembles the population and contain similar information.
3. The judging claim activity, although can help the students to conceive an initial understanding, still cannot fully foster the idea of randomization. Even though the students can detect bias in the data collection, they still have a hard time coming up with a randomized way of collecting data.

4. The concept of sample size and randomization to determine sample quality did not really help with the students' inference. Most of the time the students produce deterministic statement. In cases when they expressed a statement of uncertainty, they did not contribute their certainty or uncertainty to the ability of their sample to represent the population. In the first cycle, the students contributed it to the sampling variability, which seemed to be unintentionally introduced to the students by the comparing group charts activity.
5. The mini research project enables the students to experience how researches are done and statistical result are derived. They learn to base their argument with data as well as being critical to others about doing so.
6. The use of dot plots enable the students to explore data visually and easily detect the features of the data. Instead of being dependent on formula and calculation, the students can see the data as aggregate instead of a bunch of individual and numerical values.
7. The use of informal language create a relaxed environment where the students can present their own thoughts in their own language. It also gives light to interesting cases involving language, for example the tendency of the students to treat the charts like they are a geometric figures. However, the density of the topic and the limited time caused the informal language to be neglected. Even though the teacher made an effort to emphasize on the informal words the students produced over the course of the lesson, she did not introduced the formal equivalence of those words to the students, for fear that the students might become overwhelmed.

8. The students' reluctance to express their opinion either spoken or written becomes a major drawbacks, especially during the mini-research project.

To conclude, the designed learning instruction, although can support the first two characteristic of IIR, which are 1) making inference about an unknown population and 2) base this inference with data, cannot fully support the third characteristics, which is the use of expression of uncertainty. The students who expressed uncertainty regarding their inference, referred to the sampling variability as the cause of their uncertainty. Other students produce deterministic statements as inference, instead.

As the aim of this research is to contribute to the local instruction theory, we provide the LIT in the following table.

Table 6.1 Local Instruction Theory on Developing IIR

Tools	Imagery	Activity	Concepts
Population box, red and blue stickers, red and blue markers.	Enable the students to collect the data easily and give them a sense to actually take some data from a population.	Collecting data from the population box and represent them in the form of dot plots.	Collecting and representing data.
Dot plots, a group chart, the students' height data, red and blue markers.	Helps the students to predict the population based on a sample.	Predicting the class chart based on the group chart.	Introduction to inference.
Several sample charts from the same population	Demonstrating that a sample chart can have similar characteristics as	Comparing the group charts and determine which one can represent	A part of data can represent the whole.

in the form of dot plots.	the chart of the population.	the class chart the best.	
A sample chart, red and blue markers, a population box.	Demonstrating that the bigger the size of the sample, the more it resembles the population.	Growing sample activity.	Effect of sample size.
Students' height data, biased data collection context.	Demonstrating a biased and unfair way in data collection.	Judging whether or not the claim made over the sample about the population is suitable.	Randomization.
Mini-research form, a research question.	Introduce the students to scientific research.	Conducting an actual field research.	Using their informal knowledge to make informal inference about a population.

6.2 Weaknesses of the study

During the implementation of the study, we faced some obstacles that cannot be avoided and affect the study quite significantly. We described them as weaknesses of the study, as follows.

1. Time limitation and the school availability.

The study was conducted in a fairly short time span. We were faced against the school exams and the dense curriculum. Hence, the school did not give us enough time to conduct the study. Therefore, we stopped until the second cycle. Methodologically, the cycle should be continued until the HLT can accommodate the students' conjectures and also reach the goal for every lesson. The effect of time limitation on the result of the study was also magnified by the fact that this is a very dense topic, hence the aspect of informal statistical language could not be paid attention to.

2. The social norms and the socio-mathematical norms.

Even though in general the students are quite outspoken, they were not used to expressing their opinion, either spoken or written. They hesitated a lot and everything has to be confirmed as right or wrong. They also had a hard time getting out of question-and-answer type of mathematics problem. In the mini-research project, this becomes a major drawback because the students list their findings instead of writing, and present their research result like answering questions to a homework.

3. The students' learning style

In this study, we consistently use *exploratory data analysis* (EDA), which mostly about analyzing and summarizing a set of data mainly with visual methods (Paparistodemou & Meletiou-mavrotheris, 2008). The students are encouraged to loot at their chart and determine its features visually. Therefore, this study did not focus on the learning style of the students. We designed the activities and contexts without considering the different learning style that the students might have.

4. The unfamiliarity of the topic

Informal Inferential Reasoning is a new topic in statistics education. Although being discussed widely in International Mathematical Journal, its implementation in classroom have not happened that long yet. It is even a much stranger topic in Indonesia. To bridge this, the researcher provide an explanation about IIR in the teacher guide, but the teacher still have a hard time to actually discuss this topic in the classroom. It is also not included in the curriculum, hence the teacher hesitated to add this on top of the already dense teaching plan.

6.3 Suggestion

Based on the obstacles and constraints happened during the course of the study, the researcher propose s several suggestion that might be useful for the future researcher who might be interested in carrying out this study.

Firstly, time. This study wanted to develop such a dense concepts in a very short period. More meeting in each cycle might be useful, and it will be better if the students have a chance to focus on simply making inference about an unknown population for several meetings, instead of jumping to statistical ideas like sample size or randomization. On the second thoughts, every big ideas in this study can be a topic of an individual study.

Secondly, as an implication of the first one, this study only offer one activity for the students to learn about the effect of sample size and randomization. They need more time to learn about effect of sample size and randomization before they can connect them to the quality of their sampling and use it in their certainty or uncertainty.

Thirdly, the teacher's understanding. Informal Inferential Reasoning is a new topic internationally and even stranger in the world of Indonesian education. It is suggested for the teachers to learn about IIR individually before conducting the lesson.

And then, the last, in this study we focus on the students' height data that we believed is a familiar context for the students hence can engage them further in the problem. However, various domain of contexts and subjects can also be considered. Therefore, we suggest for the further researchers to study on different contexts and application, keeping in mind the ability of the context to engage the students within it. The context in the third lesson, even though can evoke the intended learning goal, is not realistic and might encourage an indecent characteristics among students. Therefore, a change in the context is suggested.

REFERENCES

Bakker, A., & Derry, J. (2011). Lessons from Inferentialism to Statistics Education. *Mathematical Thinking and Learning, 13*(1-2), 5-26. doi:10.1080/10986065.2011.538293

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.), *the challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht: Kluwer Academic Publishers.

Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt, & J. Ainley (Ed.), *Reasoning about Inforla Inferential Statistical Reasoning: a collection of current research studies.*

Biehler, R. (2014). On the delicate relation between informal statisical inference and formal statistical inference. In K. Makar, B. de Sousa, & R. Gould (Ed.), *Sustainability in statistic education. Proceedings of the ninth international conference on teaching statistics (ICOTS9).* Arizona, USA: International Statistical Institute (ISI/IASE), Voorburg, The Netherlands.

del Mas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi, & J. Garfield, *The challenge of developing statistical literacy, reasoning and thinking* (pp. 79-96). Dordrecht: Kluwer Academic Publishers.

Eerde, D. v., & Hajer, M. (2005). Language sensitive math teaching in a multicultural classroom: students' talking and writing enlightens hidden problems. *Fourth Congress of the European Society for Research in Mathematics Education (CERME 4)*, (pp. 1215-1225).

Eerde, H. v. (2013). Design research: looking inot the heart of mathematics education. *Proceeding The First South East Asia Design/Development Research (SEADR) International Conference* (pp. 1-10). Palembang: Sriwijaya University.

English, L. d. (2014). Establishing statisical foundations early: data modelling with young learners. In K. Makar, B. de Sousa, & R. Gould (Ed.), *Sustainability in statistics education. Proceedings of theNinth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, the Netherlands: International Statistical Institute.

Gal, I. (2002, April). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review, 70*(1), 1-25. doi:10.1111/j.1751-5823.2002.tb00336.x

Garfield, J. (1995, April). How Students Learn Statistics. *International Statistical Review, 63*(1), 25-34. Retrieved from http://www.jstor.org/stable/1403775

Garfield, J., & Ben-Zvi, D. (2004). Statistical liteacy, reasoning, and thinking: goals, definition, and challenges. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-15). Dordrecht: Kluwer academic publisher.

Gravemeijer, K. (1994). *Developing Realistic Mathematics Education.* Utrecht: Freudenthal Institute.

Gravemeijer, K.P.E., & Cobb, P. (2006). Design research from a learning design perspective. In J. van den Akker, K. P. E. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). London: Routledge.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and Teachers' Knowledge of Sampling and Inference. In C. Batanero, G. Burril, & C. Reading, *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer Netherlands. doi:10.1007/978-07-1131-0

Leung, C. (2008, December 22). Mathematcial vocabulary: fixers of knowledge or points of exploration. *Language and Education, 19*(2), 126-134.

Makar, K., & Rubin, A. (2009, May). A Framework for Thinking about Informal Statistical Inference. *Statistics Education Research Journal, 8*(1), 82-105.

McPhee, D., & Makar, K. (2014). Exposing young children to activities that develop emergent inferential practices in statistics. In K. Makar, B. de Sousa, & R. Gould (Ed.), *Sustainability in statistics education. Proceedings of the ninth international conference on teaching statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, the Netherlands: International Statistical Institute.

Paparistodemou, E., & Meletiou-mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statsitics Education Research Journal, 7*(2), 83-106.

Pfannkuch, M. (2006). Informal Inferential Reasoning. *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazil* , (pp. 2-7).

Pratt, D., & Ainley, J. (2008, November). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal* , *7*(2), 3-4.

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008, November). Local and Global Thinking in Statistical Inference. *Statistics Education Research Journal, 7*(2), 107-129.

Rossman, A. J. (2008, November). Reasoning about Informal Statistical Inference: One Statiscian's View. *Statistics Education Research Journal, 7*(2), 5-19.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114-145.

Thompson, D. R., & Rubenstein, R. N. (2000). Learning Mathematics Vocabulary: Potential Pitfalls and Instructional Strategies. *The Mathematcis Teacher*, 568-574.

Utts, J. (2003). What Educated Citizens Should Know about Statistics and Probability. *The American Statistician, 57*(2), 74-79. doi:10.1198/0003130031630

Wallman, K. K. (1993, March). Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association, 88*(421), 1-8. doi:10.1080/01621459.1993.10594283

Walpole, R. E. (1995). *Pengantar Statistika* (3 ed.). (B. Sumantri, Trans.) Jakarta: PT. Gramedia Pustaka Utama.

Walshaw, M., & Anthony, G. (2008, September). The teacher's role in classroom discourse: a review of recent research into mathematics education. *Review of Educational Research, 78*(3), 516-551.

Watson, J. (2008). Exploring beginning inference with novice grade 7 students. *Statistics education research journal, 7*(1), 59-82.

Watson, J. M., & Kelly, B. M. (2008). Sample, random and variation: the vocabulary of statistical literacy. *International Journal of Science and Mathematics EDucation*, *6*, 741-767.

Weinberg, A., Wiesner, E., & Pfaff, J. T. (2010). Using informal inferential reasoning to develop formal concepts. *Journal of Statistics Education, 18*(2), 1-23.

Wilkinson, L. (1999). Dot Plots. *The American Statiscian, 53*(3), 276-281. doi:10.1080/00031305.1999.10474474

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.

Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008, November). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.

Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008, November). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal , 7*(2), 40-58.

Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.

APPENDICES

---------------*o0o*---------------

Teacher's interview scheme
Classroom observation scheme
Pre-test
Post-test
Learning line
Students' worksheet
Teacher's guide

---------------*o0o*---------------

TEACHER INTERVIEW SCHEME

THE TEACHER'S BACKGROUND

- What is your educational background? Was your education in mathematics teaching, education in general, or pure mathematics?
- How long have you been teaching?
- Do you used to like statistics as a student or in college?

THE TEACHING

- What is your experience in teaching mathematics in the 7th grade?
- What kind of textbook do you use? Are you satisfied with it?
- Do you think following textbook is a must?
- Do you usually use media in the classroom? If yes, what are they?
- Can you explain briefly your style of teaching?
- What are your opinion on group's discussion?
- How often do you assign students to work in groups?
- What kind of norms you usually establish in the classroom?
- What do you do to enforce these rules?

THE TOPIC

- Have you had any experience in teaching statistics before, especially for 7th grade?
- How many meetings do you usually need to teach the content of statistics mandated by the curriculum?
- The statistics content in 7th grade is a new addition due to the Curriculum 2013. What is your opinion on the statistics content for the 7th grade?
- What kind of difficulty do you have in teaching statistics to 7th grade? Do you have any idea how this difficulty occur?
- Do you think mathematical and specifically statistical vocabulary are important?

- What do you usually do to introduce technical words to students?
- What do you think of statistical literacy?
- Do you think being statistically illiterate is important for people to function in society?
- Have you heard about Informal Inferential Reasoning before?

PMRI

- Do you use contextual problems in the classroom? Why?
- Have you ever heard about PMRI?
- Where did you find out the information regarding PMRI?
- What is your opinion on PMRI as an approach to teach mathematics?
- Do you think PMRI is a suitable approach for 7th grade statistics?

THE STUDENTS

- How old are your students?
- How varied are the heights of the students in the class?
- Is there any Scout? How many of them are?
- How varied are the mathematical ability of the students in the class?
- Are your students opinionated?
- How did you group your students?

CLASSROOM OBSERVATION SCHEME

THE CLASSROOM ENVIRONMENT

- How does the teacher start the lesson?
- How is the sitting arrangement in the classroom? Is there any specific arrangement?
- How close are the teacher and the students? Are the students intimidated by the teacher?
- Where does the teacher usually position themselves in the classroom?
- Do the students work in groups or individually?
- Does the teacher establish and/or enforce any kind of social norms in the classroom?
- Do the students follow this rule?
- What does the teacher do to enforce the rules?
- How is the students' interaction with each other?
- Are the students talkative and opinionated?
- Are the students engaged with the teacher's teaching? What does the teacher do to ensure this?
- How does the teacher end the lesson?

TEACHING AND LEARNING PROCESS

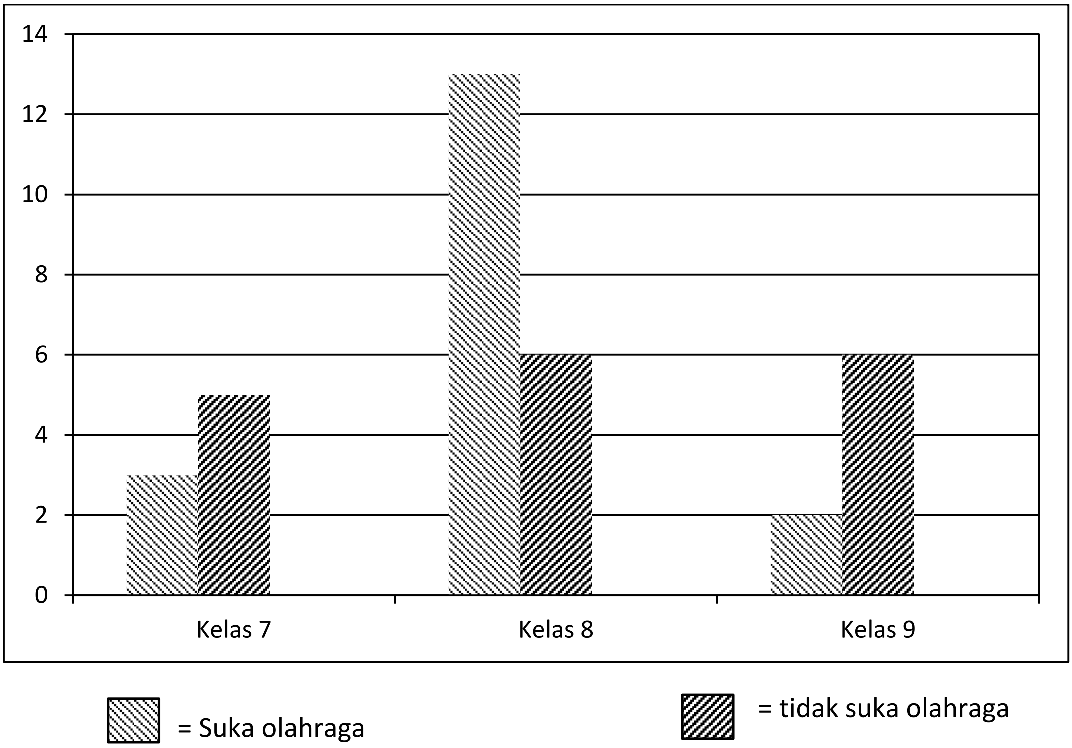
- How do the teacher teach mathematics in the classroom?
- How unconventional is the teacher? Does he or she make sure to follow the textbook? Does he or she improvise with classroom condition?
- How responsive is the teacher with the student's response, curiosity in learning mathematics?
- Is there any use of media or tools?
- Is there any use of contextual problems? Does the teacher make sure to connect their content to everyday life?
- How does the teacher deal with time limitation? Is there any strict time management in the classroom?

- Is there any classroom discussion? How does the teacher guide them?
- How does the teacher facilitate the difference solutions between the students?
- Does the teacher emphasize any form of sociomathematical norms in the classroom?
- Does the teacher pay attention to mathematical terminologies?
- Is there any specific way the teacher use to introduce mathematical vocabulary in the classroom?

PRE-TEST

Ketua OSIS SMP Lab mengadakan survey tentang minat baca diantara siswa. Dari 250 orang siswa di SMP Lab, Ketua OSIS memilih 35 siswa yang diambil dari kelas 7, 8, dan 9, kemudian bertanya apakah mereka suka membaca atau tidak.

Grafik berikut adalah hasil wawancara tersebut.



a) Ada berapa banyak siswa kelas 7 yang suka membaca? Bagaimana dengan kelas 8?

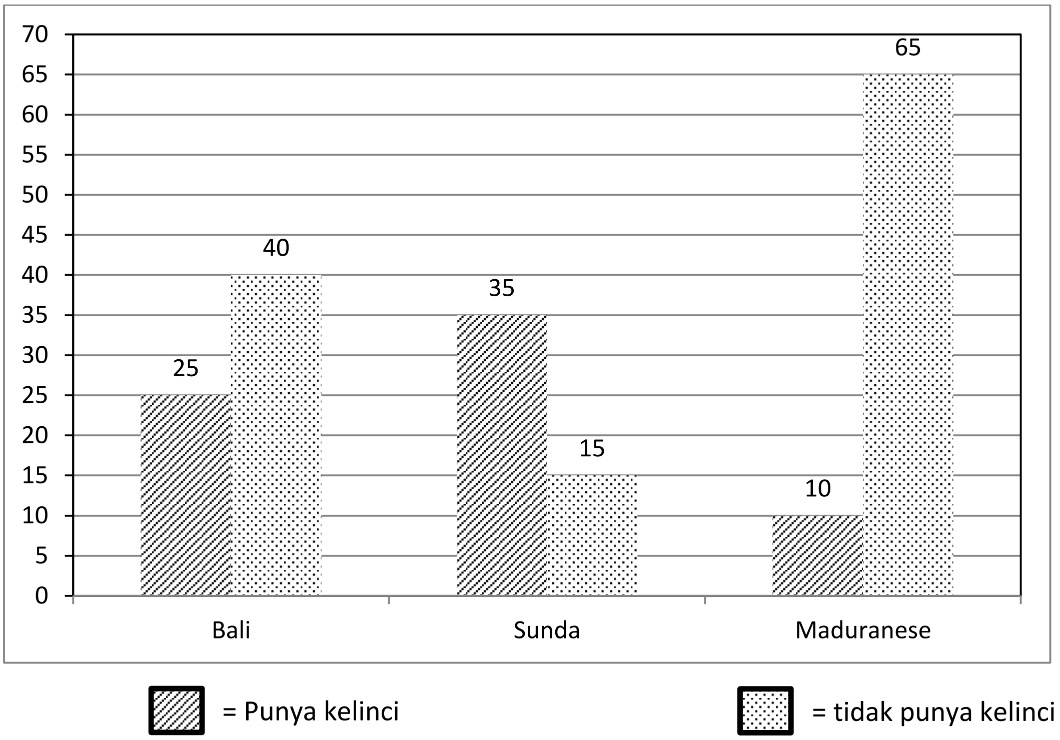
b) Apa informasi menarik yang bisa kamu peroleh mengenai minat baca siswa di SMP Lab?

c) Ketua OSIS hanya memilih 35 orang siswa, dari 250 orang siswa. Jadi jika kamu bertemu seseorang di aula dan ia berkata bahwa ia suka membaca, menurutmu siswa kelas berapakah dia?

d) Jelaskan alasanmu untuk jawaban c.

POST TEST

Pemerintah menyelidiki kepemilikan kelinci di masyarakat. Untuk memulai investigasi, mereka memilih 96 orang secara acak dari berbagai macam suku, yaitu Bali, Sunda, dan Madura, dan bertanya apakah mereka memiliki kelinci atau tidak. Grafik berikut adalah hasilnya.



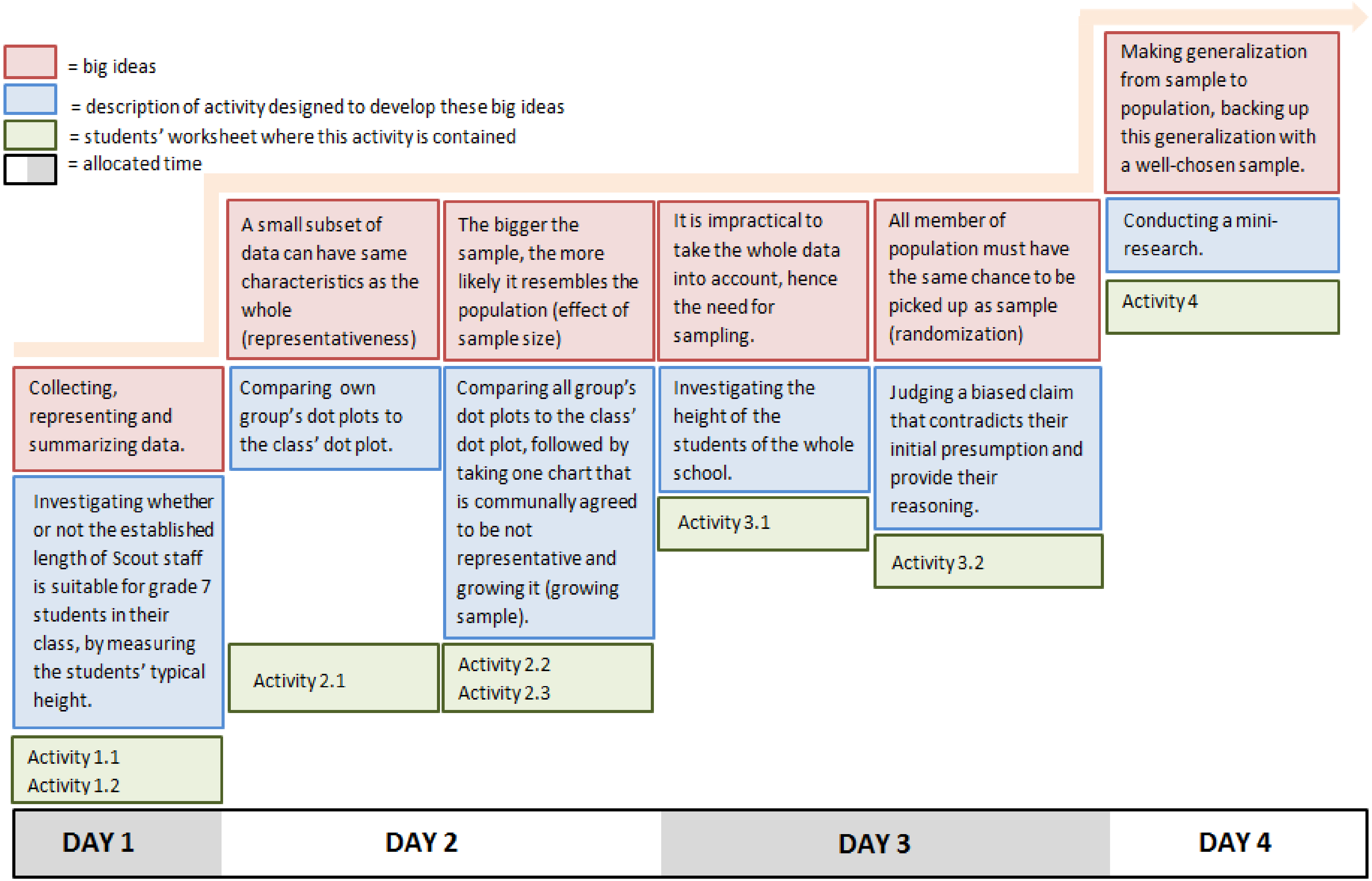
a) Ada berapa banyak suku Madura yang memiliki kelinci? Bagaimana dengan suku Sunda?

b) Apa hal menarik yang bisa kamu perhatikan dari grafik diatas? Coba ceritakan grafiknya dalam satu kalimat saja.

c) Andaikan kamu bertemu seseorang dari Madura. Apakah menurutmu orang tersebut punya kelinci di rumahnya?

d) Jelaskan alasanmu untuk jawaban c.

LEARNING LINE



STUDENTS' WORKSHEET

TEACHER'S GUIDE