Analyzing the Statistical Reasoning Levels of Pre-Service Elementary School Teachers in the Context of a Model Eliciting Activity

Cigdem Alkas Ulusoy, Mesture Kayhan Altay
Hacettepe University

To cite this article:

Analyzing the Statistical Reasoning Levels of Pre-Service Elementary School Teachers in the Context of a Model Eliciting Activity

Cigdem Alkas Ulusoy, Mesture Kayhan Altay

Article Info	Abstract
	The purpose of this study is to analyze the statistical reasoning levels of pre-service elementary school teachers. With this purpose, pre-service teachers consisting of 29 groups worked on a model eliciting activity (MEA) in scope of an elective course they were taking. At the end of the class, they were asked to present their solutions while working on the MEA in form of a detailed report. The data of the study consisted of these reports and solution sheets. Content analysis method was used in the analysis of the data. As a result, it was found that when the participants were asked to interpret a data set in a table, they could not establish a relationship between measurements of central tendency and variation, and their reasoning was limited or mistaken. The general tendency when pre-service teachers encounter a data set is that they think the only value representing the data set is the arithmetic mean. Additionally, it was found that, although the pre-service teachers were able to correctly compute the measures of variation such as standard deviation and interquartile range, they did not have sufficient knowledge about what these measures tell us about the variation of the data set.

Introduction

Even if we are not a statistics expert, while we are examining data in our hands, some points might get our attention and inspire curiosity in us. Do the data consist of values close to each other, or are there outliers? If we wanted to represent this value with only one number, what would it be? How do we proceed to reach this value that will represent the data set and maybe lead us to reach the correct result? Such questions may run through our heads. As a matter of facts, all of us encounter large amount of data in our daily lives without knowing about it, and use statistics to make a correct decision regarding these data. In the decisions we make, while some of us use statistically correct reasoning, some of us may tend toward wrong decisions and be mistaken because of the limitations in our knowledge or experience. The importance of statistical reasoning, which is significantly effective in the decisions we make, is undeniable. So, how can we define statistical reasoning?

Chervany, Collier, Fienberg, Johnson and Neter (1977), who were some of the people who defined statistical reasoning first, defined it as: a) what a student is able to do with statistical content (for instance, recalling, recognizing, distinguishing statistical concepts), and b) the skill shown by students in using statistical concepts in specific problem solving steps. On their definition, the researchers also added the property of statistical reasoning that it is not a directly observable process. The researchers, starting with this property of statistical reasoning, indicated that this skill may only be observed while working on a specific task. Statistical reasoning includes making sense of statistical information and defining statistical ideas by ways of reasoning (Garfield & Gal, 1999). These include presentation of data, interpretation of the data set, or statistically summarizing data. Ben-Zvi and Garfield (2004) also included relating one statistical concept to another (for instance tendency and variation) in the definition of statistical reasoning.

Ben-Zvi and Garfield (2004) reported that statistics education is found difficult by students and terrifying by teachers. While teachers were summarizing why students find it difficult to learn statistics, they stated that many statistical ideas and rules are difficult, complicated and illogical, and thus they experience difficulty in motivating students for learning. They also said many students found it difficult to understand the underlying mathematics of statistical concepts, the contexts statistical problems are presented in was misleading, and they had to rely on their experiences or instincts instead of choosing a suitable statistical process while making decisions. They also emphasized that students see statistics as mathematics and therefore focus on numbers, operations, formulas and a single correct answer; and situations such as dispersion in data, different interpretations that change according to different assumptions, disturb them.

The thing needed to get rid of the difficulties students experience regarding statistical concepts is to establishment of an effective learning environment where relationships are established among concepts. The biggest task in achieving conceptual learning is that of the teachers. Therefore, it is important to observe the statistical reasoning of teachers as much as it is important to observe that of students. In this study, statistical reasoning of pre-service teachers who were soon to become teachers was investigated in the theoretical framework for elementary school students developed by Jones, Langrall, Mooney and Thornton (2004). This theoretical framework was described in detail below.

Theoretical Framework

Statistical Reasoning Levels

Jones et al. (2004) described statistical reasoning of elementary school students in four processes: (1) describing data, (2) organizing data, (3) representing data, and (4) analyzing and interpreting data. The process of *describing data* includes reading the data presented as raw or in the form of a graph, table or scheme. Additionally in this process, two sub-processes as showing awareness regarding presentations and determining the units of data values are included. The process of *organizing data* includes organizing, classifying and summarizing data. Additionally, there are three sub-processes regarding organizing data as grouping the relevant data, summarizing the data using central tendency measures, and describing the variation of the data. The process of *representing data* is related to description of the data in form of a graph. Sub-processes of representing data are establishment of a style of presentation for the given data set and evaluating the suitability of the selected style. The final process of *analyzing and interpreting data* includes determination of patterns and tendencies in the data and estimation and inference based on the data. In scope of this study, we approach the process of *analyzing and interpreting data* (Jones et al., 2004) which represents the essence of statistical reasoning.

The theoretical framework developed by Jones et al. (2004) consists of four hierarchically advancing levels as: *(i) idiosyncratic (Level 1), (ii) transitional (Level 2), (iii) quantitative (Level 3), and (iv) analytical (Level 4)*. These hierarchical levels defined regarding the process of analyzing and interpreting were defined in detail over the given sample problems. Students' reasoning on *Level 1* is limited and they usually focus on their own experiences and personal beliefs independent of the data. The student responses on this level may be described as the following in the context of a problem developed by Jones, Thornton, Langrall, Mooney, Perry and Putt (2000) (Sam's friends). Sam was visited by his friends during a week in summer holiday. Number of visiting friends and the days they visited were given in the form of a line plot. Elementary school students were asked to guess the number of friends Sam would be visited by during four weeks. A student on Level 1, while making a decision, stated "*four, because four friends visited me this summer*" based on his personal experience, without attempting to make a guess. Students on *Level 2* start to have reasoning in a quantitative way; however, they are inconsistent while using it. Additionally, a student on this level focuses on only one side of a problem situation. In a problem used by Jones et al. (2004) (Best Concert Tour problem), elementary school students were given three bar graphs including concert tour information of four music bands (Barbra Streisend, Boyz II Men, Eagles and Rolling Stones). The first of these graphs showed the total concert revenue for each band, these second showed total number of concerts, and the third showed the number of cities the concerts took place. Students were asked which band had the most successful concert tour. It was found that an elementary school student on Level 2 looked only at the height of the columns on the graphs to find the most successful concert tour. The reasoning of students on Level 3 is like the following: in problem situations, they can make multiple quantitative comparisons without approaching issues unrelated to the problem. However, while a student on this level is working on a problem, he does not gather the related mathematical ideas. Reasoning of students on this level may be explained by an example used by Mooney (2002). In this example, numbers of medals (gold, silver, bronze) won by five countries in the 1998 Winter Olympics were shown in bar graphs and tables. Secondary school students were asked to compare the numbers of medals for five countries. Students on this level, while comparing the medals won, calculated both gold medals and the total number of medals. However, they were not able to relate the two. It may be argued that students on Level 4 are able to establish the relationship among different aspects of a problem. Students on this level are able to make valid inferences by using the context the problem is presented in. With the Best Concert Tour problem, the students on this level are aware of the fact that a relationship must be established between the money the bands earn and the number of concerts. On the other hand, secondary school students are also able to make comparisons by reducing relevant ratios from the context.

Cobb and Moore (1997) suggested that statistical data do not consist only of numbers, but are meaningful only when they are presented in a context, and therefore statistics requires a different way of thinking. Understanding the context of the problem while interpreting data and making inferences plays an important role in statistical reasoning. Similarly, Ben-Zvi and Garfield (2004) emphasized the importance of working with real data for a good statistics education. The researchers also emphasized that using alternative assessment methods is needed to understand the learning of students better in a good statistics education. Garfield and Chance (2000) also provided some suggestions to assess the statistical reasoning of students. In these suggestions, it is advised that students should take part in case studies or authentic tasks presented based on a real context that will reveal their strategies and guesses in solution. In this study, the statistical reasoning processes of pre-service teachers was analyzed in the frame of a model eliciting activity that provided them with real data and a meaningful context. Further information is included below regarding mathematical modeling and model eliciting activities.

Mathematical Modeling and Model Eliciting Activities

Mathematical modeling, being used in various fields such as medicine, science, engineering, architecture and economics, is also an important part of mathematics education. Assuming that mathematical modeling supports meaningful learning and helps establish relationship with real life (Lesh & Doerr, 2003), the necessity of modeling in mathematics education may be more clearly understood. Any kind of effort to create information, skill or structure towards relating the world of mathematics and the real world is a part of mathematical modeling. Mathematical modeling is a non-linear process which contains elements belonging to the real world and the world of mathematics (Zbiek & Conner, 2006).

Lesh and Doerr (2003) defined mathematical modeling as a process which occurs during activities of model eliciting. The meaning of model eliciting activity mentioned here is, problem situations that are related to the real world which, as opposed to the traditional problems, encourage the students to establish a generalizable model and produce possible different and creative solutions, rather than creating a single correct answer. Students start with real life situations presented to them in model eliciting activities. These activities which require more time, thinking and concentration in comparison to routine problems, should be implemented in small groups so that the necessary discussions take place. Model eliciting activities require students to express their ways of thinking clearly. They also clearly define the conceptual strengths and weaknesses of the students (Lesh & Doerr, 2003). Following the group work by students during the model eliciting activity, the activity continues with presentations and in-class discussion. The purpose of these presentations and discussions is, expression of their ways of thinking, establishing awareness of students of each other's way of thinking, seeing strong and weak aspects of different approaches and creating strategies for improvement of the work of others or of their own. The teacher who follows this process as a guide gets an idea about the ways of thinking of all students, determines the strong and weak aspects of his/her students, and becomes able to assess his/her students' performances as if he had a one on one interview with them (Lesh & Doerr, 2003)

Problem Situation and the Purpose of the Study

The purpose of this study is to determine the statistical reasoning levels of pre-service teachers in the process of *analyzing and interpreting data* in the context of model eliciting activity, as well as how well they use concepts on these levels while making statistical decisions. The reason for conducting the study in the context of a model eliciting activity is that model-eliciting activity is seen as a suitable method to reveal the thoughts of pre-service teachers, as much as it provides a context and data that is based on the real life situations. For this purpose, the research question was determined as the following:

> What are the levels of statistical reasoning of pre-service elementary mathematics teachers (sophomore and junior students) in the context of a model eliciting activity, and how are they using statistical concepts on these levels while making statistical decisions?

Method

Research Design

A case study was conducted using qualitative research methods in this study. The situation aimed to be analyzed thoroughly as a necessity of the case study is; the levels of statistical reasoning in pre-service teachers studying

for their second (sophomore) and third years (junior) in elementary school mathematics teaching at a large state university in the province of Ankara, and how they use statistical concepts on these levels.

Participants

This study was conducted in the spring semester of the school year of 2015-2016 on second (sophomore) and third year (junior) students of the department of elementary school mathematics teaching at a state university in Ankara. Participants are pre-service teachers who were taking the course of *Mathematical Modeling in Elementary School,* in which the second author was instructor, as an elective course. The first author attended the class with the instructor of this course during the semester and observed the class. The course is an elective course in the undergraduate curriculum of the department of elementary school mathematics teaching. The distribution of pre-service teachers according to school year, sections and gender is given in Table 1 below.

Table 1. Distribution of pre-service teachers according to school year, sections and gender

	Year 2, Section 1	Year 2, Section 2	Year 3, Section 1	Total
Female	25	38	39	102 (83%)
Male	14	3	4	21 (17%)
Total	39	41	43	123 (100%)

Heterogeneous groups of four were established in the beginning of the semester in compliance with the modeling perspective. The researchers considered the GPAs and genders of the pre-service teachers while establishing the groups. A total of 29 groups were established including 9 groups from the first section of the second year, and 10 groups from each of the second section of the second year and the third year. Group numbers in the findings section were assigned and presented in the range of 1-19 for second year students and 1-10 for third year students. Comparing the statistical reasoning levels of pre-service elementary school teachers according to their school year is not included within the scope of this research.

As a requirement of the course, four model eliciting activities were studied with pre-service teachers for first four weeks of the semester. The participants tried to solve each modeling activity for 90 minutes by discussing with their group members. In the following 60 minutes, the groups discussed their solution methods under the supervision of the instructor. While the groups were working on an activity, both researchers walked around the classroom and tried to understand the way of thinking for each group by asking questions, without providing any leads. The groups were asked to talk about their own methods of thinking at the end of each model eliciting activity. In addition, the pre-service teachers were asked to describe their solutions (models) as a report in detail. At the end of these model eliciting activities, discussions took place for two weeks on modeling perspective and the principles that need to be followed while establishing model eliciting activities. In the last four weeks, as one of the requirements of the course, pre-service teachers were asked to develop model eliciting activities. In the scope of this study, we used the data obtained from one of the four model eliciting activities, the On-Time Arrival problem. The On-Time Arrival problem is described in detail below.

Model Eliciting Activity: On-Time Arrival Problem

The On-Time Arrival problem used to reveal the statistical reasoning of pre-service teachers was developed in the scope of Small Group Mathematical Modeling (SGMM) by Purdue University School of Engineering Education (for this model eliciting activity and other model eliciting activities, you may visit the website: https://engineering.purdue.edu/ENE/Research/SGMM/CASESTUDIESKIDSWEB/ontimearrival.htm). The On-Time Arrival problem was translated into Turkish by the researchers and revised. (The researchers only changed the names of the airlines as A, B, C, D and E without changing the delay times.) The reason for choosing the On-Time Arrival problem was that it is related to statistical concepts, as well as the idea that the thinking of pre-service teachers regarding statistical concepts may be evoked more easily while working on this problem. In the problem, the pre-service teachers were given information on the number of delay times by five airline companies in 30 flights. The objective of the problem is to find the airline with the lowest probability of delay. The thing expected of the pre-service teachers was for them to: assess the data set in detail, notice outliers in the data set and interpret them, decide on the central tendency and variation measures most suitable for this data set, and be able to calculate and interpret these values.

Data Collection Instrument

The data of the study consist of the reports and solution sheets written by the 29 groups at the end of the On-Time Arrival problem. The pre-service teachers were asked for their reports to describe the decided solutions (models) at the end of the activity and mention all concepts they used.

Data Analysis

Content analysis method was used in analyzing the data. The reports and solution sheets obtained from the groups were examined together, and each group's reasoning, the statistical concepts they used, and how they used those concepts were presented in tables. Later, the data were analyzed in the scope of the theoretical framework regarding statistical reasoning developed by Jones et al. (2004). Each group's reasoning was coded by both researchers using the determined codes (personal experience, looking only at flights that were not delayed, computing arithmetic mean only, computing central tendency measure, computing variation measure, having difficulty in making more than one comparisons and finding a relationship, determining the relationship between the context and the data, etc.). These determined codes were grouped based on their similarities and gathered under descriptors which represent statistical reasoning levels defined by Jones et al. (2004). Therefore, the levels of statistical reasoning of group were determined based on this coding. Descriptive analyses (calculating frequency for each level) were conducted for the determined results. Additionally, how level characteristics were presented in the problem context and how statistical concepts were interpreted during reasoning, were also examined.

Validity and Reliability

External validity in qualitative studies is related to how much research findings may be transferred to other situations (Fraenkel, Wallen & Hyun, 2012). In order to increase external validity in this study, the properties of the sample and the study environment were defined in the method section in detail. For reliability, the data were coded by the two researchers, and the disputed points (such as statistical reasoning levels of groups) were decided on by discussion. Data analysis process was also described in detail. Therefore, the reader was given the opportunity to follow the claims and interpretation presented by the researchers based on the data.

Findings

It was aimed in this section to analyze the statistical reasoning of pre-service teachers who were studying elementary school mathematics teaching in the scope of *analyzing and interpreting data* within the context of the given model eliciting activity. The findings obtained for this reason were analyzed under four levels determined according to the theoretical framework developed by Jones et al. (2004). Table 2 shows the distribution of statistical reasoning of the pre-service teachers based on the levels.

Table 2: Statistical reasoning levels of pre-service teachers regarding the process of analyzing and interpreting data

Levels School Year	Level 1	Level 2	Level 3	Level 4
Year 2	0	14	5	0
Year 3	2	6	2	0
Total	2	20	7	0

As it may be seen in the table above, only the first three of the levels determined in accordance with this framework were seen in the pre-service teachers. Level 4, which is the highest among these hierarchically-ordered levels, was not seen in the sample of implementation in the context of the given problem. The vast majority of the pre-service teachers (20 groups) were on Level 2. While the statistical reasoning of 7 groups was determined as Level 3, 2 of the groups were found to be on Level 1. There were no groups observed to be on Level 4. The statistical reasoning determined in the context of this modeling activity for each level are presented quoting the responses by the pre-service teachers. Additionally, short descriptions of all levels are provided in Table 3.

Level 1 – Idiosyncratic: According to Jones et al. (2004), the students on this level are not able to appropriately analyze data, whereas they support their responses based on examples from their own experiences. For instance, a group making the statement "We think Airlines C is ideal, because a vast majority of its delay times are between 0 and 19 minutes. A 125-minute delay was encountered only once. We recommend this airline for you, as we think this value was due to a technical problem" (3rd year, Group 6) and a group making the statement "we thought 0-10 minutes of delay for an airplane may be ignored while even a bus departs with an average of 5 minutes of delay" (3rd year, Group 9) were evaluated to be on this level. When we examined these groups' solutions it can be said that they interpret the meaning of statistical concepts (for instance recognizing 125 minutes delay as an outliers) intuitively. As it may be seen from the responses of the groups, each group tried to reach a conclusion relying solely on their personal experiences.

Level 2 – Transitional: According to Jones et al. (2004), students on this level start to use quantitative information in statistical problems, therefore starting to show quantitative reasoning; however, they do this in a limited, mistaken way. For example, a group (2nd year, Group 3) on this level firstly calculated the arithmetic mean of delay time for each company. They felt the need to calculate standard deviation because the values of arithmetic mean for each company closest to each other. They chose the company with a standard deviation closest to the average while deciding. As a reason, they stated that "*the closeness of standard deviation to the* arithmetic *mean shows that the company is consistent,*" mistakenly. However, closeness of standard deviation to the arithmetic mean provides us with no information statistically.

When the incorrect solution method of another group on this level was analyzed, it was seen that they did not consider the situations without delays (shown in tables as 0) while calculating the arithmetic mean of delay time for each company. All data set should be considered while computing arithmetic mean. This group removed only the "no delay" times (0 minutes) from the data set while calculating the arithmetic mean of delay time for each company and defined this value as "arithmetic mean without zero values" (2nd year, Group 9). It is a noteworthy fact that the group did not remove the highest lengths of delay times (such as 123) from data set while calculating the arithmetic mean. This shows that pre-service teachers cannot conceptualize the arithmetic mean.

Table 3. Examples regarding the levels of statistical reasoning in the context of the On-Time Arrival problems regarding the process of analyzing and interpreting data

	Level 1 Idiosyncratic	Level 2 Transitional	Level 3 Quantitative	Level 4 Analytical
Process of analyzing and interpreting data	Supporting response by providing examples from own experiences • Deciding on the importance of short or long delay times based on own experience.	Focusing on only one aspect of the data and limited reasoning • Deciding based on arithmetic mean or no delays only Thinking completely wrong in terms of statistics (thinking that standard deviation value should be close to the arithmetic mean)	Making more than one comparison, difficulty to establish relationship Without thinking about the problem context, computing measures such as mode, median, arithmetic mean, standard deviation and interquartile range, but being unable to establish relationships among them.	Establishing relationship between the context and the data • Deciding on the most suitable measure to interpret the data on the table and being able to interpret the meaning of this measure in the problem's context. Being able to use the data and the context together in order to make consistent and clear predictions.

Another group considered on these level (2nd year, Group 18) made a mistaken comparison by horizontally analyzing each company's delay times in itself after calculating the arithmetic mean of delay times for each company (see. Figure 1). This group determined the airline that delayed the least in each flight, and values below the arithmetic mean of delay time calculated separately for each flight, and showed them with symbols "*" and "√" in the table. For example, the delay times of the airlines in their 14th flights were like the following: Airlines A 5 minutes, Airlines B 10 minutes, Airlines C 9 minutes, Airlines D 55 minutes, Airlines E 10 minutes. The arithmetic mean of delay times was calculated to be 17.8 minutes for this flight. For values below this, the "*" symbol was used to mark Airlines A, Airlines B, Airlines C and Airlines E. Airlines D was not marked at all as it delayed 55 minutes. Additionally, Airlines A was marked with "√" because it delayed the least in the 14th flight. Then the number of markings for these two values was calculated for each company. The company with the highest amount of marks was selected.

Analyzing each flight separately does not provide a meaningful result in the context of this problem, because the following way of thinking by this group is an indicator of their mistaken reasoning: In the 28th flight, the delay times were as the following: Airlines A 123 minutes, Airlines B 10 minutes, Airlines C 5 minutes, Airlines D 75 minutes, Airlines E 0 minutes. The arithmetic mean of delay time was found as 42.6 minutes for this flight. While there was a big difference between the delay amounts of Airlines A and Airlines D above this level, both firms were thought as the same and removed from the data set. However, flights are independent of each other. Pre-service teachers are expected to assess 30 flights on the basis of airlines, instead by flights. A typical value representing each of the airline companies should be determined in order to assess the performance of airlines. Another aspect of this level is that they start to focus on the relevant characteristics of the problem situation, but this focus is limited to considering only one factor.

* = airlines below the average delay time in each flight
√ = the airline with the least time of delay in each flight
x = the least time of delay between airlines A and airlines D

Figure 1. Solution method of 2nd year, Group 18

For example, a group on this level (2nd year, Group 5) calculated the total delay time of each airline. However, after realizing that total delay times were very close to each other, they determined the number of flights without delays for each airline. They considered the number of flights without delay while deciding on their answer. 5 of the 20 groups found to be on Level 2 chose Airlines D using this method. Figure 2 shows the way of solution by one of the groups using this method.

Airlines A; were not delayed in 15 out of 30 flights
Airlines B; were not delayed in 6 out of 30 flights
Airlines C; were not delayed in 4 out of 30 flights
Airlines D; were not delayed in 17 out of 30 flights
Airlines E; were not delayed in 10 out of 30 flights

Figure 2. Way of solution by 2nd year, Group 5

Level 3 – Quantitative: According to Jones et al. (2004), students on this level, as opposed to those on Level 2, use quantitative reasoning more effectively in a consistent way. However, while they are able to make multiple comparisons, they have difficulty establishing connections among them. For example, pre-service teachers on this level generally tried to find a typical value to assess this data set. Pre-service teachers on this level (7 groups) calculated central tendency and variation measures such as mode, median, arithmetic average, standard deviation, interquartile range for each company. However, they found it difficult to determine a relationship among these data. The way of solution by 3rd year Group 4 given in Figure 3 may be an example. This group firstly determined the airline with the lowest standard deviation (Airlines B). Then they determined the best company as Airlines A in terms of interquartile range and mode. However, they chose the company with the lowest standard deviation while making a decision (Airlines B).
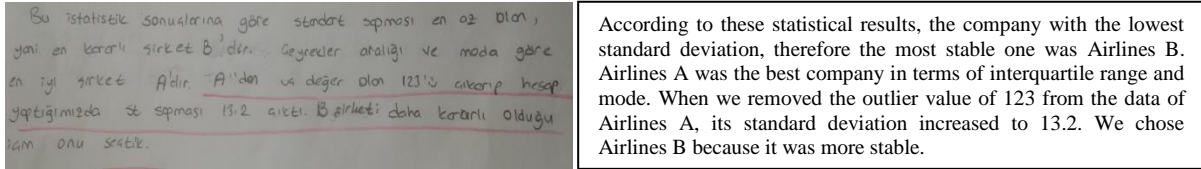
According to these statistical results, the company with the lowest standard deviation, therefore the most stable one was Airlines B. Airlines A was the best company in terms of interquartile range and mode. When we removed the outlier value of 123 from the data of Airlines A, its standard deviation increased to 13.2. We chose Airlines B because it was more stable.

Figure 3. Way of solution by 3rd year, Group 4

All groups used standard deviation at the end while they made their decisions. They chose the company "Airlines B", which had the lowest standard deviation. They stated that low standard deviation was an indicator of low deviation from the arithmetic mean and low risk. However, as this data set contains outliers, standard deviation is not a suitable value that represents it.

Level 4 – Analytical: According to Jones et al. (2004), students on this level decide on the best measure to interpret the data, and are able to interpret the meaning of this measure in the context of the problem. This level, as opposed to Level 3, uses inference based on data, and requires the combined usage of both quantitative and contextual knowledge. Pre-service teachers on this level are expected to assess the centers and variations of data sets together. In the case of the closeness of data set centers, an analysis should also be made on the variation of the data. No groups were on this level in the scope of this study.

Discussion

The purpose of this study is to investigate the statistical reasoning processes and levels of pre-service elementary school teachers in the theoretical framework developed by Jones et al. (2004) for elementary school children. These reasoning processes and levels determined for elementary school students provided similar findings for pre-service teachers. This framework, which explains statistical reasoning with hierarchical levels, is an important resource as it provides information on statistical reasoning of pre-service teachers.

The findings show that a vast majority of pre-service elementary school teacher had very low levels of statistical reasoning in relevance to the process of *analyzing and interpreting data.* Generally low levels statistical reasoning of pre-service teachers was also reported in other studies' findings. Although studies on statistical reasoning usually focus on students understanding statistical ideas, it may be argued that studies towards teachers and pre-service teachers also increased in recent years (Mickelson & Heaton, 2004; Gürel, 2016). Reasoning of teachers on statistical concepts and their knowledge of teaching were also investigated in these studies, and the findings were parallel to those found in this study.According to Jones et al. (2004), the process of *analyzing and interpreting data* is in the heart of statistical reasoning. This process requires awareness of tendency and patterns in the data and ability to make predictions and inference from the data. It is a subject that is worth consideration that a vast majority of pre-service teachers were on Level 2 in terms of the process of *analyzing and interpreting data*. The typical characteristic of this level is the mistaken or limited reasoning when asked to interpret a data set on a table. The pre-service teachers on this level, firstly, did not take the data set as a whole and ignored the distribution of the data. They did not discuss how outliers in the data set may affect the arithmetic mean. Pre-service teachers are expected to firstly examine the distribution of the data set and notice outliers. An outlier for a data set is a part that does not suit the other items on a data set. Whether there are outliers while evaluating a data set carries importance while deciding on the measure that represents the data set in the best way (Chamberlin, 2013). For example, in the On-Time Arrival problem, it is expected to notice whether the 123 minutes of delay in the 28th flight of firm A is an outlier or not, and discuss it in the group formally or informally.

When statistical reasoning of the pre-service teachers on this level was examined, it was seen that they made statistical mistakes in determining a suitable value to represent a data set. The pre-service teachers, while calculating arithmetic mean, removed some values in the data set (only the lowest ones among outliers) arbitrarily. This finding shows that the pre-service teachers have serious limitations in the conceptual meaning of arithmetic mean. This limitation in the conceptual meaning of arithmetic mean was also shown in other studies (Leavy & O'loughlin, 2006; Toluk-Uçar & Akdoğan, 2009). Leavy and O'loughlin (2006), in their study with pre-service teachers, found that pre-service teachers had the necessary procedural knowledge in relation to calculating the arithmetic mean, they knew the formulas and rules; however, they were not able to make the connection between procedural and conceptual knowledge. Likewise, in Toluk-Uçar and Akdoğan's study on secondary school students, it was found that the vast majority of the students did not understand the power of the arithmetic mean to represent the data.

It is noticeable that a vast majority of the pre-service teachers calculated firstly the arithmetic mean while interpreting a data set. The general tendency in the pre-service teachers on this level is to calculate the arithmetic mean algorithm right away without looking at the data as a complete structure when encountering a data set. However, the data in this model eliciting activity contain outliers. When there are outliers, the arithmetic mean cannot be used as a value to represent the data set, as the arithmetic mean is heavily affected by outliers. In such situations, calculating the median, which does not easily get affected by outliers as a measure of central tendency, instead of arithmetic mean, represents the data set better (Gravetter & Wallnau, 2004). This result in this study is parallel to Groth's (2006) finding that students have a belief that the only and most suitable central tendency measure for a data set is the arithmetic mean. Groth (2006) in their study asked high school students to estimate the typical income in a company with personnel salaries on different levels. They reported that the high school students have a belief that the typical income may be determined only by calculating the arithmetic mean despite the data containing outliers. However, the students are expected to look at the median value of the personnel salaries instead of the arithmetic mean while finding the typical income, and be able to explain why the median is more usable and why it does not get affected by outliers. Groth emphasizes that teachers should not send the message that the arithmetic mean is the only and most suitable central tendency measure while teaching central tendency measures.

Another important finding in the study was about the central tendency and variation measured. As opposed to those on Level 2, the pre-service teachers on Level 3 thought that a value other than arithmetic mean should be calculated, and chose to find a typical value in order to assess the data set. However, they separately calculated central tendency measures such as mode and median, and variation measures such as standard deviation, range and interquartile range, without considering the data within the context of the problem. While calculating standard deviation, which is a variation measure, in the context of the On-Time Arrival problem provides us with information on whether the delay times for each airline are homogenous or heterogeneous, it does not show us whether they are high or low. For example, when the standard deviation of delay times is calculated for each company this value may be low. This shows that delay times in different flights are close to each other, meaning there is low variation. However, this does not provide any information on whether the delay times are high or low. In other words, calculating standard deviation, which is one of the measures of variation, does not provide a meaningful result in this problem's context, because standard deviation and variance get affected by outliers as they are calculated using squares of deviation amounts (Gravetter & Wallnau, 2004).

No pre-service teacher in the sample showed characteristics of Level 4. This result is parallel to the findings of the study by Mooney (2002) on secondary school students. Mooney reported that no student was found on the fourth level in the study. In the fourth level of statistical reasoning regarding the process *of analyzing and interpreting data*, it is expected that a relationship is established between central tendency and variation measures in the context. No pre-service teachers in the sample provided any explanations about which situations variation measures are used for. Çakmak and Durmuş (2015) obtained similar results in their study on secondary school students. They found that most of the students experienced difficulties in interpreting data and choosing one of the central tendency or variation measures suitable for the purpose. They also reported that the students, who did not experience difficulty in calculating the interquartile range, had no knowledge about interpreting this concept. As there are outliers (very high and very low) in the model eliciting activity in this study, the pre-service teachers are expected to calculate the interquartile range, which is one of the variation measures suitable for this data set. Interquartile range measures the distance of the middle point of the distribution to the boundaries of the 50% in the middle. Therefore, interquartile range and interquartile mean are not affected by outliers. Thus, it is a more stable measure than range (Gravetter & Wallnau, 2004).

This very low level of statistical reasoning of pre-service teachers who will start teaching statistics soon, poses a serious limitation to create an effective learning environment when they start their jobs. When the curriculum of middle school mathematics for 5-8th grades seen in Ministry of Education document is examined, there are many learning outcomes under the name of data processing regarding calculation and interpretation of central tendency measures, as well as understanding which of these values are more useful for a specific data set (MoNE, 2013). In order for pre-service teachers to create an effective learning environment, firstly they are expected have procedural and conceptual knowledge regarding these concepts. In the scope of the Statistics and Probability I-II courses and Special Instruction Methods classes they are offered during the undergraduate degree of elementary school mathematics teaching in Turkey under the framework of the Higher Education Council, it is aimed for pre-service teachers to acquire subjects such as basic concepts regarding statistics, measures of central tendency and applications (CoHE, 2006). However, low levels of statistical reasoning in pre-service teachers may be due to the formula-based approach to statistics in these courses. In their study with secondary school students, Çakmak and Durmuş (2015) determined the parts where the students experienced difficulties in statistics and probability concepts, and they reported the approach on statistics and probability, as

well as their nature as a reason. Therefore, based on the results of this study, it may be suggested that pre-service teachers should work on problems that include real data, containing risk situations and outliers problems during their undergraduate degree.

Statistical reasoning processes of pre-service teachers regarding *analyzing and interpreting data* were investigated in the context of a single model eliciting activity. Studies may be conducted on whether or not the statistical reasoning levels of pre-service teachers will change in the context of different modeling activities. It is considered that model eliciting activities that may be implemented in connection with statistical concepts are useful not only in revealing the statistical reasoning of pre-service teachers, but also improving it. Lesh, Amit and Schorr (1997), reported in their study that structuring statistical concepts and developing statistical reasoning processes are difficult or impossible with traditional methods, and emphasized the importance of model eliciting activities in statistics education. While pre-service teachers are working on real data, they improve their statistical reasoning and thinking by finding the chance to discuss different ways of solution for the presented statistical problems (Garfield & Gal, 1999). It has been emphasized in numerous studies that such rich activities are needed to conceptually understand statistics concepts such as central tendency and variation measures (Ben-zvi & Garfield, 2004; Garfield & Gal, 1999; Garfield, delMas & Zieffler, 2010; Leavy & O'loughlin, 2006).

References

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), The challenge of developing statistical literacy, reasoning and thinking (pp. 3-17). Dordrecht: Kluwer Academic Publishers.

Chamberlin, S. A. (2013). Statistics for kids: Model-eliciting activities to investigate concepts in statistics. S. Redmond (Ed.). Waco, Texas: Prufrock Press Inc.

Chervany, N. L., Collier, R. O. Jr., Fienberg, S. E., Johnson, P. E., & Neter, J. (1977). A framework for the development of measurement instruments for evaluating the introductory statistics course. American Statistician, 31 (1), 17-23.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics and teaching. The American Mathematical Monthly, 104 (9), 801-823.

Çakmak, Z. T., & Durmuş, S. (2015). İlköğretim 6-8. sınıf öğrencilerinin istatistik ve olasılık öğrenme alanında zorlandıkları kavram ve konuların belirlenmesi. [Determining the concepts and subjects in the area of learning statistics and probability that 6-8th grade math students have difficulties.]Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 15 (2), 27-58.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to design and evaluate research in education (8th ed.). New York: McGraw-Hill.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges, Mathematics Thinking and Learning, 2 (1&2), 99-125.

Garfield, J., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L. V. Stiff (Ed.), Developing mathematical reasoning in grades K-12 (NCTM 1999 Yearbook) (pp. 207-219). Reston, VA: National Council of Teachers of Mathematics.

Garfield, J., delMas, R., & Zieffler, A. (2010). Developing tertiary-level students' statistical thnking through the use of model-eliciting activities. In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

Gravetter, F. J., & Wallnau, L. B. (2004). Statistics for the behavioral sciences. Belmont, CA: Thomson.

Groth, R. E. (2006). An exploration of students' statistical thinking. Teaching Statistics, 28 (1), 17-21.

Gürel, R. (2016). Ortaokul matematik öğretmenlerinin merkezi eğilim ve yayılım ölçülerine ilişkin öğretim bilgilerinin incelenmesi. [An investigation of middle school mathematics teachers' knowledge for teaching measures of central and tendency and variation] (Unpublished doctoral dissertation). Hacettepe University, Ankara.

Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2004). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), The challenge of developing statistical literacy, reasoning and thinking (pp. 97-117). Dordrecht: Kluwer Academic Publishers.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. A. (2000). A framework for characterizing children's statistical thinking. Mathematical Thinking and Learning, 2 (4), 269-307. DOI: 10.1207/S15327833MTL0204_3

Leavy, A., & O'Loughlin, N. (2006). Preservice teachers understanding of the mean: Moving beyond the arithmetic average. Journal of Mathematics Teacher Education, 9, 53–90.

Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning, and problem solving. In R. Lesh & H. M. Doerr (Eds.), Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching (pp. 3-33). Mahwah, NJ: Lawrence Erlbaum.

Lesh, R., Amit, M., & Schorr, R. Y. (1997) Using "real-life" problems to prompt students to construct conceptual models for statistical. In I. Gal & J. B. Garfield (Eds.), The Assessment Challenge in Statistics (pp. 65-83). Netherlands: IOS Press.

Mickelson, W., & Heaton, R. (2004). Primary teachers' statistical reasoning about data. In D. Ben-Zvi & J. Garfield (Eds.), The challenge of developing statistical literacy, reasoning and thinking (pp. 327-353). Dordrecht: Kluwer Academic Publishers.

Ministry of National Education (MoNE) (2013). Ortaokul Matematik Dersi (5.,6.,7.,8. Sınıflar) Öğretim Programı [Middle School Mathematics (5, 6, 7 and 8th Grades) – Teaching Curriculum.]. Ankara: MEB Yayınları.

Mooney, E. S. (2002). A framework for characterizing middle school studens' statistical thinking. Mathematical Thinking and Learning, 4 (1), 23-63. DOI: 10.1207/S15327833MTL0401_2

Purdue University College of Engineering. Retrieved August 28, 2016 from https://engineering.purdue.edu/ENE/Research/SGMM/CASESTUDIESKIDSWEB/ontimearrival.htm

Toluk-Uçar, Z., & Akdoğan, E. N. (2009). 6.-8. sınıf öğrencilerinin ortalama kavramına yüklediği anlamlar. [Middle School Students' Understanding of Average] İlköğretim Online, 8 (2), 391-400.

Council of Higher Education (CoHE). (2006). Eğitim Fakültesi Öğretmen Yetiştirme Lisans Programı: İlköğretim Matematik Öğretmenliği Lisans Programı Ders İçerikleri. [Faculty of Education, the Undergraduate Program in Teacher Training: Curricula of the Undergraduate Program in Elementary School Mathematics Teaching.] Retrieved August 28, 2016 from

http://www.yok.gov.tr/documents/10279/49665/ilkogretim_matematik/cca48fad-63d7-4b70-898c-dd2eb7afbaf5

Zbiek, R. M., & Conner, A. (2006). Beyond Motivation: Exploring mathematical modeling as a context for deepening students' understandings of curricular mathematics. Educational Studies in Mathematics, 63, 89-112. DOI: 10.1007/s10649-005-9002-4.

Author Information

Çiğdem Alkaş Ulusoy	**Mesture Kayhan Altay**
Hacettepe University	Hacettepe University
Ankara, TURKEY	Ankara, TURKEY
Contact e-mail: cigdemalkas@*hacettepe.edu.tr*	