MIDDLE SCHOOL STUDENTS' STATISTICAL REASONING ABOUT DISTRIBUTION IN THEIR STATISTICAL MODELING PROCESSES

TUĞÇE BALKAYA
Ministry of National Education
tb.balkaya@gmail.com

GAMZE KURT
Mersin University
gamzekurt@mersin.edu.tr

ABSTRACT

Statistical reasoning about a population through samples can be achieved by modeling the relationship between population and sample. One way to do this is to model real data situations in a technology-integrated environment. With this view, we aimed to investigate how middle school students formed distributions and examined their statistical modeling processes through the informal reasoning process within the Reasoning with Informal Statistical Models and Modeling (RISM) framework. The case study reported in this paper focuses on how the conjecture and data models students designed throughout three activities evolved and how their inclusion of a fundamentally probabilistic mechanism matured. Findings show the students approached the distribution probabilistically with their inferences during the modeling process and grasped the statistical concepts they would encounter at a more advanced level. Therefore, we claim that students shifted from understanding empirical distributions to understanding theoretical distributions.

Keywords: *Statistics education research, Distribution; Informal statistical inference; Statistical modeling; Statistical reasoning; TinkerPlots*

1. INTRODUCTION

Distributions are conceptual, organizational structures or mental tools that facilitate the development of statistical thinking and reasoning. These structures are complex, and their understanding requires development and cultivation. Thus, many questions arise about the conceptual, pedagogical, and research-related aspects of reasoning about distributions (Pfannkuch & Reading, 2006; Rumsey, 2002). Statistics teaching focuses on building and defining distributions particularly with an increasing emphasis on making and testing data-based assumptions, and the introduction of the term "distribution" (National Council of Teachers of Mathematics, 2000). The interpretation of distributions and their editing and manipulation to provide more information from the data are often not mentioned in statistics classes (Lehrer & Schauble, 2004). Students, however, can explore relevant contexts if they understand different distribution structures built on the same data (Bakker, 2004a). First, students must understand what the various graphical data representations and data distributions mean to interpret and evaluate the data (Bakker & Derry, 2011; Bakker & Gravemeijer, 2004).

Interpreting data and their distribution is crucial and a contemporary global necessity. Studies have shown that students and teachers/prospective teachers do not understand or focus on distribution to make statistical inferences about data (Canada, 2004; Hammerman & Rubin, 2004; Makar & Confrey, 2002, 2005). Even when students have been taught the fundamentals of statistics (measures of central tendency and variability), they seldom understand its relevance to data distributions and variation (Watson, 2009). In elementary school, the topic of data handling is heavily focused on calculations of statistics, such as the mean or median, despite the fact students will need to advance towards an understanding of theoretical distributions in later school years.

In statistics teaching, students will meet two kinds of distributions. The first are empirical (or data) distributions, for which students learn to generate, define, and interpret graphs using empirical distributions. The second are theoretical (or probability) distributions, such as the normal or binomial distributions. While empirical distributions may be explained by defining and interpreting variability, theoretical distributions are models that explain and predict data variability (Wild, 2006). While both are similar in terms of characteristics like shape, students must explain the distinction between a theoretical and empirical distribution using the idea of variability (Garfield & Ben-Zvi, 2008). They should also focus on variability in the context of sampling (Garfield et al., 2008; Shaughnessy, 2019).

Wild (2006) used distribution as a *lens* to see variability. In statistics teaching, sample, distribution, and probability distribution types must be employed to facilitate reasoning about distributions (Noll & Shaughnessy, 2012). The distribution of the sample may serve as a bridge between empirical and theoretical distributions in statistics learning, since it provides students with a sense of the variability between samples when they go from considering empirical (i.e., data) distributions to theoretical (i.e., sampling) distributions (Garfield & Ben-Zvi, 2008). As a result, it is critical to examine the type of distribution students choose to represent their data to comprehend how they reason about the distribution. The research presented here provided students with statistical modeling exercises and instructions to create, review, compare, and conjecture about the data models utilised. The aim was to investigate their statistical reasoning to see how they formed the notion of distribution.

Statistics deals with data as aggregate and looks for patterns in data distributions to make inferences (Ben-Zvi & Arcavi, 2001). As a result, it is critical to distinguish between data and data distributions. Numerous studies have demonstrated that students view data as individual values rather than aggregates (Konold et al., 2007; Ben-Zvi & Arcavi, 2001) and that the use of technological tools can facilitate students thinking about data (Chance et al., 2007). Biehler et al. (2012) claimed students can understand statistical concepts more quickly if computation or graphing is reduced via the use of software. Additionally, Konold et al. (2007) argued that the use of the divider tool in *TinkerPlots*™ (https://www.TinkerPlots.com), which divides the sample into as many regions as desired, represented by boxes on the screen with case numbers written in the corner of each box, assists students in exploring the patterns of distributions. Additionally, the researchers asserted that data and chance can be linked to simulations created with *TinkerPlots*™ sampler tool, which is used to develop and run probability simulations, thereby creating data factories (Konold et al., 2007). Simulations can assist students in comprehending the probabilistic nature of data and act as mediators between data and chance. Given the critical role of context in statistics, connecting data, chance, and context with statistical modeling enables a meaningful connection (Pfannkuch et al., 2018).

While statistics is the science of inferring meaning from data, it encompasses data, data analysis, and statistical inference (Moore, 1997), and statistics instruction is increasingly contextualized using authentic activities that are meaningful to students (Wild et al., 2011). For this study, it is asserted that statistical inference enables an assessment of a more general cognitive development toward the use and comprehension of distributions (Pfannkuch, 2006; Reading & Reid, 2006).

If a model is created for a statistical purpose, it should have two distinctive features: the first is that the phenomenon of interest includes variability, and the second is that the existence of probabilistic predictions should use variability (Brown & Kass, 2009). In the modeling process, students consider variability and uncertainty (Pfannkuch et al., 2018). If the modelers are young students, probabilistic thinking brings informal statistical reasoning. Since probability feeds uncertainty and is the state of measuring uncertainty, informal reasoning is indispensable (Dvir & Ben-Zvi, 2018). *TinkerPlots*™ is accepted as a powerful tool for statistical modeling since it provides learners with both data analysis and sampling in statistical problems as they model data to produce explanations (Pfannkuch et al., 2018).

In the study reported in this paper, we aimed to examine how students learn and make inferences during the process of generating representations of distributions of quantitative data considered in the modeling process, which they develop by expanding their assumptions in terms of their real-life inquiries. We examined how students perform the reasoning process by modeling the data simulated through *TinkerPlots*™. We also analysed students' comments about the sample according to their data distributions. While using *TinkerPlots*™ in this study, we investigated the following research questions:

"How do middle school students make a comparison of conjecture and data models in a real-world setting?" and "How can we explain students' distributions via statistical modeling with a lens of informal inferential reasoning?"

2. THEORETICAL FRAMEWORK

Exploratory Data Analysis (EDA) refers to analyzing data in exploratory ways. Some recent research on EDA has focused on students' informal inferential reasoning processes between the real and probabilistic worlds (Manor & Ben-Zvi, 2017). A probability approach, stressing how statisticians employ modeling activities to solve problems, lacks some true data exploration elements, such as the formulation of a research question (Manor & Ben-Zvi, 2017). The integrated modeling approach (IMA) combines these two approaches with modeling-based guidelines. It is an approach that perceives statistical inference as a spiraling process between probability and the real world, with the goal of students understanding the link between sample and population (Aridor & Ben-Zvi, 2017). IMA was developed to guide the design and analysis of experimental tasks, deepen students' modeling while making informal statistical inference (ISI), and guide the assessment of that reasoning (Braham & Ben-Zvi, 2017).

In contrast, informal inferential reasoning (IIR) is defined as generalizing about the population from random samples using informal statistical tools (Makar et al., 2011). Makar et al. further argued that the statistical context chosen to prepare the authentic environment and determine the authentic aspects in statistical research would provide a common language for researchers to better express statistical reasoning. Students are expected to present both formal and informal representations about the concept being investigated simultaneously in the statistical inference process. The fact that this process is carried out both simultaneously and in the form of explaining the models makes it difficult to determine what students have or have not learned. Based on these definitions, the comparison model developed between the participants' conjectures and data models was considered by the research as a form of IIR. The statistical modeling process (Dvir & Ben-Zvi, 2018), presented as sub-modeling process, also forms the basis of the model comparison framework. In addition, the framework does not distinguish between the data and conjecture models, instead suggesting the modeler used similar considerations when exploring both the data and conjecture models.

One of the theoretical frameworks in which students are essential as modelers is Reasoning with Informal Statistical Models and Modeling (RISM; Figure 1). This framework, created by Dvir and Ben-Zvi (2018, p. 1185), advocates statistical modeling as an alternative modeling method, including statistical reasoning. It is particularly useful to divide the informal modeling process into three separate, though not independent, modeling processes: the process the data model goes through, the process the conjecture model goes through, and the process used in comparison through the comparison model. Consequently, simultaneous evaluation of many aspects of the context will enable students to explain the complicated and challenging nature of data handling (Arnold & Pfannkuch, 2014).

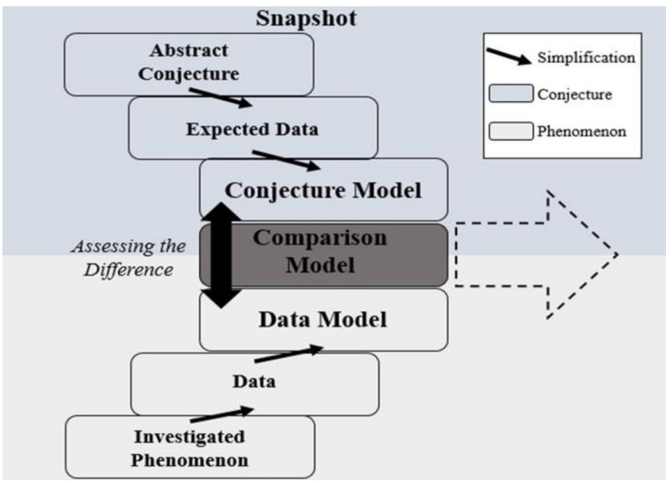


Figure 1. The RISM snapshot of an informal statistical modeling process (Dvir & Ben-Zvi, 2018).

In the RISM framework, the "conjecture model" is used. Whether the conjecture under consideration is a statistical model is important for this framework. The compatibility of the conjecture model with the probabilistic mechanism supports the belief that students reason with statistical models. The modeler transforms the obtained conjecture model into a data model. This process can be defined as the process of transition to data suitable to the investigated phenomenon. The modeler translates the patterns observed in the data into a data model. While students focus on creating models for both assumptions and data, an informal model simplifies the process of evaluation: students use the comparison model as a tool to evaluate the consistency between the data they create and the assumption model. We see this dual purpose as two planes where the statistical modeling process is carried out. In a plane, the phenomenon under investigation (for example, the relationship between height and weight) is examined and simplified into a model; in another plane, a more general assumption (e.g., a linear relationship) is symmetrically simplified. The comparison model acts as a catalyst in the frame. It adds movement to the models that were applied. Each snapshot describes the current version of all three models and recognizes a difference in one (or more) elicits the construction of a new snapshot. The result is then a discrete set of snapshots, each containing a multitude of organized information (Dvir & Ben-Zvi, 2018). Dvir and Ben-Zvi (2023) also explained the dual nature of conjecturing, in which one can lead contradictory/biased or true predictions, shows the potential of pedagogical benefits of teaching the shift between real-world-modeling to probability-world modeling.

Students can use any statistical model they create in RISM to run probabilistic queries on multiple samples. This not only builds students' confidence in their models, but also allows them to test their assumptions probabilistically. In Figure 2, Dvir and Ben-Zvi (2019, p. 925) show the median model's transformation from a data model to a conjecture model using symbols.

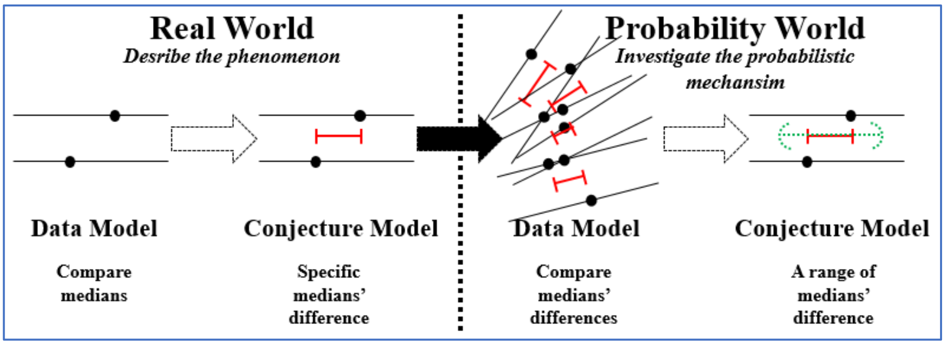


Figure 2. Illustration of the median model (Dvir & Ben-Zvi, 2019).

In the study conducted by Dvir and Ben-Zvi (2019) to examine the RISM process, they considered students' modeling from an in-depth perspective. Researchers were able to inform the student about the process by instantly detailing the representation in Figure 2 through the interview with the participants. For this purpose and when Figure 2 is examined, it is explained whether the students need any informal statistical inference in order to transform the data models they deal with during the activity into conjecture models; and, in cases where students want to generalize the situations they deal with and cover all the inferences they make, they change the model they have created or re-create them and switch to the probability world (Dvir & Ben-Zvi, 2019).

3 METHODOLOGY AND DATA COLLECTION PROCEDURE

This qualitative case study focuses on a specific topic, event, or situation, and in-depth information is obtained (Creswell & Poth, 2016). As Patton (2014) interpreted, "case study involves organizing data for specific cases for in-depth study and comparison," and deductive analysis examines "data to illuminate pre-determined sensitive concepts" (p. 534). The teaching setting here is the case in this case study.

3.1. STUDY CONTEXT

The study was conducted in the eastern region of Turkey at a school with a computer lab. The first researcher went to the school and informed the seventh-grade teachers about the study and the use of *TinkerPlots*™. Teachers recommended several students as volunteers. The study group consisted of 22 students who volunteered and whose parents provided permission to participate. Only eight students attended on a regular basis; the others came occasionally. Therefore, we analyzed the data of eight 11- to 13-year-old students: Aylin, Fatih, Hakan, Selim, Utku, Halit, Yaman, and Reyhan (all pseudonyms). With the exception of Reyhan, the students were in the seventh grade. Before instruction, we had a few lab sessions to practice *TinkerPlots*™. The research team worked hard to get the students to talk because they were initially shy. The students were required to attend the lab sessions. The students were eager to learn *TinkerPlots*™ and complete the worksheets. The study group met once a week for two hours after classes, for a total of six hours. A group of students sat side by side in front of a computer and discussed their responses. Seating plans were made, and the same groups worked together throughout the lab.

The students were all taught statistics in mathematics courses. In their primary school years between 1st and 4th grades, students learned about frequency tables, tally tables, pictograms, bar charts, and reasoning on bar charts. In the 5th grade (beginning of middle school), they learned about creating research questions, collecting data, organizing data in a bar chart or frequency table, and double bar graphs. In the 6th grade, they learned about data analysis, such as minimum and maximum values, mean and range. Students learnt line graph, pie chart, mode, and median in 7th and 8th grades. The curriculum excludes histogram and boxplot. The only probability concepts taught in first eight grades were probability of a basic event. Despite some shortcomings in this curriculum, there is evidence that students could learn the concepts intuitively through activities designed to their level (Kazak et al., 2014).

3.2. ACTIVITIES AND DATA COLLECTION

This study is part of a larger study investigating students' modeling processes using statistical concepts over six weeks. It summarizes the findings of three activities focused on the distributions of quantitative variables using activities modified from those found on the *TinkerPlots*™ website: *Sketching Distributions*, *Mystery Mixers,* and *Fish-Length Distributions* (http://www.tinkerplots.com/activities/data-analysis-and-modeling-activities). Activities were translated into Turkish and some questions were modified for length or to split them into multiple parts.

Activity 1: Introduction to distributions. This activity was based in the *TinkerPlots*™ Sketching Distributions Activity and involves describing the shape of a data distribution (left skewed, normal, right skewed, uniform, etc.) using its center and variability, and identifying distribution properties. In the first part of the activity, the teacher handed out the activity sheets showing 5 distributions with different shapes. The teacher showed larger versions of the distributions one by one and asked the students to match the distribution shown on the screen with the distributions on their worksheet. the bigger ones of these shapes of the distributions to the class and let students match them. Then students opened a file in *TinkerPlots*™ and created a distribution for the variable, height. The researcher then discussed the shape of that distribution along with other potential shapes. The research then asked the students to find variables in the data set whose distributions matched the shapes of the distribution on the worksheet. This activity was designed to help students create their own distributions for the variables they chose and recognize the differences between them. We waited for students to discuss "which variables form a left- (or right-) skewed distribution." When the researcher realized the students understood the shapes of distribution, we handed out the worksheets. The students had to match the distribution models they created in the worksheets to the displays on their screen. The student was asked to write the variable he/she chose for the distribution on the worksheet.

Activity 2: Discovering the distributions. This activity was modified from the *TinkerPlots*™ Mystery Mixers Activity and aims to help students discover the features of the distribution of the sample when increasing sample size. In this activity students interact with the sampler tool. This activity contains four different data sets, each containing 500 positive integers between 0 and 100, but with different distributions. There data in the first and third mixers followed a normal distribution. The second mixer contained a left-skewed distribution, and the fourth mixer contained a right-skewed distribution. Each mixer's sampler generator was set to draw five numbers from the set and add them to the sample. Students were asked to estimate the center of the population distribution from the distribution of the sample using the minimum sample size.

The goal is to increase the size of the sample until you get a sense of the center of the hidden distribution, but each observation has a 'cost' and students were challenged to find the center of the population distribution keeping the cost as low as possible. We asked them to decide on the sample sizes in the worksheet and the distribution's center to better understand the distributions they created. This way, students could estimate the data clumping region and discover different distribution displays while creating a sample which is representative of the population.

Activity 3: Comprehending the distributions. This activity was modified from the *TinkerPlots*™ Fish-Length Distribution Activity and teaches how to compare two distributions using means and medians. Students had to decide whether to buy genetically modified (GM) fish from a farmer. There were two types of fish: fish farmers' fish and GM fish. *TinkerPlots*™ contains a dataset representing a population of 625 fish of both types combined. Students would select samples without replacement, first of size 130 and then of size 15, and would compare the fish lengths in two groups using their center and variability. The students were instructed as follows: 'Sample 130 fish and calculate their mean and median. Repeat 2 or 3 times. Do mean and median change after running the sampler for 130 fish? If so, how? How long should the sampler be to tell if GM or non-GM fish are longer? Why? Compare the sample averages to the population average. What do you get?' While comparing, students used several tools such as averages (mean and median) and hat tool. In the second part of the activity, students were asked if samples of 15 and 130 fish were sufficient to conclude that GM fish were longer. We expected them to confirm their answers while comparing them with the size of the population, that is, 625 fish. The goal was for participants to see the relationship between the distribution of the population and the sample and explain it by relating it to centers and variability.

The data collection tools used in this study were three worksheets for each student, video recordings of their screens, and field and observation notes of the researcher. The video recordings included the students' work on *TinkerPlots*™ during the teaching. The lesson started with distribution of activity worksheets, and running *TinkerPlots*™, then we started to record their work on the computers. Field and observation notes included the researchers' notes taken during the classes and their pair work. During the data analysis, we triangulated the data sources and confirmed the results with each other.

3.3. ANALYSIS

Descriptive data analysis was used to analyze the collected data in a deductive approach. While focusing on the concept of distribution, the descriptive analysis was done according to Table 1, which shows the conjecture, data, and comparison models for each activity according to the RISM framework laid out by Dvir and Ben-Zvi (2018). The data analysis aims to present the findings directly to the reader in an edited and interpreted manner. Therefore, the data obtained is first described systematically and clearly, then these descriptions are explained and interpreted, and direct quotations are often given to reflect the participants' views.

Table 1. Data analysis

	Conjecture Model	Data Model	Comparison Model (IIR)
Introduction activity	Distribution curve of the data according to the 'height' variable.	Other displays obtained with different variables similar to the distribution displays included in the activity sheet.	Classification of distribution displays by matching the obtained distribution curves and the distribution displays on the activity sheet.
Discover Activity	The minimum sample size determined in the mixer specified on the *TinkerPlots*™ window.	Determining the central aggregate regions using *TinkerPlots*™ tools for distribution displays obtained from the data via the mixer.	The relationship between the distribution displays determined by sample sizes obtained in the mixer and the central aggregate regions.
Comprehend Activity	Comparing the lengths of normal fish and GM fish (understanding the difference).	Generating and comparing distribution displays for normal and GM fish with a sample size of 15 and 130.	Sampling by comparing the distribution displays of samples created with a sample size of 15 and 130 and population.

Rather than providing quotations from each participant in each activity, we summarized the students' responses to ascertain their understanding of the statistical concepts discussed in the activities. As a result, we did not provide all of the participants' responses; rather, we analyzed how students reached at the concepts from their perspectives and presented their summaries.

4. FINDINGS

In this study, we examined the conjecture models, data models, and the RISM processes arising from the comparison of these two models to examine how they sense, discover, and comprehend the distribution concept.

4.1. FIRST ACTIVITY: INTRODUCTION TO DISTRIBUTIONS

We explained to the students the stages involved in generating the distribution for the variable height, as illustrated in Figure 3. This display of the distribution was created by all students. We were curious as to how our participants would transfer the distribution they saw on the screen to a blank piece of paper.

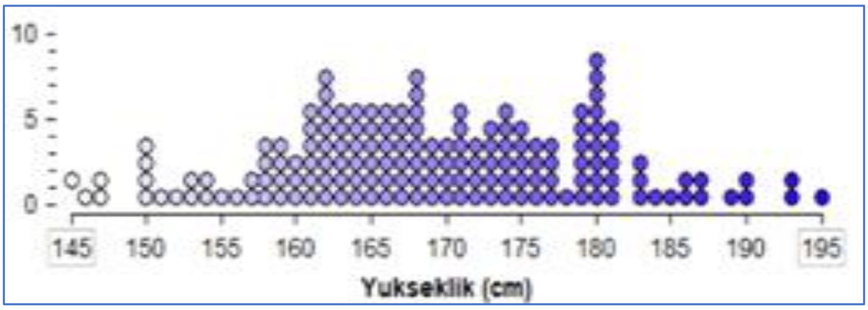


Figure 3. The distribution of the height data.

Students then sketched distributions of other variables in the dataset and attempted to depict the distribution as a curve by illustrating the increases and decreases in the groups into which the data were stacked. Aylin and Selim's sketches can be seen in Figure 4. In her sketch (Figure 4(a)), Aylin addressed the variable 'TV time watched on weekends.' As seen by the erasures, Aylin first drew a jagged curve and then created a smooth curve. Selim retained the jagged features in his sketch (Figure 4(b)). Additionally, he placed the Y-axis to indicate the region in which the data is stacked.

Students mostly focused on the increase and decrease in the stacks of data in specific regions as a result of using incorrect variables in their conjecture model. Aylin and Selim created a conjecture model by sketching the researcher's curve rather than the distribution of the variable they chose and graphed on the *TinkerPlots*™ screen. Additionally, it was noted that only Reyhan was unable to create a model in *TinkerPlots*™, even though she paid attention to the instructions and followed the researcher's directions.

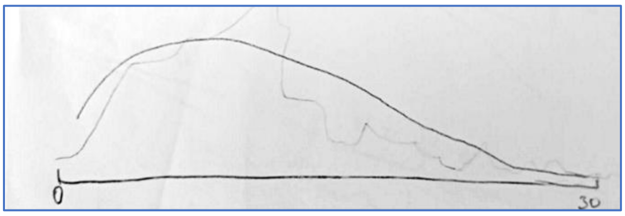


Figure 4(a): Aylin's sketch.

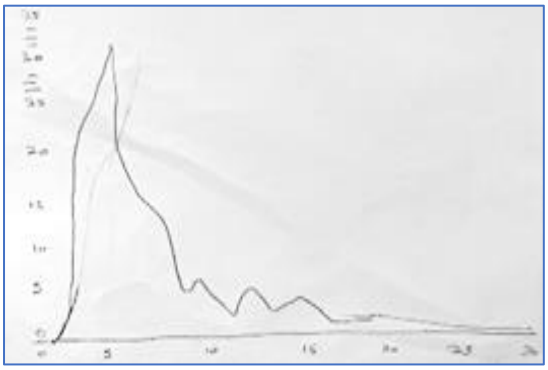


Figure 4(b): Selim's sketch.

Students discussed their sketches in class in accordance with the researcher's instructions and agreed on the different shapes of distributions and the essential characteristics required in an appropriate sketch of the distribution. At this point, Fatih and Reyhan chose the variables birth month and grade level and created the graphs shown in Figure 5. Though they could match their displays to the shapes discussed, the students were unaware that the variables they chose were categorical, and not continuous.

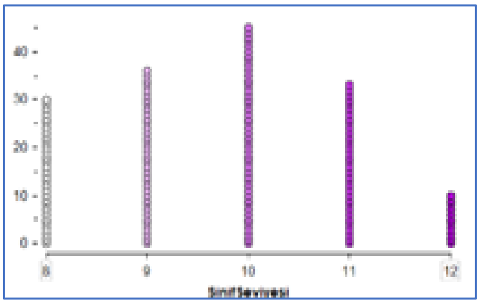


Figure 5(a). A section from Fatih's work on TinkerPlots™ (data model).

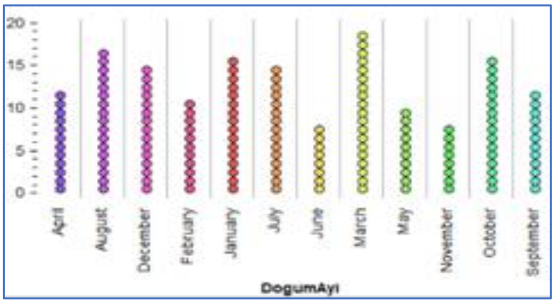
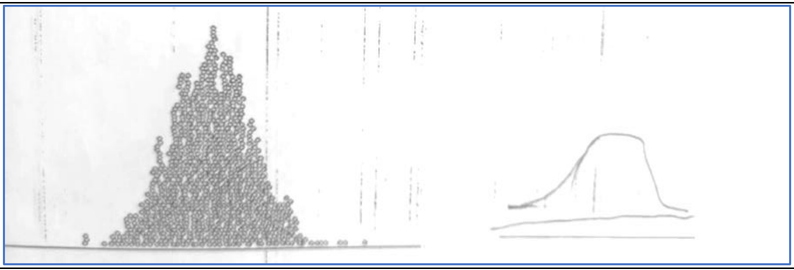
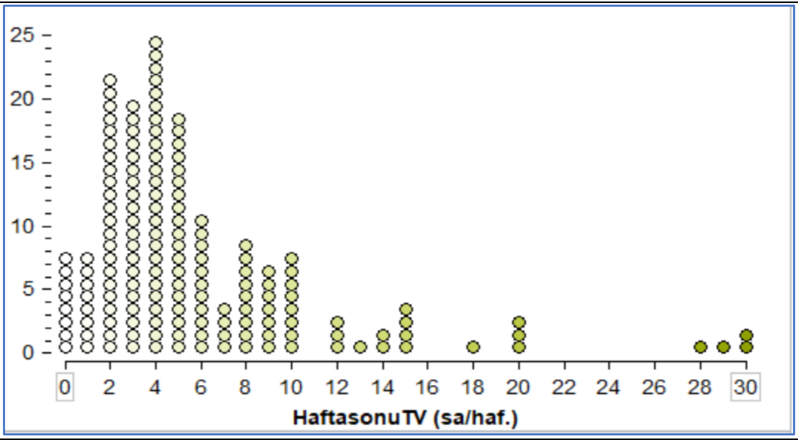
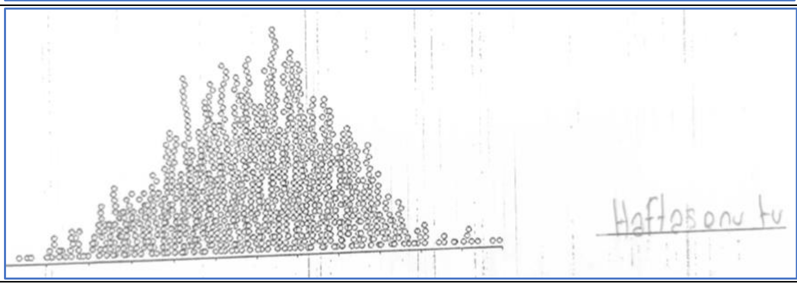


Figure 5(b). A section from Reyhan's work on TinkerPlots™ (data model).

Furthermore, Fatih asserted that the distribution he constructed for the grade level variable was a normal distribution, which it was not. Reyhan asserted that the representation in Figure 5(b), which she generated using the birth month data, demonstrated a distribution. Fatih perceived the data as dispersed and ensured the data overlapped with the stacking tool he has used.

The distributions presented in the worksheet belonged to larger data groups and were different from the data models created by the students. For instance, Hakan experienced the modeling process as seen in Table 2 by expanding his distribution display in *TinkerPlots*™ for the variable 'TV time watched at the weekend' according to the conjecture model in the worksheet. His inference was that the display similar to the conjecture model on the worksheet could be classified and named for each type of distribution that was learned, such as skewed or normal distribution (Table 2).

Table 2. Hakan's modeling process

Conjecture Model	
Data Model	
Comparison Model	

From Table 2, we see that he represented the data increase region as more rounded in his conjecture model's distribution curve, since he sketched a normal curve. Then he separated the *HaftasonuTV* in the data model he constructed, allowing the data to form a distribution representation. When he reviewed the comparison model, he reasoned that the distinct placements of the values in the data group in the worksheet and the data in the data model he constructed should be same. He was unconcerned with the minimum and maximum values, but also with the conjecture model generated by the distribution representation as a whole.

In general, students made inferences by focusing on the displays' peaks, as shown in Figure 5. While they were determining the shape of distribution according to the increase and decrease of the peaks, some thought of only the data stacked at specific regions as distribution. For instance, Aylin determined the variables in the activity according to those who could create a distribution. Yaman made something similar to what Aylin did, and he explained as follows:

> Yaman: When some of the dots [cases] were stacked on top of each other, I had to drag them sideways and align vertically, but for others, this was not necessary. After creating the distribution that I prepared for the variables, I matched them normal, left- and right-skewed

This finding brought the understanding that students now thought of the distributions according to how data spread. We claim that they viewed the distribution intuitively by seeing the data display holistically.

When we examined the participants' inferences overall, we concluded that Hakan, Halit, Utku, and Fatih, who were able to sense and express the notion of distribution, compared the variables and distribution shapes. Reyhan, Yaman, Aylin, and Selim, however, did not demonstrate they had developed a sense of distribution. They based conclusions on the conjecture models they generated from the distribution representations provided in the activity sheet, without comparing distribution shape.

4.2. SECOND ACTIVITY: DISCOVERING THE DISTRIBUTIONS

The students' conjecture models had the sample sizes (for each of the four mixers in the activity) determined by the distribution process. Selim decided to group the data using the divider tool. The sample sizes generated by Selim for each of the four mixers were consistent with each other. Selim said that "when the distribution is on the left or right side, I use the divider tool more to get more regions. When it is normal distribution, I do not need to get many regions." It can be concluded that Selim discovered the concept of distribution while he was considering the center of the distribution. It was observed that while using too many dividers to predict the location of the center for a right- and left-skewed distribution, he used the divider tool less because he knew that the center of the normal distribution would be in the middle. In addition, when deciding on the distribution shape, Selim specified the regions where data were stacked. Therefore, we can claim that Selim successfully discovered the distribution of the samples he created by assuming consistent sampling for all four mixers, as shown in Figure 6.

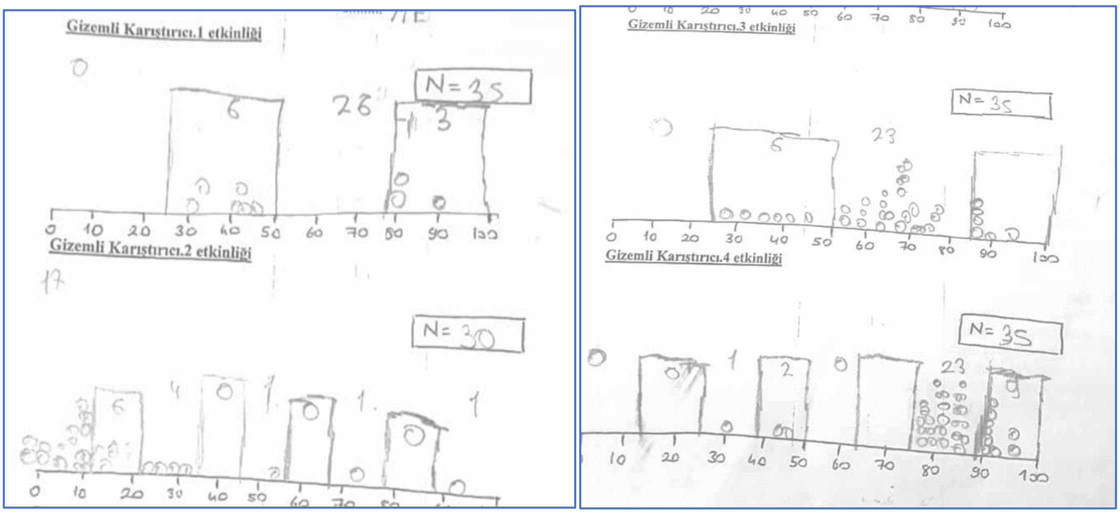


Figure 6. Selim's sketches on the paper for four mixers during the second activity.

Reyhan and Halit selected the same sample size for each mixer (30 and 40 for the sample size, respectively). Although they chose the same sample size for each mixer, their statistical reasoning was different. For example, the following dialogue took place between the researcher and Reyhan:

Reyhan: I try to take the highest number [of the dots] in one box (in the region) and the lowest number in the others.
Researcher: Why do you always choose the same sample size?
Reyhan: Because you asked us to choose the smallest sample size. When I set a smaller number [for sample size], it does not work that way. There is not too much in one box (in the region) and less in the others.

Since Reyhan could not establish a sense of distribution, she did not approach the data holistically. Therefore, she could not make an inference about the center of the sample. Unlike Reyhan and Halit, Utku was the student who assumed the most inconsistent and largest sample about distributing the data or the center of the data.

Due to the randomness of the samples created from four mixers, each student would come up with a different sample size. Therefore, each student was analyzed according to their work without being compared with others. The important thing is that the students could see the data distribution with the same amount of repetition. Selim decided on how the data was spread by discovering the shape of the distribution. It enabled him to select smaller samples without being affected by the randomness of the samples. In contrast, Yaman made a correct prediction about the central region of the sample distributions he created, but because he could not discover the distribution of the population, he decided to sample it with more data. Due to this situation, Yaman did not make a decision based on the data in the fourth mixer on the central region. The central region was the region in which data was stacked the most.

We expected the students to make their conclusions using the divider tool in *TinkerPlots*™ and associate the center of the distribution shape with the increase in the sample size. The important thing here was not to what extent the students divided the data or which size they decided to sample. We wondered how many regions the students would divide the distribution into according to the sample size to decide on the distribution shape.

Fatih estimated the center of the distribution by dividing all the sample distributions he created into ten regions. Although his reasoning made it easier for him to see that the specific center of the data model that he generated consisted of ten divisions, the sample distribution in the fourth mixer was not according to this order and consequently caused him to predict the center inaccurately. Even though Reyhan worked in all sample distributions with the same number of divisions, she never changed the number of data and divisions for each sample in the mixers. Selim chose to use a small number of dividers for normal distributions and more dividers for left- and right-skewed distributions. Figure 7 below shows Hakan's data model while he was trying to find the central region. He tried to show the divider tool by expanding the region where the data was mostly stacked.

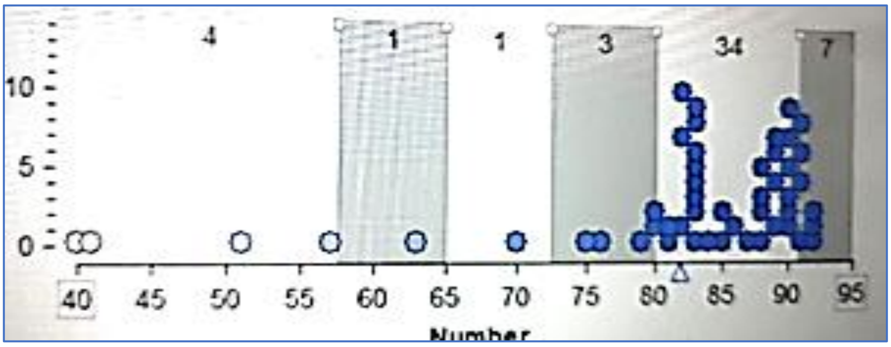


Figure 7. Hakan's data model based on the fourth mixer sample.

Hakan enlarged the region rather than increasing the sample size of the fourth mixer sample to ensure the data entered the intended region. He adjusted the region's size because the regions created with the divider tool were not what he wanted. As seen in Figure 7, the first section is larger than the others, and their sizes may be adjusted by dragging the dots in the gray boxes' corners. Noticing this incident, the researcher inquired as to why Hakan picked that particular tool. "I'm not sure," Hakan said. "I couldn't locate the spot where it was heaped up. I considered many possibilities for this. When I picked this, it felt natural for me to pick the section with the data layered on top of one another. That is why I choose it."

The procedure by which students increased the sample size using the 'RUN' key on the sampler and decided which regions to include using the divider tool was supposed to be the data modeling process. Students arrived at the center aggregation region by engaging in the modeling process and exploring the central aggregation region using either the sample size or the number of divisions.

4.3. THIRD ACTIVITY: COMPREHENDING THE DISTRIBUTIONS

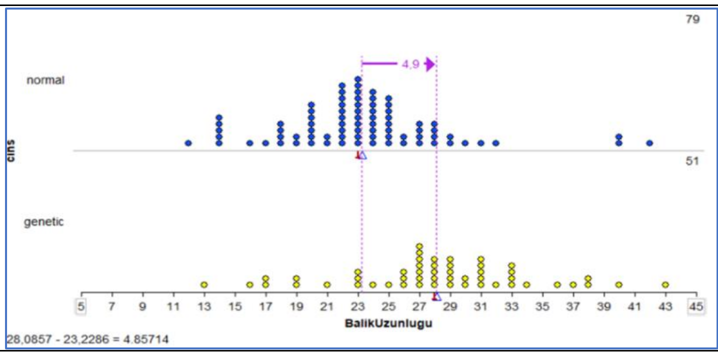
The students built their conjecture models with *TinkerPlots*™ to determine whether GM fish or normal fish were longer, according to a sample of 130 fish. In these models, we observed that some students responded with their ideas, while others still did not fully understand the distribution concept. While Aylin, Hakan, Utku, and Reyhan each prepared a dot plot (see for example Figure 3), Fatih, Yaman, and Halit did not produce a dot plot. The students who created dot plots answered the research question and saw that GM fish were longer. Of the students who did not create a dot plot only Fatih claimed that normal fish were longer than GM fish, when later he constructed a distribution of length variable. None of the other students who did not create dot plots answered that question correctly.

Anticipating that students knew little about variability, we used the hat tool (a tool that makes a hat plot of data) for groups of GM fish and normal fish to determine how they decided on how variable the distribution was. Except for Utku, students said that the distribution of GM fish was more varied than normal fish. Utku answered the question by stating, “There was neither an increase nor a decrease in normal fish, but it gradually increased and slightly decreased, which means that it is unstable, which means that it is variable.” By looking at the distribution of normal and GM fish of different sample sizes using another *TinkerPlots*™ tool (such as divider, averages tool), we wondered what inferences the students would make for sample size this time. The students first examined two different sample distributions composed of 130 fish. Students created data models using different *TinkerPlots*™ tools when comparing distributions. It was Aylin, Hakan, and Yaman who chose both mean and median using the Averages tool (it is a tool with which users can find the mean, median, or mode of the data set). Although these students found at least five different measurement results, we noticed they took notes from the results whose arithmetic mean and median values were the same or very close to each other. For instance, Aylin explained the reason why her measurements were similar as follows:

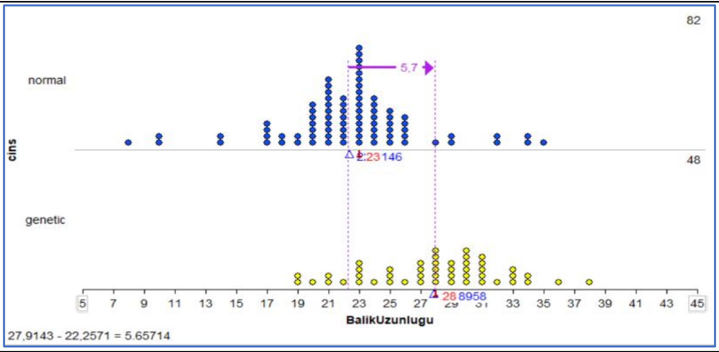
Aylin: Every time I press the RUN button, the values of the fish-types change. But sometimes this [causes little change], sometimes there is a lot of difference. I am looking to see how much the difference has changed. I write on a paper when I find close results

Comparing the sample values, Yaman compared the distribution of the same fish groups with different sized samples. Aylin discovered variability by interpreting the averages according to only one of them, even though she showed both of them on the distribution. Table 3 below shows Aylin’s modeling for 130 fish. Throughout the modeling procedure, Aylin only employed tools to determine the arithmetic mean and median values. She customized the tool she used to convert the mathematical mean and median values she determined to their numerical equivalents. Additionally, while she determined the arithmetic mean and median values for each distribution representation, she was primarily interested in determining the difference between the arithmetic mean values when comparing two distinct distributions.

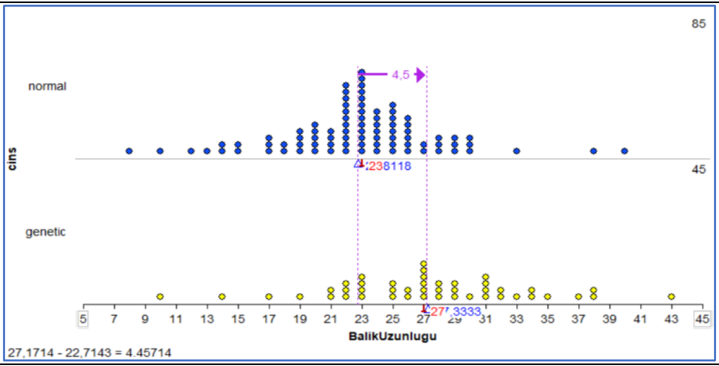
Table 3. Aylin’s modeling for 130 fish (the variable seen in x-axis is the fish length)

Modeling 1	

Modeling 2



Modeling 3



Along with Aylin's modeling process (Table 3), Hakan considered the difference between the median and arithmetic mean values produced for each distribution representation he constructs. When the researcher inquired as to why Hakan kept repeating himself, Hakan stated, "The two symbols (for mean and median) over there should be near to each other. Then it is true." Then, we asked the students the question of, "How do you comment on the distributions you created for 15 and 130 fish when you compare them to the distribution you created for all the fish in the lake (population of 625 fish)?" We examined the inferences they made by referring to the question. When Aylin chose a different sample size, she made her explanations with variability. She said that for 15 fish, the center regularly changed; for 130 fish, the center moved less; for 625 fish, the center did not change at all in any of the random distributions. Therefore, Aylin thought that by selecting all 625 fish, she could make a better decision about whether the normal or GM fish would be longer. Halit made his decision based on the distributions of the samples he created and made an inference from the data stacked in these distributions. He said that among 15 fish and 130 fish, he could find a more accurate answer to the research question with 130 fish because 15 fish did not show any center. When he examined the population, however, he said he would decide to use that distribution because it gave more insight into the 625 fish in the population.

Hakan, who was interested in the displacement of the center of the resulting distributions (he could not identify the center from the distribution of data), said that "there was no distribution with 15 fish," which was similar to Halit's explanations. He discovered, as Aylin did, that the center of each distribution was also different. Hakan however, reasoned with the distance between the median and mean of data sets, unlike other students, as seen in Figure 8 below. He explained that 130 fish were better at deciding which type of fish would be longer because when he worked with 130 fish, the averages were closer to each other. Among the students, only Reyhan could not produce any ideas about the relationship between the sample and the population, or the meaning of the distribution display she generated, or about noticing the variability. She said that she had seen the distribution with 15 fish in greater detail, found the center more easily, and struggled to have the averages of each sample she created to be the same.

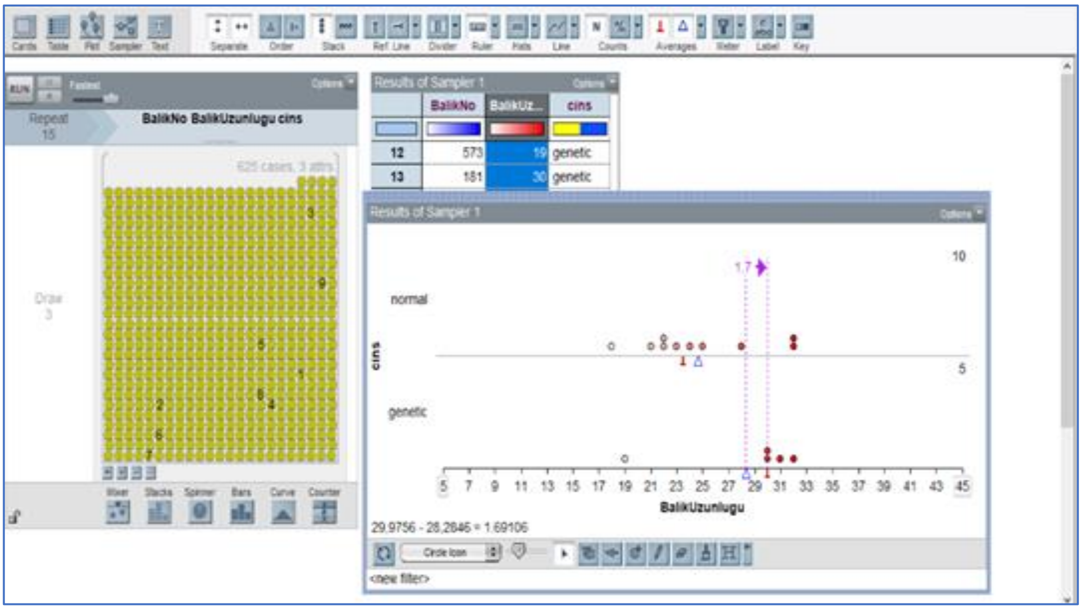


Figure 8. A screenshot showing Hakan's modeling process in Fish Length Distribution activity.

When we examine the findings of Fatih, Halit and Yaman, we claim that since they could not create dot plots, they could not make comparisons about the lengths of normal and GM fish. We noticed that only Aylin was using the distribution curve for the representation of the data, while Halit and Utku paid attention to the columns where the data were stacked. Only Yaman wanted to compare the data using a table. When the informal inferences of the students are examined, we can see they follow different processes.

The findings are summarized in Table 4 using the RISM framework. The students participated in three distinct tasks that focused on feeling the distribution, understanding the distribution, and grasping the distribution, in that order. In addition, the table indicates how the students developed distribution models throughout the activities. During the process of distribution modeling, the students used informal inferential reasoning on statistical concepts such as center and variability. Moreover, we may assert that it is possible to teach fundamental statistical ideas without focusing solely on statistical calculations on the data set.

Table 4. Summary of overall findings

Activity	Conjecture Model	Data Model	Comparison Model (IIR)
First Activity	The distribution display, which was seen as individual cases, began to be seen holistically.	The distribution display was created by stacking the data vertically and spreading it horizontally.	Distribution displays belonging to different data groups were grouped according to distribution shape.
Second Activity	Distribution display was obtained by increasing the amount of data with data selection using the sampler tool.	The location of the center was estimated according to the distribution shape.	The outlier found on the distribution display was eliminated when estimating the location of the center.
Third Activity	Two different distribution displays were compared with respect to a particular variable.	The distribution shape formed according to the size of the data obtained from the sampler tool was determined.	According to the statistical inference, the distribution display was constructed with the data group whose size was determined by the student.

5. DISCUSSION

In this research, the findings revealed the inferences students had obtained in their modeling processes with data simulated by *TinkerPlots*™. We looked at their statistical modeling through the RISM framework's window to examine the students' reasoning processes. In this section, we discuss the contributions or differences that the findings can make with regard to the theoretical framework of RISM, under the headings of the meanings that stand out in the activities. We discuss and interpret how the students' emerging statistical reasoning evolved through the modeling process they experienced. The findings of this study show that the students constructed alternative distribution models, and we will discuss these reasons and how students' statistical reasoning was facilitated by this in the following part.

5.1. DEVELOPMENT OF STUDENTS' MODELING

We observed in the first activity that the students created the conjecture models by showing the given variable type and aggregated regions, especially according to the increases and decreases in the data set's peak values, and by attending to the minimum and maximum values of the data set. The findings given in Figure 4 show that students express the data, which they regard as individual cases, as a line graph, as they show the distribution with the line between peak values at first sight. As Cobb et al. (2003) claimed in their study, students tend to see the data sets as hills. Therefore, students were trying to use the connect these hills in the form of a line graph.

We did not see any difference in students' design of data models in *TinkerPlots*™, but the fact that students thought the models seen in Figure 5 form a distribution suggests they did not consider the types of variables, because the displays were of categorical variables, and the students who made these models did not attend to how the data were clustered. Concerning this, Watson (2005) stated that young children form distribution displays and that it is typical for complex concepts and characteristics of distribution to be included in these displays to a large extent.

In their study, Dvir and Ben-Zvi (2018) interpreted their participant's (Erez) modeling process by exploring how they made a transition from the conjecture model to the data model. According to the RISM framework, when the students' conjecture and data models were compared in the second activity, the divider tool used by Selim when switching from the conjecture model to the data model established the relationship between the models. According to this relationship, Selim noticed the stacks according to the different shapes of distribution presented in the second activity.

5.2. DEVELOPMENT OF STUDENTS' INFORMAL INFERENTIAL REASONING

In our study, the students' inferences about the shape of distributions were based on matching them with different distribution displays from the data models. The literature states that students need to focus on the characteristics of the data sets to think holistically about distribution, which begins as individual cases in their eyes (Ben-Zvi & Amir, 2005; Konold & Higgins, 2003). Furthermore, students were confused when they were matching the types of non-perfect probabilistic distributions of different data sets while they were easily matching more perfect distribution displays. We agree with the conclusions of research involving the mapping of shapes, histograms, and distributions (Scheaffer et al., 2004; Rossman & Chance, 2005); furthermore, this situation enabled students to see concepts such as variability of a distribution and center by expanding their thinking about distribution in their assumptions.

It is impossible reason completely about distribution without considering its variability (Garfield & Ben-Zvi, 2008). In the second activity in our study, the students formed the conjecture models with the sample size they chose to reason about the variability of a distribution. Students decided on conjecture models based on the sample size, type of distribution, the multiplicity of data, and the number they randomly obtained. Bakker (2004b) claimed, in his growing sample activity, that students are better at predicting what will happen to the graph by showing that variability decreases as the sample grows and is more similar to the population. We initially told the students to estimate the minimum sample size, and we aimed to estimate the display generated by the addition of cases. Thus, we can argue that when students grew the sample themselves, they found that they did not need to know the size and population distribution to notice that the variability decreased. This claim led us to think that students could make inferences about the population from the given sample.

Furthermore, technological tools in students' distribution modeling process allowed us to see their reasoning on the sample size (Mills, 2002). Therefore, we claimed that students who never changed the sample size were unwilling to take risks, while students who increased the sample size could not estimate the stacks of data. Students who were in this situation reasoned more roughly and superficially. These students might become more precise with a few, minor instructions. We can argue that the students could establish the link between the sample and variability, which was the focus of the third activity. Therefore, the activity of "discovering the distributions" could be a good teaching point to highlight the representativeness of the samples (Watson & Moritz, 2000). Besides, since there is no formal rule for normal, left- or right-skewed distributions in this activity, and because the activity was presented in a game-like manner, students could be motivated to participate and discover intuitively the distributions that Bakker (2004a) stressed in his study.

In the third activity, we expected students to draw inferences to examine how they made sense of the distribution they created. Students who were able to construct the distribution in the conjecture model could compare the distribution of GM fish with normal fish by commenting correctly on the variability. While students commented on the variability of distribution, although they used the hat plot tool, they chose to explain the difference between the data and both distributions. They informally defined variability by their statements, such as that GM fish showed a more unstable form, an increase, or a decrease. We can say that the students had an understanding of variability in search of different regions by defining two different distributions. Lehrer and Schauble (2004) studied natural variability where children seek plateaus as low variability indices. The results from that study match our findings and support our claim. In contrast, students were unable to transcend their sense of variability. Therefore, it is necessary to approach sampling probabilistically in order to see the distribution as a *lens* through which to view the variability, as explained by Wild (2006). We recognize that students lack an understanding of variability when comparing the distributions of various samples obtained via sampling. According to Shaughnessy's (2019) findings in his "candy sampling task", probabilistic thinking may be beneficial for comprehending variability. This finding could be explained by the fact that our participants have not received formal instruction in probability, and thus their probabilistic reasoning is limited. Garfield et al. (2008) emphasized this point by highlighting the need for students to focus on the nature of variability via sampling.

We examined students' reasoning with sample distributions to make sense of distribution for different activities. We noticed that students thought more deeply about their distributions using the averages, ruler, divider, and line plot tools of *TinkerPlots*™ in the data models they dealt with the sampler tool. Although the tools they used made it easier to understand their data models, their inferences were valuable. We thought that Hakan made his inference for 15 fish in Figure 8 during the data modeling process by associating it with the tools he used. Hakan sought variability between the data while making sense of the representation of the data and brought the concept of variability to the fore. The concept of distribution forms the basis of nearly all statistical reasoning methods related to variability (Wild, 2006). Although participants were at the university level in studies on variability, they were shown to pay attention to the variability criterion of data and expressed their representations over the variability it represented (delMas et al., 2007; Konold & Pollatsek, 2002).

The students drew inferences on sample-population balance by comparing their results with different samples in the sampler tool with conjecture models and data models. They conveyed their inferences with different statistical concepts. These concepts were in the form of the center, variability, or spread of distribution display. We cannot argue that students made right or wrong inferences based on the basic concepts they addressed, but we observed how students wanted to see them in distribution. Aylin compared the randomly generated sample distributions and the center of the population when making an inference, as she only saw the center of the distribution. Since the centers of the samples created varied, Aylin did not trust that all data would be represented. Fatih struggled because he often used categorical instead of quantitative variables. This student made inferences by creating the least number of different distributions among the students and selecting the tool less. Despite this, he gave the correct answer as he saw the variability of different sample distributions. Hakan made inferences by associating them with both centers and variability of the different distributions he created. Hakan experienced the representability of the display based on the sample size, understanding how little or much the sample changed the center of random data, and the relationship between the sample and the population by making sense of the distribution displays.

6. CONCLUSION AND IMPLICATIONS

We have investigated the students' modeling processes through the RISM framework by comparing the data model with the conjecture model and the pros and cons of the inferences obtained in this dynamic process. While students could easily make their distribution displays in *TinkerPlots*™, they could not accurately relate these models to the real-world settings. The reason could be because they did not have inferences on statistical concepts in their modeling processes.

It is becoming increasingly important to focus on the best ways to use technological tools in the classroom, as we concluded in this study. Examples of some of the practical uses of technology are seen in statistics classes, and such tools enable students to answer, "What happens if?" type questions while students follow the changes seen on their screen (Chance et al., 2007). The students' statistical modeling process is complicated because of the simultaneous use of real-world problems and the *TinkerPlots*™ software in the activities. *TinkerPlots*™ provides students with the opportunity to see the connection between the data and chance. The activities, including some tasks such as repeated measures, can help students to see the idea of data as signal and noise (Biehler et al., 2012). Biehler et al. also claimed that students can understand statistical concepts more quickly if they deal with computation or graphing less via the software.

To understand the data, the students experienced their relationship with the basic concepts of graphs, distributions, and statistics, not through the eyes of a statistician but purely for informal reasoning. However, students cannot use only informal reasoning to understand the data. They need to experience the relationship between the data display, distribution, and fundamental statistical concepts to reach informal reasoning. Therefore, students do not have to learn statistical concepts in a specific order, and what is essential is not the order of understanding statistical concepts, but the relationships of how students comprehend these basic concepts must take place in a particular order. Examining the students' usage of the idea of distribution reveals several statistical concepts, such as variability and center, and cycles of feeling, understanding, and interpreting the distribution, based on our findings. According to Pfannkuch and Reading (2006), the distribution reasoning process should be investigated in the context of statistical inference, because it is essential for comprehending the statistical concepts and recognizing how reasoning is performed. Rumsey (2002) also claimed that statistical reasoning and thinking are at a higher level than statistical competence, which means that the level of thinking, rather than knowledge, is essential.

Students can make inferences about the basic concepts of statistics if the guidelines and questions enable them to reason about statistical activities. Because the students' comparisons of the conjecture and the data models they created will cover the entire activity, the cycle between the order, relationship, and usage of statistical concepts are complex. For instance, Garfield and Ben-Zvi (2005) concluded in their review for variability that "understanding variability is much more complex and difficult than what the prior literature suggests" (p. 93).

The students' probabilistic approach to distribution through their self-inference during the whole modeling process will enable them to make better sense of the complex statistical concepts they will learn at a more advanced level. In summary, the discussion shows how students comprehend basic statistical concepts such as variability, center (mean, mode, median), and shape of a distribution in an empirical distribution generation process and how they feel about theoretical distributions (Garfield & Ben-Zvi, 2008). This relationship coincides with the transition from their conjecture models to data models that stands out in the RISM framework (Dvir & Ben-Zvi, 2018). With the activities implemented in this study, students' modeling skills improve more effectively when they make sense of statistical concepts. Therefore, we think that it would be meaningful and valuable to provide conceptual knowledge at an early age so that students can perform data analysis and statistical inference simultaneously during statistics teaching.

REFERENCES

Aridor, K., & Ben-Zvi, D. (2017). The co-emergence of aggregate and modelling reasoning. *Statistics Education Research Journal*, *16*(2), 38–63. https://doi.org/10.52041/serj.v16i2.184

Arnold, P., & Pfannkuch, M. (2014). Describing distributions. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education.* Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona. https://icots.info/icots/9/proceedings/pdfs/ICOTS9_8G1_ARNOLD.pdf

Bakker, A. (2004a). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64–83. https://doi.org/10.52041/serj.v3i2.552

Bakker, A. (2004b). *Design research in statistics education: On symbolizing and computer tools*. [Doctoral dissertation, Universiteit Utrecht]. CD-β Press.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*(1–2), 5–26. https://doi.org/10.1080/10986065.2011.538293

Bakker, A., & Gravemeijer, K. P. (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). Springer.

Ben-Zvi, D., & Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies.* Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-4), Brisbane, Australia.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, *45*(1–3), 35–65. https://doi.org/10.1023/A:1013809201228

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2012). Technology for enhancing statistical reasoning at the school level. In M. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds) *Third international handbook of mathematics education* (pp. 643–689). Springer. https://doi.org/10.1007/978-1-4614-4684-2_21

Braham, H. M., & Ben-Zvi, D. (2017). Students' emergent articulations of statistical models and modeling in making informal statistical inferences. *Statistics Education Research Journal*, *16*(2), 116–143. https://doi.org/10.52041/serj.v16i2.187

Brown, E. N., & Kass, R. E. (2009) What is Statistics? *The American Statistician*, *63*(2), 105-110. https://doi.org/10.1198/tast.2009.0019

Canada, D. (2004). *Preservice teachers' understanding of variation*. [Doctoral dissertation, Portland State University].

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1). https://doi.org/10.5070/T511000026

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and instruction*, *21*(1), 1–78. https://doi.org/10.1207/S1532690XCI2101_1

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches.* SAGE Publications.

delMas, G., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2), 28–58. https://doi.org/10.52041/serj.v6i2.483

Dvir, M., & Ben-Zvi, D. (2018). The role of model comparison in young learners' reasoning with statistical models and modeling. *ZDM Mathematics Education*, *50*(7), 1183–1196. https://doi.org/10.1007/s11858-018-0987-4

Dvir, M., & Ben-Zvi, D. (2019). Repurpose and extend: Making a model statistical. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Eleventh Congress of the European Society for Research in Mathematics Education* (CERME11) (pp. 922–929). Utrecht, The Netherlands.

Dvir, M., & Ben-Zvi, D. (2023). The double-edged sword of conjecturing. *Mathematical Thinking and Learning*, *25*(2), 153–176. https://doi.org/10.1080/10986065.2021.1940427

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, *4*(1), 92–99. https://doi.org/10.52041/serj.v4i1.527

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). Learning to reason about variability. In J. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning* (pp. 201–214). Springer.

Hammerman, J., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17–41. https://doi.org/10.52041/serj.v3i2.546

Kazak, S., Fujita, T., & Wegerif, R. (2014). Year six students' reasoning about random "bunny hops" through the use of TinkerPlots and peer-to-peer dialogic interactions. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education.* Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217–230. https://doi.org/10.1007/s10758-007-9123-1

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). National Council of Teachers of Mathematics.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289. https://doi.org/10.5951/jresematheduc.33.4.0259

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635–679. https://doi.org/10.3102/00028312041003635

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1), 152–173. https://doi.org/10.1080/10986065.2011.538301

Makar, K., & Confrey, J. (2002). Comparing two distributions: Investigating secondary teachers' statistical thinking. In B. Philips (Ed.), *Developing a statistically literate society.* Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS-6), Cape Town, South Africa.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54. https://doi.org/10.52041/serj.v4i1.524

Manor, H., & Ben-Zvi, D. (2017). Students' emergent articulations of statistical models and modeling in making informal statistical inferences. *Statistics Education Research Journal, 16*(2), 116–143. https://doi.org/10.52041/serj.v16i2.187

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1). http://www.amstat.org/publications/jse/v10n1/mills.html

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–137.

Noll, J., & Shaughnessy, J. M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education, 43*(5), 509–556. https://doi.org/10.5951/jresematheduc.43.5.0509

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.*

Patton, M. Q. (2014). *Qualitative evaluation and research methods* (4th ed.). SAGE Publications.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, *5*(2), 27–45. https://doi.org/10.52041/serj.v5i2.498

Pfannkuch, M., & Reading, C. E. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal 5*(2), 4–9. https://doi.org/10.52041/serj.v5i2.496

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM Mathematics Education, 50*(7), 1113–1123. https://doi.org/10.1007/s11858-018-0989-2

Reading, C. E., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal, 5*(2), 46–68. https://doi.org/10.52041/serj.v5i2.500

Rossman, A., & Chance, B. (2005). *Workshop statistics student toolkit*. Key College Publishing.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910678

Scheaffer, R. L., Gnanadesikan, M., Watkins, A., & Witmer, J. (2004). *Activity-based statistics: Student guide* (2nd edition). Key College Publishing.

Shaughnessy, J. M. (2019). Recommendations about the big ideas in statistics education: A retrospective from curriculum and research. *Cuadernos de Investigación y Formación, 18*, 44–58.

Watson, J. (2005). Developing an awareness of distribution. In K. Makar (Eds.), *Reasoning about distribution: A collection of current research studies.* Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy, Brisbane, Australia.

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal, 8*(1), 32–61. https://doi.org/10.52041/serj.v8i1.456

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44–70. https://doi.org/10.2307/749819

Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10–26. https://doi.org/10.52041/serj.v5i2.497

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society Series A: Statistics in Society, 174*, 247–295. https://doi.org/10.1111/j.1467-985X.2010.00678.x

TUĞÇE BALKAYA
GAMZE KURT
Mersin University