STUDENTS' EMERGENT ARTICULATIONS OF STATISTICAL MODELS AND MODELING IN MAKING INFORMAL STATISTICAL INFERENCES

HANA MANOR BRAHAM
LINKS I-CORE, University of Haifa
hana.manor@gmail.com

DANI BEN-ZVI
LINKS I-CORE, University of Haifa
dbenzvi@univ.haifa.ac.il

ABSTRACT

A fundamental aspect of statistical inference is representation of real-world data using statistical models. This article analyzes students' articulations of statistical models and modeling during their first steps in making informal statistical inferences. An integrated modeling approach (IMA) was designed and implemented to help students understand the relationship between sample and population, as well as reasoning with models and modeling. We explore the articulations of a pair of primary school students, who had previously participated in the Connections Project exploratory data analysis (EDA) activities, and suggest an emergent conceptual framework for reasoning with statistical models and modeling. We shed light on ideas of statistical models and modeling that can emerge among primary students and how they articulate those ideas. Implications for teaching and research are discussed.

Keywords: *Informal statistical inference, Sample and population, Statistical model, Statistical modeling, Statistical reasoning, Statistics education research*

1. INTRODUCTION

The understanding that an inference can be made about a population from a sample is a fundamental aspect of statistical inference. However, studies indicate that students can hold contradictory views about the relationships between samples and their populations (Pfannkuch, 2008). For example, when focusing on sampling representativeness students may believe that a sample completely represents the population. On the other hand, while focusing on sampling variability, students may believe that a sample does not represent the population at all. One way to facilitate students' understanding of the relationship between samples and population is by modeling real-world situations with computerized tools and generating many random samples from these models (Garfield, delMas, & Zieffler, 2012).

Two types of settings have frequently been used in statistics education research to examine young students' reasoning with modeling: 1) scientific inquiry-based learning environments in which students are engaged in real-world data investigations (e.g., Lehrer & Romberg, 1996); and 2) probability inquiry-based learning environments in which students are engaged in manipulating chance devices such as spinners (e.g., Pratt, 2000). Modeling in the first setting refers to the construction and use of data to explain real-world phenomena. Modeling in the second setting refers to the construction of a distribution

Statistics Education Research Journal, 16(2), 116-143, http://iase-web.org/Publications.php?p=SERJ

model in a computer-based simulation to create reasonable approximations of phenomena. We suggest that there is a need to integrate these approaches to support students' understanding of the relationship between samples and population, and their reasoning with statistical models and modeling in the context of informal statistical inference (ISI).

In the next section we review the relevant theoretical background of modeling in formal and informal statistical inference. In section three, to explain the research method, we initially describe the research question and the participants. Next, we present our pedagogical approach that guided the design and analysis of the experimental tasks. Finally, we describe our data analysis method. In section four, to put the results of the study in context, we present the entire learning trajectory of the study and briefly describe what happened before the focused episodes of this article. Section five provides the main results of this study according to a conceptual framework that emerged as a result of the data analysis. In section six we discuss our insights regarding the chronological order of our conceptual framework dimensions throughout the learning trajectory. Theoretical implications and limitations are also discussed.

2. THEORETICAL BACKGROUND

Informal Statistical Inference (ISI) and Informal Inferential Reasoning (IIR) have recently become a focus of research (Makar, Bakker, & Ben-Zvi, 2011; Pratt & Ainley, 2008). The goal is to give students, even at a relatively young age, a sense of the power of drawing reliable statistical inferences from samples, particularly due to the fact that statistical inference is challenging for most students (Garfield & Ben-Zvi, 2008). ISI is a data-based generalization that includes consideration of uncertainty and does not involve formal procedures (Makar & Rubin, 2009). IIR consists of the reasoning processes that lead to the formulation of ISIs, which includes "the cognitive activities involved in informally drawing conclusions or making predictions about 'some wider universe' from patterns, representations, statistical measures and statistical models of random samples, while attending to the strengths and limitations of the sampling and the drawn inferences" (Ben-Zvi, Gil, & Apel, 2007, p. 2).

Understanding the logic behind ISIs includes 'juggling' several ideas, such as random sampling, sampling variability, and the relationship between samples and populations. However, students can hold two contradictory ideas about these relationships: 1) sampling representativeness: the expectation that a sample taken from a population will have characteristics similar to that population; and 2) sampling variability: the expectation that different samples taken from a population vary from each other and do not match the population (Rubin, Bruce, & Tenney, 1990). Although students can possess portions of these contradictory ideas, they may not understand the integration between them. Rubin et al. (1990) showed that senior high school students do not integrate these two ideas during their reasoning with distributions of sample outcomes, but instead focus on one idea at a time depending on the given task. To integrate these contradicting ideas, students need to envision a process of repeated sampling and understand "that the values of a statistic are distributed somehow with a range of possibilities" (Thompson, Liu, & Saldanha, 2007, p. 209). One way to achieve this goal is by modeling real-world situations and simulating models by drawing many random samples (Garfield et al., 2012).

Mathematics education and science education advocate modeling approaches because model-based reasoning can serve as a bridge that facilitates the shift from personal, intuitive knowledge to a more mathematical and scientific understanding of the world (Lehrer, Horvath, & Schauble, 1994). When students are engaged in constructing their own models they can develop conceptual understanding through a repetitive process by which they continuously construct data from phenomena and reason about these data. Indeed, the generation, testing, and revision of models are at the very heart of what it means to think statistically (Lehrer & Romberg, 1996).

One common approach to statistical modeling activities, the Exploratory Data Analysis (EDA) approach, involves students in scientific inquiries in such a way that they create surveys to study a question of interest (Ben-Zvi, 2006; Makar et al., 2011; Makar & Rubin, 2009; Pfannkuch, 2006; Tukey, 1977). Using survey construction, students resolve modeling issues such as the translation of real questions into survey questions and statistical research (Lehrer & Romberg, 1996), as well as representations of data used to persuade others. Another approach, the probability-based approach to modeling activities, emphasizes how probability is used by statisticians in problem solving. For example, in a study done by Konold, Harradine, and Kazak (2007), students built models using computer-based simulations to create reasonable approximations of phenomena, taking into account signal and noise. The first approach (the EDA approach) lacks probabilistic considerations that are important for understanding the relationship between samples and populations. The second approach (the probability approach) lacks aspects of an authentic data exploration such as posing a research question. Our suggested *integrated modeling approach* (IMA, described below) integrates these two approaches with modeling-based instruction.

3. METHOD

This research is part of *Connections* (2005-2015), a longitudinal design and research project created in order to develop an inquiry-based and technology-enhanced environment for learning statistics in grades 4–6, and to study students' emergent statistical reasoning. We focus here on the question, *how can students' articulations of models and modeling emerge while making ISIs*?

To address this question, design-based research was implemented (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Two iterations of interventions were conducted in a primary school in northern Israel according to a pedagogical integrated modeling approach (IMA) learning trajectory that we designed to help students understand the relationship between samples and their population. During a five-week intervention, two pairs of seventh grade students, who had previously participated in the EDA activities of the *Connections* Project, experienced the IMA learning trajectory. Using the innovative software TinkerPlots2 (Konold & Miller, 2011), the activities emphasized reasoning with samples, and sampling in making ISIs. Based on insights from this first intervention, several significant modifications and improvements were made in the design of the IMA learning trajectory. A second six-week intervention was conducted the following year with three pairs of sixth grade students from the same school who had participated in the EDA activities of the *Connections* Project in fifth grade.

In this article, we describe the work of one pair of students in the second iteration. They were chosen because our data analysis showed that they held conflicting notions regarding the issue of representativeness of random samples. While one argued that a random sample was likely to resemble the population, the other one argued that during random sampling any result was feasible and that it was impossible to predict the occurrence of a random sample result. These conflicting arguments are important elements in reasoning with sampling and modeling in the IMA learning trajectory, and assist us in categorizing articulations of models and modeling. In this article we study how the students built hypothetical probabilistic models of their explored population and compared these models with generated data from the Sampler in TinkerPlots2. The Sampler is a probabilistic simulator that can be used to model probabilistic processes and generate data. Learners can build a data distribution of a population and draw random samples from this population in an animated and visual way. The students also compared real collected data and generated data from these models. We used an interpretive micro-analysis (Meira, 1998), a microgenetic method (Chinn & Sherin, 2014), to analyze students' articulations of statistical models and modeling.

3.1. PARTICIPANTS

This study involved one pair of students (Grade 6, age 12), Ohad and Ido, from a primary school in northern Israel. It is a school that focuses on science and the environment and utilizes curricula that present a holistic image of the natural world to the students. Ohad and Ido were selected due to their advanced communication and analytical skills that provided a window to their statistical reasoning. They had participated in *Connections* activities in fifth grade where they collected and investigated data about their peers, using TinkerPlots1 (Konold & Miller, 2005), and were gradually introduced to samples of increasing size to support their reasoning with ISI, as well as sampling following the growing samples heuristic (Ben-Zvi, Aridor, Makar, & Bakker, 2012). According to the growing samples heuristic, students explore small data sets to infer about a wider set of data. They are gradually given more data and asked what can be inferred regarding the bigger sample or the entire population. Therefore, by the teacher's "what-if" questions, students learn about the limitations of inference (Bakker, 2004, 2007; Ben-Zvi, 2006; Konold & Pollatsek, 2002).

3.2. THE SETTING

The IMA was developed to guide the design and analysis of the experimental tasks, help us deepen students' reasoning with sampling and modeling when making ISIs, and guide the evaluation of this reasoning.

The IMA rationale The *Connections* project was based for many years on the EDA pedagogic approach toward ISI. Students were drawing ISIs from real samples following the statistical inquiry cycle (the left cycle in Figure 1). To foster students' appreciation of the power of inferences, we added a model-based perspective (the right cycle in Figure 1) to the EDA approach. During this inquiry cycle students build a model (a probability distribution) for an explored (hypothetical) population, and produce data of random generated samples from their model. Analyzing generated random samples and comparing them with the model, students can learn about the relationships between samples and populations. The aim of the right cycle is to enable students to explore key statistical ideas, such as sample–population relationships and sampling variability.

The IMA aims to assist students in making connections between those two cycles. The students move from the left cycle to the right cycle to learn about the behavior of random samples. This move is initially facilitated by a guided activity in which the conclusion becomes a hypothesis. Subsequently, the students use the insights from the right cycle about sample–population relationships to improve their conclusions in the left cycle which is performed by increasing sample size or explaining their confidence level in the conclusions.

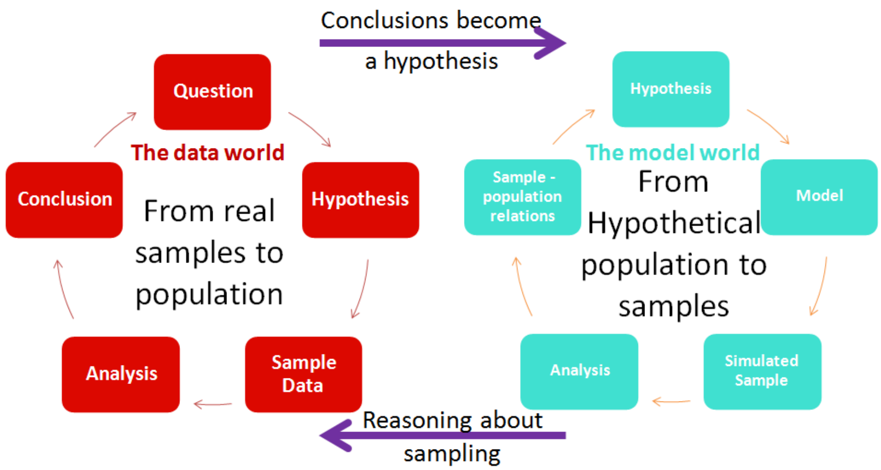

Figure 1. The rationale of the integrated modeling approach (IMA)

The IMA in detail The IMA is comprised of data and model worlds to help students learn about the relationship between sample and population, as well as model and modeling. In the data world (red and dotted line in Figure 2), students collect a real sample using a random sampling process to study a particular phenomenon in the population. In this world, students choose a research theme, pose questions, select attributes, collect, clean and analyze data, make informal inferences about a population, and express their level of confidence in the data. In doing so, students begin to create statistical models from real-world phenomena by shifting from informal questions to statistical ones. However, they may not account for probabilistic considerations, such as the chance variability that stems from the random sampling process.

In the model world (blue dashed line in Figure 2), the students build a model (probability distribution) for an explored (hypothetical) population and generate random samples of data from this model. They pay attention to the model and the random process that generates the sample outcomes from this model. Due to randomness, the details vary from sample to sample, but the variability is controlled. Given a certain distribution of the population, the likelihood of specific results can be estimated.

In the IMA, students iteratively create connections between the two worlds (purple continuous line in Figure 2) by working on the same problem context in both worlds and with the researchers' guiding questions as support. These guiding questions include questions about the minimal sample size needed to draw conclusions about the population with high confidence level, or "what if" questions on optional real data results while exploring model-generated random samples. Being engaged in an authentic inquiry, students keep building, revising, and explaining models of phenomena before and after collecting real data, and simulating their models to verify the data collected in relation to the model and to real sample data.

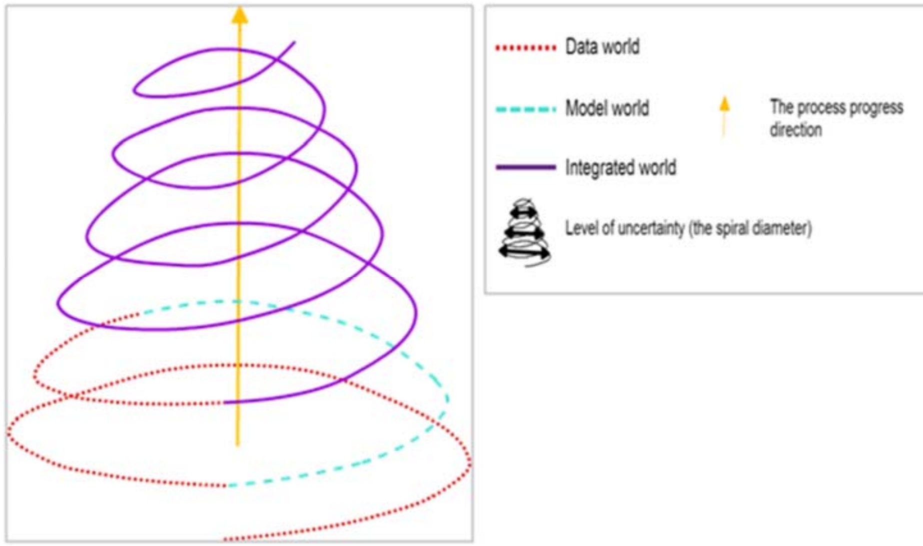

Figure 2. The integrated modeling approach (IMA) learning process

3.3. DATA ANALYSIS

We performed a retrospective analysis after each session (to re-direct the next session) and following completion of the entire teaching experiment. Empirical data collection included students' responses and gestures (captured using *Camtasia*), researchers' observations, and students' artifacts (e.g., their data representations). All student verbalizations were carefully transcribed. They were all translated to English because they were originally said or written in Hebrew. As part of the micro-analytic method, we closely examined the meaning of every word to make sure the translation was as close as possible to the original contributor's intention. Interpretive micro-analysis (e.g., Meira, 1998), a microgenetic method (Chinn & Sherin, 2014), was used to analyze the data. It is a systematic, qualitative, and detailed analysis of the transcripts that takes into account verbal, gestural, and symbolic actions within the situation in which they occurred.

The goal of the data analysis was to interpret student articulations of statistical models and modeling and to illuminate the students' processes of learning while they built TinkerPlots2 models and drew from them random samples based on their conjectures. The data analysis was performed in three stages. During the first stage, the primary researcher (first co-author) transcribed all the *Camtasia* movies while searching for segments that were relevant to the research question. In this stage the exploration allowed the data to tell the story. At the end of this stage the researcher had a collection of segments for in-depth analysis. In the second stage, the primary researcher meticulously transcribed the relevant segments and analyzed them according to the principles of the microgenetic method. In order to interpret students' articulations and emergent ideas, the researcher looked at the "history" of the emergent ideas, i.e., their previous appearance, origin (student or researcher) and context. The researcher formulated temporary localized assumptions about students' emergent reasoning and validated them by exploring what happened before and after. Repeated students' performances and similar verbalizations strengthened these assumptions and contradictions weakened them.

In the third stage, triangulation was performed by a small group of expert statistics education researchers (including the second author) and novice peers. The primary researcher presented the chosen segments to this group without her opinion or interpretation to enable unbiased interpretations by these experts. These researchers discussed, presented, advanced, and/or rejected hypotheses, interpretations, and inferences about the students' reasoning and articulations. During these triangulation meetings, hypotheses that were posed by the researchers were advanced or rejected, until a consensus was reached. The group of researchers looked for agreed interpretations, but also focused on disagreements and examined contradicting interpretations, Triangulation was attained only after multiple sources of data validated a specific result (Schoenfeld, 2007), achieving "trustworthiness" (Lincoln & Guba, 1985).

4. PUTTING RESULTS IN CONTEXT

4.1. THE LEARNING TRAJECTORY

The first co-author observed and guided the students during eight activities that lasted about 15 hours (Table 1). In the data world, the students planned a statistical investigation, which consisted of choosing a research theme and population, posing research and survey questions, formulating a conjecture, and deciding upon the sampling method and sample size (Activity 1). Due to the fact that some of the students' sampling methods in Activity 1 were biased, we added an activity to explore the meaning of biased sampling versus random sampling (Activity 2). This activity also served as the first interaction with the model world and the idea of sampling distribution. As a result of Activity 2, students refined their sampling methods, reformulated their conjectures, and implemented a sample survey in their school (Activity 3). Using TinkerPlots2, the students explored their real sample data to make conclusions about the population in response to their research question (Activity 4). In their second encounter with the model world, they used the Sampler to build a hypothetical model for the population distribution. Based on their conjecture, they drew random samples from this model, compared them to the model and their real sample data, and explored sampling distributions (Activity 5, which is the focus of the current study). To encourage them to examine the connections between the worlds, they were asked "what if" questions on real data results while exploring generated random samples.

Because students found it difficult to connect between generated random samples and the real sample, they were given a sixth activity in which they were asked to use the TinkerPlots2 Sampler to draw many random samples from a hidden sampler in order to make ISIs. (Activity 6, see Manor Braham, Ben-Zvi, & Aridor, 2014). A hidden sampler, one of the TinkerPlots2 options, hides some attributes' distributions of a model. The hidden sampler can hide a representation of the distribution of a particular attribute of a model. Thus, the hidden model can stand in for a hidden population. Including a hidden population in this activity causes the sample-population relationship to resemble the relationship in real-life situations in which the population is unknown. Then the students returned to their own investigation, explored different sample sizes from their TinkerPlots2 model to compare between them and decide about the minimal sample size needed to draw conclusions about the population. According to the results of their chosen sample size, the students collected more real data (Activity 7). Finally, they simultaneously explored data and models in the two worlds by examining the real larger-sample data in relation to their conclusions in the model world. For example, in their conclusions about the population from the real larger-sample data, they used likelihood estimation with the given sample size and under the hypothesis of a certain population distribution (Activity 8).

Table 1. The actual IMA learning trajectory

Act. No.	Activity Title	Session's Themes	Statistical Ideas and Concepts	Length (min)
1	Learning about teenagers (part 1)	Session 1.1: Choose a research theme and a research population, pose research and survey questions, choose sampling method and sample size	Considerations in formulating a research question, choosing sampling method, random sampling, and sample size (absolute size vs. population percentage)	45
		Session 1.2: Formulate a hypothesis and express confidence level in the results	Research hypothesis (verbal and visual expressions of flow chart) and confidence level	45
2	Random sampling (The Strings Activity)	Estimate the proportion of short strings in a bag that contains long and short strings by drawing strings from the bag with closed eyes. Explore the proportion of blue beads in samples of beads drawn from a bag that contains an equal number of blue and red beads	Random sampling vs. biased sampling, sampling with or without replacement, sample size, confidence level, and visual expression of sampling distribution	60
3	Learning about teenagers (part 2)	Refine sampling method, reformulate conjectures, and implement a survey in the school	Sampling methods (random vs. biased, sample size and stratified sampling), research hypothesis (verbal and visual expressions of flow chart), and confidence level	25
4	Learning about teenagers' art shows preferences (part 1)	Explore real sample data, draw conclusions about the population, and estimate level of confidence in the conclusions	Distribution, variability, context and data, sample size, and level of confidence in conclusions about the population	65
5	Modeling art shows preferences among teenagers (part 1)	Session 5.1: Build a TP2 model of relationships between several attributes in the population	Presentation of frequencies with numbers or with percentages A model as a reduced finite population Dependency between attributes	80
		Session 5.2: Draw and explore samples of different sizes from the model	Comparison of samples (from the model to the model) Comparison of samples from the model to real sample Sampling variability	120
		Session 5.3: Compare samples from the model and explore sampling distribution	Sampling distribution, probability of sample results, and degree of inaccuracy	105
6	Teenagers in social networks	Draw ISI by exploring samples of a hidden model	Relation between sample size and sampling variability, control and quantify sampling variability	120
7	Modeling art shows	Explore samples of different sizes from the model and	Relationship between sample size and sampling variability.	125

	preferences among teenagers (part 2)	decide about the minimal sample size needed to draw conclusions	The sampling distribution range and variability, sample size, and frequency of sample results	
8	Learning about art shows preferences among teenagers (part 2)	Explore real sample data in relation to a sampling distribution and draw ISI	As the population is larger, a certain percentage of the population will be more similar to the population Informal hypothesis testing	90
Total time (minutes):				880

4.2. WHAT HAPPENED BEFORE THE FOCUSED EPISODES?

In the following episodes, which took place predominantly during Activity 5, we focused on students' initial experiences with models to trace: 1) what types of models they built before and after working with the TinkerPlots2 Sampler; 2) how they explained the models of the population; 3) what changed in their modeling articulations as they began to generate random samples from their TinkerPlots2 models; and 4) what changed in their modeling articulations as they began to compare between generated and real random samples, or between generated random samples and their model.

To illustrate our case, this article focuses on one pair of students' articulations of statistical models and modeling during Activity 5. We also use examples from Activities 3 and 4 to illustrate the shift in the students' reasoning with statistical models and modeling. To put these episodes in context, in the next few sections we briefly describe what occurred prior to these activities.

Activity 1. Learning about Teenagers: From a sample to population In the first activity, we asked the students to plan a research study on a subject that is relevant to teenagers and interested them. Ohad and Ido decided to study art show preferences (music concerts, theatre, etc.) among fourth- to ninth-grade students in their school. Their goal was to be able to offer a better and more interesting choice of shows that were brought to their school. They formulated 11 survey questions to study what attributes may be related to the favorite shows. In planning their sampling method, the students were asked not to interrupt the school schedule and take out only a few students for their survey. Therefore, they were forced to decide on a minimal sample size, and were asked to explain how they chose sample students in order to draw reliable inferences about their school. Ido and Ohad suggested randomly drawing ten students from each class in fourth, sixth, seventh, and ninth grades. They did not choose fifth and eighth grades because they argued that they were well-represented by the other grades.

Activity 2. The Strings Activity: Random sampling The researchers added the Strings Activity to impart the meaning of biased versus random sampling. The students were asked to estimate the proportion of short strings in a bag of long and short strings by drawing with closed eyes a few strings with replacement. Students underestimated the proportion due to the biased sampling method (long strings are more likely to be selected). After they were told the real proportion of long strings in the bag (50%), they discussed the biased sampling method and how to avoid it. They experienced and discussed an unbiased sampling method using equal-sized red and blue beads instead of strings.

During the Strings Activity, Ido claimed that random sampling is the best sampling method, but in response to the researcher's question, he estimated his confidence level that a random sample represented the population as 50%. In the case of red and blue beads, Ido said that the proportion of red beads could be anything because every result had the same chance. In contrast, Ohad argued that there was a greater chance to get a result that was equal to the real proportion of red beads in the bag. Finally, Ido and Ohad decided to collect data randomly from 12 cases in each grade. They also prepared and implemented an online questionnaire that included their show preferences survey questions.

5. MAIN RESULTS

In this section, we provide an analysis of Ido and Ohad's reasoning with statistical models and modeling. We present the scenes not in chronological order, but according to the dimensions of a framework that emerged as a result of the data analysis (Figure 3).

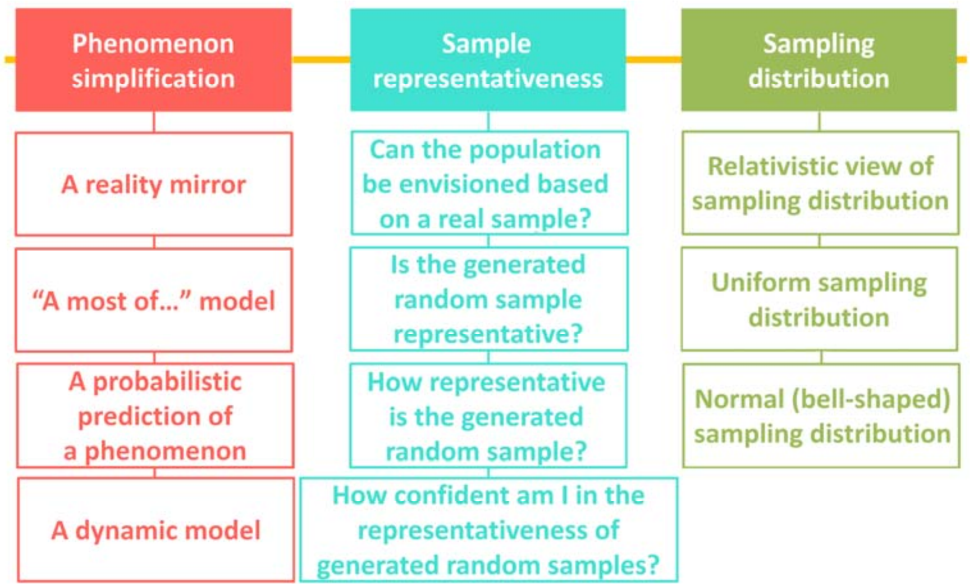

Figure 3. A framework of reasoning with models and modeling

We found three dimensions in the students reasoning with statistical models and modeling (Figure 3):

1. *Reasoning with phenomenon simplification.* This dimension includes different types of representations that the students used to simplify a phenomenon. In the data world, the students described their hypothesis regarding the phenomenon verbally and visually before and after the data collection. In the model world, using theTinkerPlots2 Sampler, the students built and ran a model that included three attributes, according to their hypothesis. The students also refined the model as a result of their simulations.
2. *Reasoning with sample representativeness.* This dimension includes various issues that the students explored regarding sample representativeness. In the data world, students explained how they compared the real sample data to their image of the population in order to decide how confident they were in the real sample data. In the model world, students compared generated random samples from the model to the model itself. This motivated them to study how to decide whether, and to what

extent, the samples and the model were similar and how confident they could be in the sample representativeness.

3. *Reasoning with sampling distributions.* This dimension includes different students' views of sampling distributions. In the model world, trying to describe a sampling distribution, the students discussed probable sample results that may stem from different types of population distributions.

5.1. REASONING WITH PHENOMENON SIMPLIFICATION

Ido and Ohad were asked to draw a representation that described their hypothesis regarding the phenomenon of art show preferences among teenagers in the population before and after they collected the data. We identified four elements of phenomenon simplification in the students' reasoning: a reality mirror, a "most of" model, a probabilistic prediction of a phenomenon, and a dynamic model (Figure 3).

A reality mirror is an exact description of the phenomenon with a detailed depiction of the attributes and their relationships. It reflects a deterministic view of the phenomenon with no uncertainty, and was typical in the students' preliminary investigation stages. For example, when students were asked to make conclusions about the population, they argued that the population would be identical to the sample data. The "reality mirror" view diminished quickly at the beginning of Activity 5 when they compared random simulated samples to the model from which the samples were drawn and became aware of sampling variability.

A "most of" model view describes the phenomenon by providing details on most of the population elements. A "most of" model includes a description of the signals in the phenomenon with a small reference to the noise. Ido and Ohad presented a "most of" model view during Activity 3 when they were asked to describe their conjecture about the population verbally and visually before they collected real data.

Because the students were accustomed to drawing flowcharts in science and computer lessons in their school, the researcher asked them to draw a flowchart of their conjecture prior to using the Sampler. The flowchart they drew (Figure 4) expressed their first conjecture. It contained four attributes: gender, grade, favorite show type, and preferred manner of watching a show (on a screen or attending the show). Even though they did not explicitly mention the attribute names, they drew the values of each attribute in a different row and drew arrows to show the relationship between the attribute values. The last three arrows at the bottom of the chart that point to "go to a show" represent the fourth attribute and the conjecture that all the students will prefer to go to a show (Figure 4).

Describing their conjecture, Ido said: "Girls in grades four to six will prefer to go to a singing concert." When they were asked whether they assumed that *all* girls prefer singing, Ido explained that, "if you are a girl in grades four to six, there is a greater chance that you prefer a singing concert." Although the first quote shows that they meant "the most" rather than all the students, the second quote actually depicts that their model was not deterministic describing most of the elements in the phenomenon with small variability.

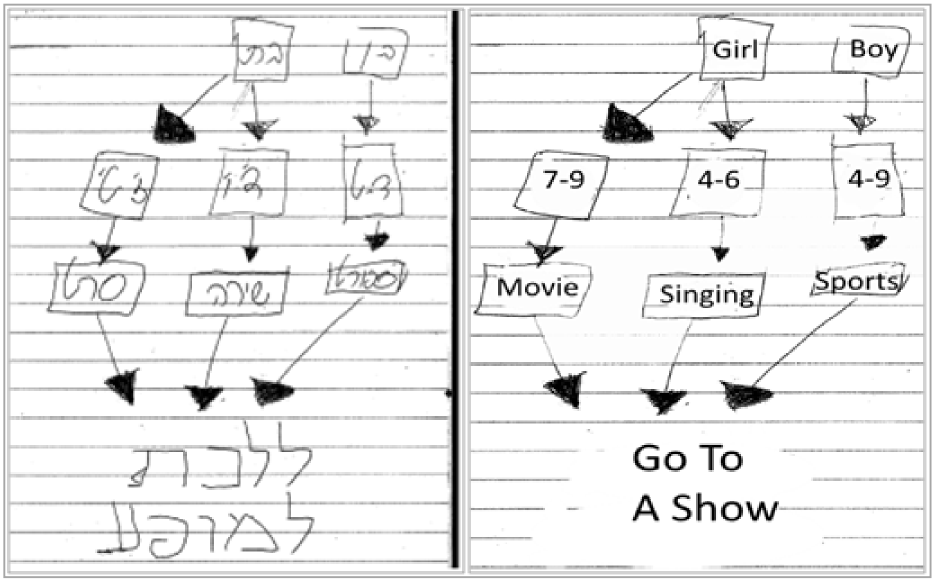

Figure 4. Ido and Ohad's conjecture regarding the relations between favorite show types, gender and grade (translated to English on the right)

A probabilistic prediction of a phenomenon In Activity 4, the students collected and explored real sample data (N=48), which they used to create a graph (Figure 5) and learn how the attribute "favorite type of show" is distributed by gender and grade. Ido and Ohad were surprised to see that in contrast to their conjecture, very few children preferred music shows and therefore they expressed low confidence in the sample results.

Figure 5. A real sample of favorite show types organized by gender and colored by grade

In Activity 5, the students were asked to learn about sample-population relationships by using the Sampler to build a model that described their conjecture. A model in TinkerPlots2 is built by several chosen attributes and the relationships among them. The students can represent the attribute distribution in the Sampler by one of optional devices (e.g., spinner, curve). The students' model (Figure 6) contained three attributes (the same as the first three attributes in "most of model", Figure 4) and the relationship among them.

After they finished building the model (Figure 6), they explained it:

589	Ido:	We actually did a flow chart of our guesses. We wrote what will happen, in our opinion, in the population that we are examining.
607	Ido:	We made here a flowchart of what will happen. If you are a male and you are [learning] in na na na [a certain] grade, then between fourth to this [sixth], the chances that you will like it [a certain show] are this and that, that is in our opinion.
608	Ohad:	This is not the chances that you will like it.
610	Ohad:	This is what we assume on the population.
611	Ido:	This is the way in our opinion it will be distributed. And if you are between seventh to ninth [grade] then it will be distributed in this way, the population.
612	Ido:	Of course it [the model, Figure 6] was not exact because it could be completely different, but in our opinion it will resemble [the population]. These are our assumptions. We don't know.
613	Ido:	We are basing ourselves here with a relatively high level [of confidence] on what we actually took from here [the real sample] and on general knowledge.

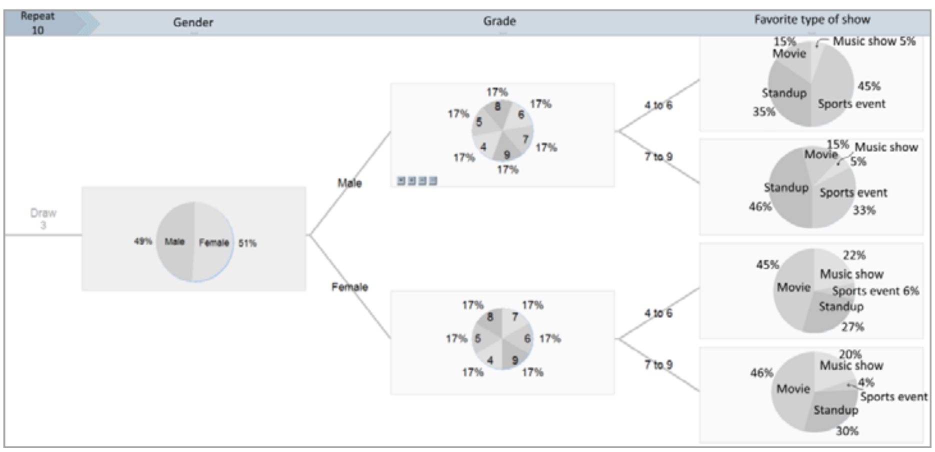

Figure 6. Ido and Ohad's first model in TinkerPlots2

Ido mentioned two different purposes for building the model (Figure 6): 1) describe their hypothesis about the population [line 589]; and 2) predict the chance of events happening [line 607]. The first purpose of the model was to describe: The model is a descriptive simplification of a phenomenon that depicts relative frequencies of specific values in the population. The second purpose of the model was to predict: The model is a predictive simplification of a phenomenon that predicts the chance that certain values occur in the population.

A dynamic model During Activity 5, after generating small random samples from the model (Figure 6), and comparing them to that model to examine the representativeness of the random samples, the students changed the devices used to represent the first two attributes, gender and grade, from spinners to counters (Figure 7).

They explained that they wanted the Sampler to generate samples in the same way they had drawn their real random sample because otherwise “it would not show in a good way what happened [the sampling method] at school” [Ido, Line 887]. In other words, they wanted the Sampler to draw an equal number of males and females and an equal number of students from each grade. We assume that this request may be an indication of the connection they began to create between sampling in the data and the model worlds. This request also hints at a dynamic view of a model with which one can generate data in order to examine the representativeness of random samples.

In this section we presented four depictions of models that the students used to simplify a phenomenon. The four depictions indicate a gradual shift from a deterministic to a probabilistic to a dynamic view of a model.

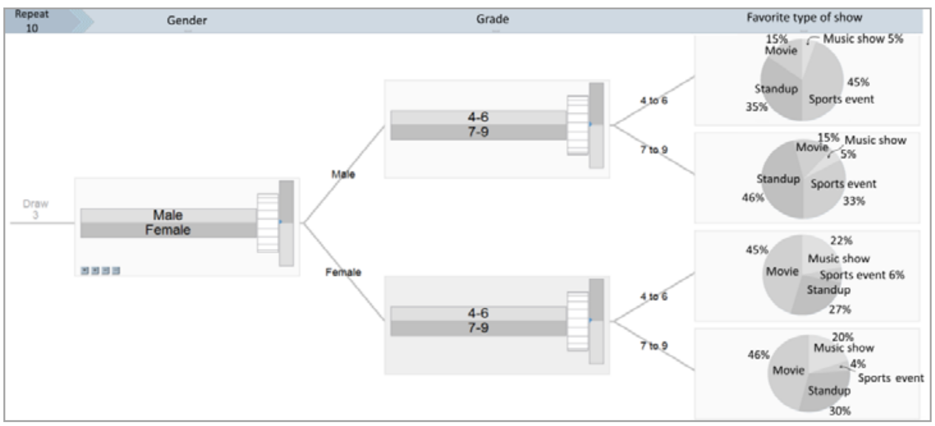

Figure 7. Ido and Ohad’s second model in TinkerPlots2

5.2. REASONING WITH SAMPLE REPRESENTATIVENESS

We identified four main questions explored by the students regarding sample representativeness: Can the population be envisioned based on a real sample? Is the generated sample representative? How representative is the generated sample? How confident am I in the representativeness of the generated sample?

Can the population be envisioned based on a real sample? The following episode happened during Activity 4 (See Table 1). When the students analyzed their real sample data (N=48, Figure 5) they were surprised to see that in contrast to their conjecture, most students did not prefer music shows. Looking for an explanation, they decided to explore students’ favorite music. They created a graph of a new attribute “favorite music show types” organized by gender and ordered by grade (Figure 8).

While exploring this graph (Figure 8), the students said that they could infer about the population only from unequivocal evidence in the sample. Noticing that all fourth grade males do not like music shows, they concluded that most of the fourth grade males in the population did not like music shows. However, because they could not find further unequivocal evidence in the sample, they stated that their confidence level in their conclusions was only 50% or 60%.

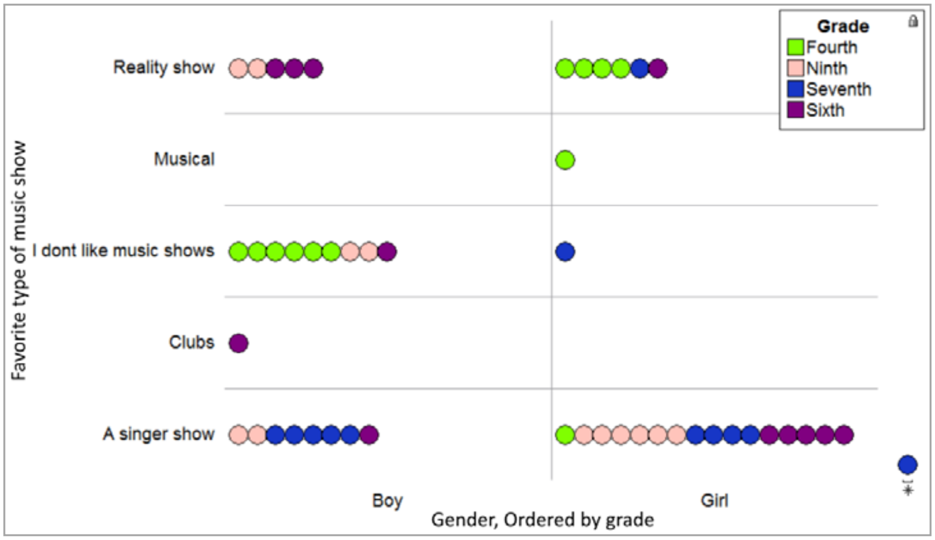

Figure 8. A real sample of favorite type of music show organized by gender and colored by grade

At this point we initiated learning about sample-population relationships (Activity 5). Initially, they were asked to think about simulations in their daily life. Then they were asked to think about a simulation they can use to check whether a sample can represent a population. In response Ido said: "We can simply multiply it [the sample size] several times until it gets to the number of students that is exactly the number [of the size of the population]… but with the same ratio of students. We multiply it [the sample] in our head and we check if this [the multiplied sample] matches [our conjecture]. This [process] works if the sample was exact." Ido described a decision rule about how to judge the representativeness of a sample. He explained that they "multiplied" their sample data to see how the population results would look if the sample was accurate and compared the multiplied sample to their conjecture.

Is the generated random sample representative? The following episode took place during Activity 5. The students drew small random samples from their TinkerPlots2 model (Figure 6) and examined sampling representativeness by comparing the random samples results and the model. The sample in Figure 9 was the first sample (size 10) that they drew from the model.

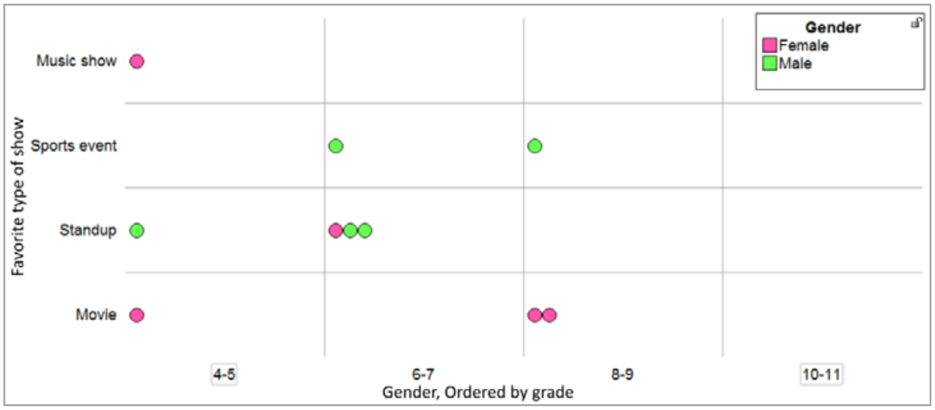

Figure 9. A first random sample of size 10 taken from the first model (Figure 6) in TinkerPlots2

Ido interpreted the sample graph (Figure 9): "So this [the generated sample] looks as we expected. Many [girls] came out [were chosen] in the movie as we really expected. But we also gave chances [set chances in the model] to sporting event [for girls]. But sporting event didn't come out [was not chosen] because we gave it [set in the model] very low chances." They noticed that the sample results were similar to the model in categories with rather large relative frequency. Furthermore, in comparison to their initial description of the model (see the third sub subsection in section 5.1), their reference to the relative frequencies in the model (Figure 6) acting as the chance of certain results occurring made more sense to them. Referring to relative frequencies as chances helped them explain why the sample represented the model.

How representative is the simulated random sample? After the students had explored simulated samples of size 10, they asked to increase the sample size to 48. They drew a sample of 48 cases (Figure 10) from the model in Figure 7 and compared between the generated sample and the model. They organized the graph of the generated sample (Figure 10) in such a way that enabled them examine the percentages of the attribute "favorite type of show", for each category, within a certain gender and grade. After comparing several generated samples to the model, they were asked, "To what extent do you think that these random samples represent the model?" Ohad responded: "They represent the model by 50%" and explained: "Half [of the percentages in the generated sample in Figure 10] are similar [to the appropriate percentages in the model in Figure 7]". Ido added: "But like, it depends upon what you consider similar. But it's not that I have a definition [decision rule]. Every [percentage in Figure 10] for example [that] is in a range of 5% above, 5% below [the appropriate percentages in the model in Figure 7] is considered as similar… and I am examining how many are similar and how many are not similar." As seen in the quote above, Ido found a way to quantify the level of representativeness of the generated random sample.

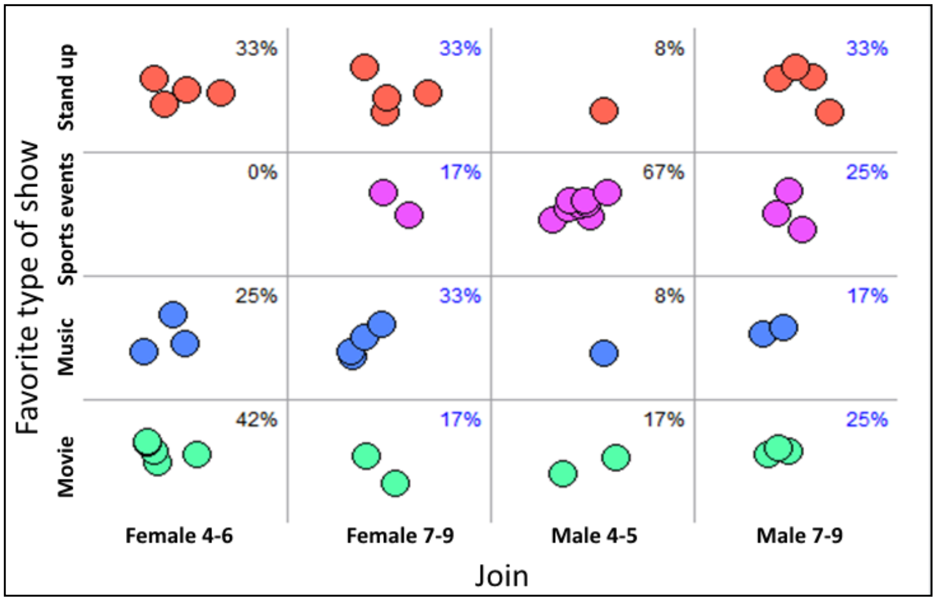

Figure 10. A first random sample of size 48 taken from the second model (Figure 7) in TinkerPlots2

How confident am I in the representativeness of generated random samples? After the students tried to compare among four samples of size 48 to decide how confident they were in the representativeness of the generated random samples, the researcher showed the students how to collect repeated sample measures to create a sampling distribution. They collected the percentage of males whose favorite type of shows was standup (%STANDUP in abbreviation). They collected measures from 100 samples and tried to define what can be considered a "good" result by organizing the sampling distribution accordingly (Figure 11).

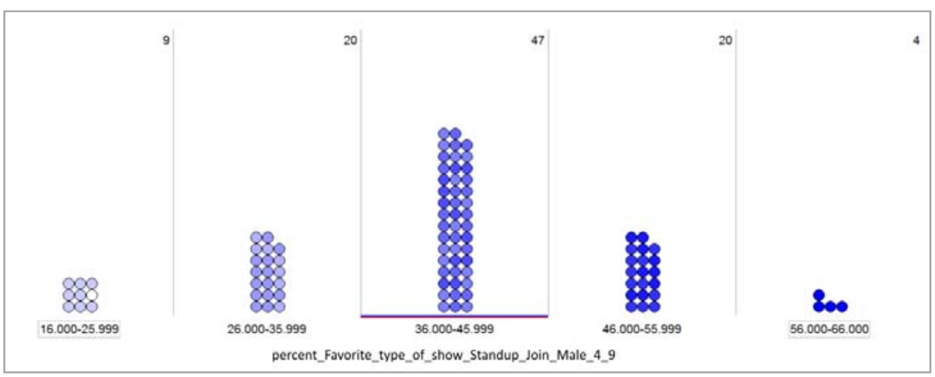

Figure 11. A %STANDUP sampling distribution, sample size 48

1135 Ido: Let's call these [results in 36-45.999], the good results [meaning close enough to the real value in the model] and they are almost 50%.

1136 Ohad: Yes.

1137 Ido: It [47%] is very good.

1139 Ido: We shall call these two [the results in the intervals 26-35.999 and 46-55.999] reasonable.
1140 Int.: Near the good [results]?
1141 Ido: Yes.
1142 Ido: And let's call these [the results in the intervals 16-25.999 and 56-66] bad [results]. We see that we have 47% here [good results], 40% [reasonable results] and only 13% [bad results].

The students defined good, reasonable and bad results. They noticed that there were many more good and reasonable results than bad results (Lines 1135, 1137, and 1142). It seems that exploring this sampling distribution (Figure 11) increased their level of confidence in sampling representativeness.

In the questions asked by the students in this dimension, we can see a gradual growth in their articulation of sampling representativeness. In the first stage, the students used an intermediate representation to judge the sample representativeness. They examined representativeness by comparing multiplied sample results to their conjecture. In the second stage, they examined sampling representativeness by comparing samples generated by the model to the model. By connecting the incidence of results in the sample to their relative frequency in the model, the students increased their confidence level in their random samples. In the third stage, they examined the influence of sample size on sampling variability and invented an initial method to quantify the accuracy of sample results (the difference between sample result statistics and the model parameter) and sampling representativeness. During the fourth stage, they refined this method to discover the chances of "bad," "reasonable" and "good" results to quantify their level of confidence in samples of size 48.

5.3. REASONING WITH SAMPLING DISTRIBUTIONS

During their learning progression, the students discussed the issue of repeated samples by "what if" questions or by exploring repeated samples. In the following section we discuss the students' articulation about signal, noise, and shape of sampling distributions. We identified three different views of sampling distributions held by the students: Relativistic view of sampling distributions, uniform sampling distribution, and normal (bell-shaped) sampling distribution. We use the term relativistic view of sampling distributions in the sense suggested by Ben-Zvi at al. (2012) to describe instances when students had no confidence in random samples, suggesting that any sample result could happen by chance.

Relativistic view of sampling distributions The following episode was taken from a scaffolding activity (see Activity 2, Table 1) that supported students' understanding of random versus biased samples. During one of the tasks in this activity, the students drew samples of size 10 from a bag that contained an equal number of red and white beads and discovered the percentage of red beads in each sample. After drawing six samples, the students documented the resulting percentage of red beads in each one of the samples. They were then asked to suggest what the results would be if they drew 100 samples of size 10.

Ido explained that "for every [result] there are equal chances. It [the samples results] will be completely randomly scattered. It could be that all [the results] will fall here [he pointed with his hand to the right side of an imaginary graph]. It could be that all [the results] will fall in the middle. It might happen that they will completely scatter." As Ido explained above, because any result was possible, he was unable to describe the graph and draw a suggested sampling distribution.

Later in the learning trajectory (Activity 5) when the students drew a second sample of size 10 from the TinkerPlots2 model (Figure 6), Ido reinforced his relativistic view about sampling distributions:

> A fourth–fifth grade was not chosen [in the random sample] (Figure 12). By chance… in principle this actually proves what I said. In situations where every [result] has equal chances [pointing at grades distribution in the model in Figure 6], anything can happen. What happened before just proves that I was right. No fourth-fifth grade [student] was selected at all.

Unlike his previous comment, Ido limited his opinion to uniform phenomenon, namely, if the distribution of the population was uniform, any result was possible.

Results of Sampler 1

	Gender	Grade	Favorite_type_of_show
1	Male	7	Movie
2	Female	9	Movie
3	Female	9	Standup
4	Female	7	Standup
5	Female	6	Music show
6	Male	7	Sports event
7	Male	9	Standup
8	Male	8	Sports event
9	Male	6	Movie
10	Female	9	Movie

Figure 12. A second random sample of size 10 taken from the first model (Figure 6) in TinkerPlots2

Uniform sampling distribution The following episode took place in Activity 5 (Table 1) after the students built their first TinkerPlots2 model (Figure 6) and before they generated random samples from the model. They explained how they decided whether to build their model (in Figure 6) based on their real sample or on their context knowledge. Pointing at the graph of their real sample (Figure 5), Ido gave an example of a situation in which he did not rely on the real sample result for building the model in the Sampler:

> So there are three boys [sixth grade students that prefer sporting events, Figure 5] out of six [sixth grade boys]. So the chance it [the real sample] represents half [of] all the 45 [sixth grade students in the population], half of them, I mean 22.5, one boy is divided, prefer sporting events, is small, since half is a lot.

To better understand Ido's view, the following discussion took place:

208	Int.:	And if I ask it [in the] opposite [way]? You know that there are half [in the population]. Okay? You have 22 students in the sixth grade population that like sporting events out of 44 [students in the population]. Now we drew a sample. What is the chance that we will get exactly such a result? That we get three out of the six students [in the sample]?
209	Ohad:	Fifty fifty.
210	Int.:	That means, there is a possibility that it will be zero (students who like sporting events). That is what I am asking.

211 Ido: It will be equal.

Ido's response to the interviewer's question reflected a different view of sampling distribution than the one he held previously. As we discussed above (see the first subsubsection in section 5.2) Ido used the "multiply method" to explain why the sample result was not representative. He multiplied the real sample result to see what the conclusions could look like in the population based on the sample. Because the result in the multiplied sample was not similar to his conjecture, he concluded that the sample result was not representative. In the current episode, the researcher asked the question in an opposite way: Given a certain situation of the population, what is the likelihood of specific results? Unlike Ido's previous answer, he now said that each sample result had the same chance to be selected, which depicted his view of a uniform sampling distribution of the proportion of students who prefer sporting events.

Normal (bell-shaped) sampling distribution Ohad also answered the interviewer's question as follows:

219	Int.:	If [we imagine that] there are 44 students in the population of sixth grade, and 22 of them like sporting events. If we drew a sample, what will we get there?
220	Ohad:	How many [children] in the sample?
221	Int.:	Like this [points to the real sample graph, Figure 5]. We took 12 students from [sixth] grade. Six boys and six girls. In your opinion, how many students of each group will be selected in the sample?
222	Ohad:	It is not possible, there isn't a large percentage that we won't see any boy in the sporting events [category].
225	Ido:	So still he [Ohad] assumes that in his opinion it [the sample result] will come out more similar to the real [result in the population].
226	Int.:	Is it more likely that it will be 3? More likely to be 2? 1? 0? Maybe 4?
227	Ohad:	Three
228	Int.:	Do you think it is more likely that it will be similar (to the population) than it will be 0? Or [than] it will be 6?
229	Ohad:	Yes.
230	Int.:	Why?
231	Ohad:	Because among the six [students] not all the children are the same. There is a possibility that it will be everybody, that all six will [like] sporting events, but then it means that we fell exactly on [got sample results of] the children that like sporting events… it is less likely that we will fall [will get results] only on [of] them. Then [the possibility] is that we will fall [will get results] on [of] 3, 2 or 4. Something in the middle.

Ohad described sample results as resembling a "bell-shaped distribution" and explained why he thought that it was more probable to get sample results that were close to the population parameter.

The students built a sampling distribution of the percentage of males that preferred stand-up comedy shows (%STANDUP in abbreviation) based on 100 samples of size 48 (Figure 13) drawn from their second TinkerPlots2 model (Figure 7). It was the first time they explored a sampling distribution with the TinkerPlots2 Sampler. Initially, they compared the mode of the sampling distribution to its appropriate parameter in their model and found that the values were very close. In the following discussion, the researcher attempted to learn what they understood about this sampling distribution (Figure 13).

1099 Int.: What do you say now if we take a sample size of 48? What can we learn from this graph [Figure 13]?

1100 Ido: These 100 samples came out very similar to this (parameter in the model). It just strengthens my feeling that, maybe not one sample, but if we examine several samples and then produce their mean, the chance it [the difference between the mean of the statistics and the parameter] would be closer is greater.

1101 Int.: Yes, but in reality you take one [real] sample. What can you see here from the graph [Figure 13] and can it help you somehow?

1102 Ido: One sample can fall here [point to an extreme result of 16% on the left side of the graph] which is not accurate or here [points to an extreme result of 58% on the right side of the graph], but it could be that it [the sample] will turn out to be very precise [close to the parameter in the model which is 40%], and the chance that it will be precise is bigger than the chance it will be imprecise. But still there is a chance that it [the sample] will be imprecise.

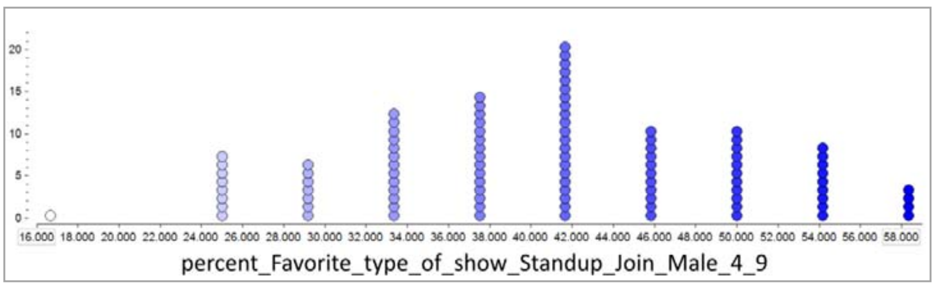

Figure 13. A %STANDUP sampling distribution of 100 samples of size 48

Ido's articulation depicts a change in his perception of sampling distributions. He began to realize that there was a greater chance to get a statistic that is similar to the parameter in the model than to get an extreme statistic that is different.

In summary of this dimension, we identified three perceptions of sampling distributions. The students' explorations in the model world enabled them to explore and refine their perceptions of sampling distributions. We were interested mostly in Ido's view of sampling distribution because he began with a relativistic viewpoint. After building a model with the Sampler, Ido presented a new insight. He distinguished between two situations, one in which the model was uniform and another in which the model was normal. In the first situation, Ido thought about repeated sample results in a relativistic way (any result can happen), and in the second situation, Ido envisioned a uniform sampling distribution (every statistic has the same chance of happening). In the last stage, after the students drew samples from the model and presented a sampling distribution of 100 samples, Ido seemed to embrace a normal view of sampling distribution. It was the first time he said that there was a greater chance for statistics that were similar to the parameter.

6. DISCUSSION

The main research question of this study was: how can students' articulations of models and modeling emerge while making ISIs? To answer this question we carefully examined Ido and Ohad's work in the IMA learning trajectory according to a framework for reasoning with models and modeling that emerged as a result of the data analysis. Describing the chronological order of the dimensions throughout the learning trajectory, we first discuss the relationship between the three dimensions of students' reasoning progression with models and modeling. Then, we examine how the students' views about modeling changed, expressed by their articulations, while they moved between the data and model worlds. Finally, we present implications of the framework and the limitations of this study.

6.1. A CHRONOLOGICAL ORDER OF THE FRAMEWORK DIMENSIONS

The framework dimensions are presented in a chronological order in Figure 14. The students entered the data world with a relativistic view of sampling distribution and a simplification of a "most of" model. After they engaged in data exploration, they thought of sample-population relationships as a multiplicative relationship. When they entered the model world, their understanding of the data converged with their context knowledge, and they described their first model in TinkerPlots2 as a probabilistic prediction of the phenomenon. However, they had two different views about this model. Whereas Ido described the model as a predictive simplification of a phenomenon that predicts the chances of values occurring in the population, Ohad disagreed and stated that the model described their hypothesis about the population distribution.

A meaningful change in the students' reasoning with models occurred when they refined their probabilistic prediction of the phenomenon model to a dynamic model. That is seen when the students asked to change the generated sampling method of the model according to the sampling method of the real sample. This refinement in their methodology implies that the students began to make connections between generated random samples from their model and the real sample they drew from the population. However, when looking at the chronological order of the dimensions, we can see that this change happened following their exploration of generated random sample representativeness. Before the students refined their model to a dynamic one, they drew repeated samples from the model and examined whether the generated random samples were representative by comparing them to the model. This may signify that the students drew an analogy between the mechanisms of random sampling in the model and the real world.

Looking at the chronological order of the dimensions, we can see two appearances of the "relativistic view of sampling distribution," in the data and in model worlds. The first manifestation occurred in the data world, with Ido's relativistic view of sampling distributions. Ido stated that in random sampling any result could happen and there is no way to predict the likelihood of results. The second manifestation occurred in the model world and was a result of a connection Ido had made between the worlds. Ido observed the generated random samples in the model world due to his desire to examine his relativistic view regarding sampling distribution, because it influenced his confidence level regarding the inferences that could be made by a real random sample in the data world. By drawing generated random samples in the model world, Ido realized that when sampling randomly from a certain model, "the chance that it [a sample statistic] will be precise is greater than the chance that it will be imprecise." This caused a change in his sampling distribution view.

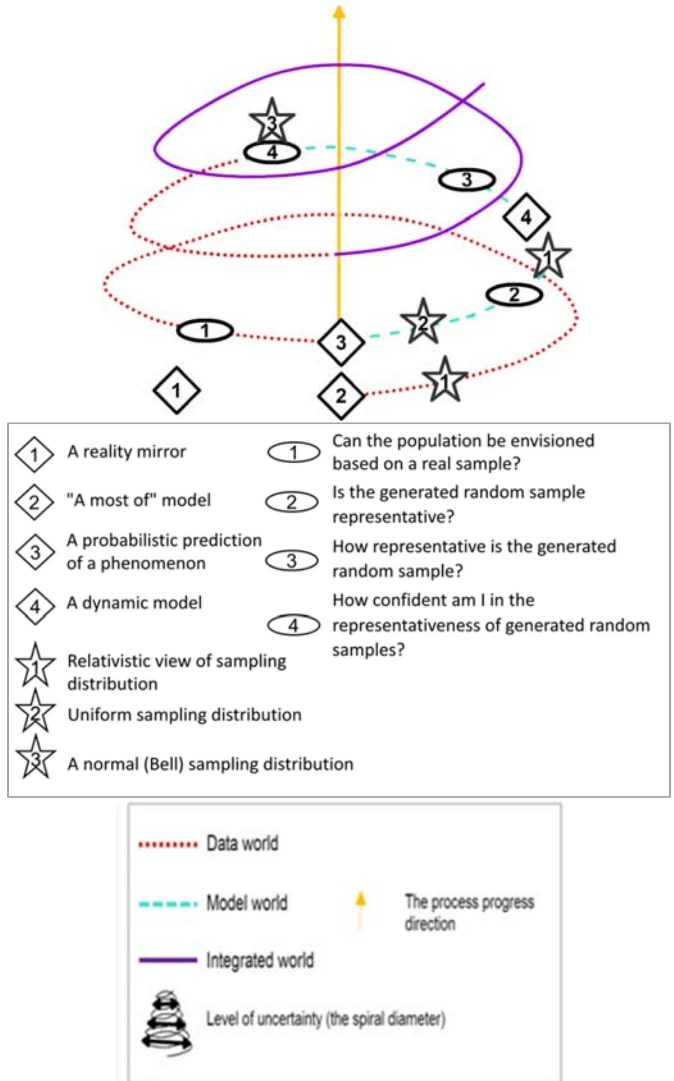

Figure 14. A chronological order of the framework dimensions' elements

6.2. DIFFERENT VIEWS ABOUT MODELS AND MODELING

We identified four main issues that evolved in the students' reasoning with statistical models and modeling during the learning trajectory: 1) a descriptive versus a predictive model view; 2) a dynamic model view; 3) subjective versus objective confidence level; and 4) intuitive versus scientific understanding of sample-population relationships.

Descriptive vs. predictive model The first time the students refined their model happened after they explored real sample data. They refined their "most of" model to a "probabilistic prediction of phenomenon" model (Figure 6) using the TinkerPlots2 Sampler. This refined model had the same attributes as the "most of" model, but instead of presenting only "most of" the students preferences, it described all the possible attribute values and percentages for the attributes' relative frequencies. When Ido explained how they built this model, he said that the model described the likelihood of event occurrence. Ohad corrected him and said that the model did not describe chances. Therefore, Ido suggested that the model described the way in which the results would be distributed.

We suggest that these two ways of describing the model (descriptive model and predictive model) depict two different views of phenomenon simplification. A descriptive model illustrates how attributes in a phenomenon are distributed. A predictive model presents a phenomenon distribution in order to predict the chance of events happening. Ohad had not yet made the connection between relative frequencies and chances, therefore, in his opinion, chances were not part of this model.

Static vs. dynamic model A dynamic model is a model that can generate samples. After the students built the first model in TinkerPlots2 and generated samples from it, they explained how the measurements of relative frequencies of particular events in the generated sample were related to chances that they set in the model. For example, Ido explained that "[the result of] sporting events didn't appear [in the sample] because we gave it [in the model] very low chances." The connection made between chances set in the model to relative frequencies in the generated sample indicates a different view, a dynamic one, the students had acquired about the model.

We refer to the first two models: descriptive and predictive models as static, and to the third one as a dynamic model. We argue that the change from a static to a dynamic model view stemmed from a change in the purpose of building a model. Although the purpose in presenting the static model was to describe the population or predict the population behavior, the purpose in presenting the dynamic model was to examine the model validity, meaning, and how confident the students were in their inference based on a random sample. Because the dynamic model the students presented was based mainly on a random sample they drew in reality, their reason to simulate samples from the model was to examine to what extent and under what conditions they could trust random samples.

Subjective vs. objective confidence level The process of modeling a phenomenon in the context of ISI is accompanied by probabilistic language (Makar & Rubin, 2009) including reference to the confidence level regarding a conclusion that is drawn from a single sample. One important finding of this study is that the students gradually moved from articulating a subjective confidence level to articulating a more objective confidence level. When the students explored their real sample, they expressed uncertainty accompanied by a low confidence level. The confidence level was usually expressed by a percentage such as 60%. However, their confidence level stemmed from a conflict between context and data and was based mainly on subjective feelings regarding the sample representativeness. The first time the students expressed an objective confidence level in random samples occurred after they explored generated random samples of size 48 (Figure 10) from their TinkerPlots2 model. They realized the need to decide whether a sample result is similar enough to the model. They gave an example of a decision rule for similarity between each statistic in the sample and its appropriate parameter in the model, a range of 5% above and 5% below the parameter (see the third sub subsection in subsection 5.2). Then they suggested a quantification of the confidence level based on the number of statistics that were similar to their appropriate parameters. Later, when they built a sampling distribution, they used this range once again and quantified their confidence level in samples of size 48.

Intuitive vs. scientific understanding of sample-population relationships This study supports the argument of scientific educators that modeling reasoning can serve as a bridge that facilitates the shift from personal, intuitive knowledge to a more mathematical-scientific understanding of the world (Lehrer et al., 1994). In the beginning of the IMA learning trajectory students thought of sample-population relationships as a multiplicative relation. The students' use of the Sampler created new possible explorations regarding random sampling. After the students drew samples from the model and realized that the Sampler drew data randomly, they decided to change the two first attributes in the model, gender and grade, from devices to counters (Figure 7). The rationale for this action was that they wanted the Sampler to draw samples from the model using the same method used by the students to draw real sample data. The real sampling method chose an equal number of females and males from each of the fourth, sixth, seventh, and ninth grade classes. We believe that the last refinement reflects the students' new statistical understanding of the world (Lehrer et al., 1994). The students realized that a model was a dynamic machine that can create generated data and that the model structure influenced the way data were created. Furthermore, they began to think about the mechanism of random data creation also in reality. In other words, the students made the connection between real random samples from a population (the data world) and generated random samples from a model (the model world).

6.3. IMPLICATIONS OF THE FRAMEWORK

Researchers draw attention to the necessity of probability models in developing informal inferential reasoning (Fielding-Wells & Makar, 2015; Rossman, 2008), and helping students integrate data and chance (Konold & Kazak, 2008). Although most studies emphasize the importance of using models in making probabilistic predictions about chance situations, this article sheds light on the role of statistical models and modeling in developing students' informal inferential reasoning with real-world phenomena. This article depicts how the transition from static models to dynamic models enables students to learn about the underlying mechanism of random sampling and thus opens a new line of inquiry about the representativeness of random samples and the validity of the model.

The suggested *framework for reasoning with models and modeling* discussed in Section 5 (Figure 3) emphasizes the importance of making students aware that the exploration of the behavior of real-world phenomena includes not only the attribute distributions and the relationship among them, but also the behavior of random sampling. Learning about the representativeness of samples (the second dimension of the framework) can help students to refine their phenomenon's simplification in the first dimension ("reasoning with phenomenon simplification") and can also change students' views about sampling distributions in the third dimension ("reasoning with sampling distribution"). Exploring sampling distributions can help develop students' understanding about repeated sampling and thus can help them validate the phenomenon's simplification. Therefore, we argue that learning about attribute distributions of the explored phenomena as well as learning about the behavior of random samples are both linked and required.

Learning about attribute distributions of the explored phenomenon includes looking for signal and noise in data, as well as searching for patterns, trends, and relationships among attributes in order to learn about real-world phenomena. Learning about the behavior of random samples includes exploration of random sampling variability, and examination of the role of random sample size on sampling variability. It is essential to learn the relationship between the two to deepen reasoning with sampling and informal inferential reasoning. This article also sheds light on the potential of the IMA in helping students integrate both concepts of attribute distributions and behavior of random samples.

Researchers and teachers who would like to repeat the activities of the IMA need to take into account that our students were involved in EDA activities during the previous year and therefore were exposed to ideas of sample size and inferences that can be drawn informally from the sample. We suggest that in the IMA learning trajectory, students' experience with an exploratory approach to data is essential for entering the model world and dealing with the complex idea of uncertainty (Pfannkuch, Wild, & Parsonage, 2012). Reasoning with uncertainty in the context of informal statistical inference is an ongoing discourse aimed to convince others regarding inferences that can be made, and the level of confidence in making those inferences. The fact that our students were accustomed to a learning environment of open discourse during the previous year prepared them to discuss and deepen their reasoning with uncertainty and inference in this study.

6.4. SUMMARY

The findings of this article are based on a pair of students with excellent communication and thinking skills. More research is needed to study students' reasoning with models and modeling using the IMA learning trajectory in a classroom format. We are currently conducting another study of sixth grade students to test the case presented in this article. Further research is also needed to refine and extend our understanding of: 1) design considerations that can support teaching or learning of informal inferential reasoning about real-world phenomenon with the addition of a modeling perspective; and 2) how students understand and use statistical models when they make informal inferences on real-world phenomena, and what ideas are needed to understand and use models.

In this empirical study we strove to deepen our understanding of the development of students' reasoning with models and modeling in the context of ISI. The suggested framework can shed light on the role of reasoning with models and modeling on students' understanding of key issues of IIR such as: sample-population and data-chance relationships. We believe that the IMA may contribute to the body of design research in statistics education by helping to develop students' understanding of modeling, sampling, and uncertainty in the context of ISIs.

REFERENCES

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Utrecht, The Netherlands: CD-β Press, Center for Science and Mathematics Education.

Bakker, A. (2007). Diagrammatic reasoning and hypostatic abstraction in statistics education. *Semiotica*, *164*(1/4), 9-29.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the 7th International Conference on the Teaching of Statistics, Salvador, Bahia, Brazil, July 2-7). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/icots7/2D1_BENZ.pdf

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM – The International Journal on Mathematics Education*, *44*(7), 913-925.

Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about Informal Inferential Statistical Reasoning: A collection of current* *research studies*. (Proceedings of the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL5), pp. 29-35). Warwick, UK: University of Warwick.

Chinn, C. A., & Sherin, B. L. (2014). Microgenetic methods. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 171-190). New York: Cambridge University Press.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9-13.

Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 1-28). Minneapolis, MN: Catalyst Press.

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

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM – The International Journal on Mathematics Education*, *44*(3), 883-898.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, *12*(3), 217-230.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1). Retrieved from
http://repositories.cdlib.org/uclastat/cts/tise/ vol2/iss1/art1/

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic Data Exploration* (Version 1.0) [Computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C., & Miller, C. D. (2011). *TinkerPlots* (Version 2.0) [Computer software]. Emeryville, CA: Key Curriculum Press. Available from www.tinkerplots.com

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.

Lehrer, R., Horvath, J., & Schauble, L. (1994). Developing model-based reasoning. *Interactive Learning Environments*, *4*(3), 219-231.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, *14*(1), 69-108.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1&2), 152-173.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82-105. Retrieved from http://iase-web.org/documents/SERJ/SERJ8(1)_Makar_Rubin.pdf

Manor Braham, H., Ben-Zvi, D., & Aridor, K. (2014). Students' reasoning about uncertainty while making informal statistical inferences in an "integrated pedagogic approach." In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13-18). Voorburg, The Netherlands: International Statistical Institute. Retrieved from
https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8C2_MANOR.pdf

Meira, L. (1998). Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education*, *29*(2), 121-142.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the 7th International

Conference on Teaching of Statistics, Salvador, Bahia, Brazil, July 2-7). Voorburg, The Netherlands: International Statistical Institute. Retrieved from https://iase-web.org/documents/papers/icots7/6A2_PFAN.pdf

Pfannkuch, M. (2008). Building sampling concepts for statistical inference: A case study. *Proceedings of the 11th International Congress on Mathematical Education* (pp. 6-13), *Monterrey, Mexico.* Retrieved from http://tsg. icme11.org/document/get/476

Pfannkuch, M., Wild, C., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM – The International Journal on Mathematics Education*, *44*(7), 899-911.

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, *31*(5), 602-625.

Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 3-4. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pratt_Ainley.pdf

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.

Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Rossman.pdf

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *School and general issues* (Proceedings of the 3rd International Conference on the Teaching of Statistics, Dunedin, New Zealand, August 19-24). Voorburg, The Netherlands: International Statistical Institute. Retrieved from https://iase-web.org/documents/papers/icots3/BOOK1/A9-4.pdf

Schoenfeld, A. H. (2007). Method. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69-107). Charlotte, NC: Information Age Publishing.

Thompson, P., Liu, Y., & Saldanha, L. (2007). Intricacies of statistical inference and teachers' understandings of them. In M. Lovett & P. Shaw (Eds.), *Thinking with data* (pp. 207-231). Mahwah, NJ: Erlbaum.

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison- Wesley.

HANA MANOR BRAHAM
Faculty of Education
University of Haifa
199 Aba Khoushy Ave.
Mount Carmel
Haifa, 3498838
Israel