EXPLORING INFORMAL STATISTICAL INFERENCE IN EARLY STATISTICS: A LEARNING TRAJECTORY FOR THIRD-GRADE STUDENTS

SOLEDAD ESTRELLA
Pontificia Universidad Católica de Valparaíso
soledad.estrella@pucv.cl

MARITZA MÉNDEZ-REINA
Pontificia Universidad Católica de Valparaíso
maritza.mendez.r@mail.pucv.cl

PEDRO VIDAL-SZABÓ
Universidad del Desarrollo
pfvidal@udd.cl

ABSTRACT

Recent research suggests the benefits of supporting a progressive understanding of concepts of inference prior to the teaching of procedures and formal calculations through the study of informal statistical inference (ISI). To contribute to the growing knowledge about the early learning and teaching of statistics, particularly regarding the development of informal inferential reasoning (IIR), we designed a learning trajectory (LT) that addresses ISI in K–4 students (ages 5 to 9 years). This article describes part of the LT in detail, in which third-grade students (n = 12) were introduced to sampling, frequency distribution, randomness and sampling variation as well as to developing a data sense in online lessons. The results of this type of teaching show that the creation and collection of authentic data in a playful context, together with an exploratory analysis of the data as a precursor to utilizing aspects specific to IIR, promoted the integration progress of IIR components in the oral and written informal inferences of students.

Keywords: *Statistics education research; Early statistics research; Learning trajectory; Informal inferential reasoning; Informal statistical inference*

1. INTRODUCTION

Using and reasoning about information extracted from data are necessary thinking skills for citizens in 21st-century society. These skills can be developed at an early age by providing opportunities for students to informally explore the behavior of data, to make conjectures, and build arguments to support inferences about a more widespread group.

Statistical inference begins with the collection and analysis of data, and it is through these processes that inferential statistical reasoning makes it possible to work effectively with data sets. Pfannkuch (2006) defined the term informal statistical inference (ISI) as “the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data” (p. 1). In general, the development of statistical reasoning involves the development of informal inferential reasoning, which is defined as the process of making probabilistic generalizations from (supported by) data that extend beyond the collected data and have a particular level of confidence (Ben-Zvi et al., 2015; Makar & Rubin, 2009).

Participation in activities that involve making (informal) inferences in the early years of school seems to facilitate more complex learning about formal statistical inference (Estrella et al., 2022; Leavy et al., 2018; Makar & Rubin, 2009; Manor Braham & Ben-Zvi, 2017). For example, reflecting on a

Statistics Education Research Journal, 22(2), Article 10. https://doi.org/10.52041/serj.v22i2.426
© International Association for Statistical Education (IASE/ISI), July 2023

simple random experiment, such as the toss of two coins, in a playful context can foster a progressive and early understanding of randomness since it both connects the observation of a random behavior that tends to regularize after many repetitions and makes evident the unpredictability of the results obtained after a few repetitions, thus regularity is recognized (Estrella et al., 2022).

In the field of education, an increasingly important construct in pedagogical research is the learning trajectories (LT) (Maloney et al., 2014; Simon, 1995, 2018). A LT is a construct that supports task designs by characterizing and identifying an instructional itinerary to develop students' thinking and reasoning processes (Simon, 1995). Arnold et al. (2011) suggested that carefully designed learning trajectories can stimulate students to gain access to inferential concepts and reasoning processes. Van Dijke-Droogers et al. (2020), however, suggested that further research is still required to show that ISI can be developed in young students. Additionally, they recommend carrying these results into theoretically grounded learning trajectories in which students are introduced to key ideas, such as sampling and the associated probability component.

The present study seeks to contribute to this line of research on ISI in the first years of school by exploring students' comprehension of sampling, frequency distribution, randomness and sampling variation. The understanding of these ideas contributes to the development of informal inferential reasoning since students collect data sets and learn to make reliable statistical inferences while focusing their attention on trends and patterns of data behavior and making judgments about those samples (Ben-Zvi et al., 2015; Makar, 2016; Saldanha & Thompson, 2002). The hypothesis of this study is that LTs—based on an ordered network of experiences that students encounter through situations that promote the generalization, refinement and successive testing of the relationships between data and chance—can lead students to develop informal inferential reasoning (IIR).

This paper describes part of the design process of an LT that aims to support students to make ISIs about the data samples obtained from a random experiment, identify the regularities and the variation they observe, and gradually develop their IIR.

2. CONCEPTUAL FRAMEWORK

The most important objective of statistics education is to make students and citizens understand that decision-making in situations of uncertainty is based on data samples (Shaughnessy, 2019). In this sense, when only partial data are available, statistical inference allows for the formulation of conclusions in situations of uncertainty (Makar & Rubin, 2018).

2.1. INFORMAL INFERENTIAL REASONING

The trend in statistics education is to place ISI at the center of the school curriculum, which requires a rethinking of how to build this reasoning about inference concepts and how to teach it (Garfield et al., 2015). In recent decades, ISI has been established as a paradigm for teaching that integrates aspects of IIR: using *data as evidence* to establish claims associated with a question or problem, prioritizing the evidence provided by the data over personal experiences or opinions (Makar & Rubin, 2009; Pfannkuch et al., 2015); *generalizing beyond data* at hand to make inferences about a broader set of cases (Rossman, 2008; Zieffler et al., 2008); *expressing uncertainty* to manifest the uncertainty underlying generalizations beyond the data considering that the statements are not given in absolute or certain terms (Ben-Zvi et al., 2012); *considering the aggregate* to view the data set as a whole, focusing on one or a few behavioral characteristics of the data (Konold et al., 2015); and *integrating contextual knowledge* to consider possible relationships present in a situation while deepening reasoning with data in context (Langrall et al., 2011).

Zieffler et al. (2008) recognized "informal inferential reasoning as the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (p. 44) or inferences dealing with a sampling process in random experiments (Pratt et al., 2008). In their research on the topic of learning to reason from samples, Ben-Zvi et al. (2015) stated that the concepts of sample and sampling were key ideas of statistical inference and that it is important to introduce these ideas to students early in their statistical learning. The authors note, "Taking representative samples of data and using samples to make inferences about unknown populations are at the core of statistics. An understanding of how samples vary (sampling variability) is crucial to make reasoned data-based estimates and decisions" (p. 292).

In addition, it is important for the learning of statistics that students have the opportunity to construct the sampling distribution and experience the process involved (Ben-Zvi et al., 2015; Estrella et al., 2023). Repeated sampling generates empirical sampling distributions created from data collected by students; this is considered key for the development of statistical inferential reasoning from an informal approach (Silvestre et al., 2022). Conversely, different data representations can be generated that reveal aspects of how the data collected in the samples are distributed, allowing students to move from examining individual values to studying aggregates, trends, and patterns in the data (Konold et al., 2015) and ultimately making inferences from the data for a whole population.

Randomness as a model describes random phenomena in which the outcome of a single repetition is uncertain, but after a large number of repetitions, it shows a regular distribution of relative frequencies (Moore, 2000). According to Moore, the best way to understand randomness is to look at random behavior, not only the regularity that appears after many repetitions but also the unpredictable results obtained after a few repetitions. In this regard, fourth-grade students are able to observe how variation decreases as the sample size increases (English & Watson, 2016).

Demonstrations of random sampling and repeated sampling allow greater confidence in generalization over a broader group. In our proposal, randomness in samples of sufficient size that is incorporated into a task with random generators, such as coins, is the basis for increasing confidence in informal inferences (Pratt et al., 2008). To develop IIR, it is necessary to give students opportunities to have authentic decision-making experiences in which they must make informal inferences based on samples and then work beyond the collected data, using language that indicates awareness of the uncertainty underlying these inferences. This implies a need to work simultaneously with the samples and the conclusions that are based on them (inferences) as a fundamental aspect of prediction and decision-making (Manor Braham & Ben-Zvi, 2017; Watson & Moritz, 2000).

For our analysis of the emergent reasoning of children when working with samples and sampling, we must address the characteristics of ISI as well as the rationale behind the design of the tasks we used to stimulate the development of IIR: make a prediction and compare it with the collected data; visualize and recognize sample variation; become aware of the regular behavior of the samples by assigning a confidence level; and generate assertions beyond the available data using expressions of uncertainty.

The research questions guiding this study are the following:

> How can we promote young students' reasoning about fundamental inferential concepts (sampling, frequency distribution, randomness, and sampling variation) through a playful learning context?
>
> How can these concepts be taught through the enactment of an LT such as the one designed and implemented in this study.

3. METHODOLOGY

LTs allow for the study of progressive levels of IIR in students in which each level is more sophisticated than the previous level. Arnold et al. (2018) suggested that many statistical ideas and processes—such as those that are considered in informal statistical inference—are not currently included in school curricula, and that carefully designed LTs can stimulate students to gain access to these ideas and processes. Many researchers in statistics education have used LTs together with design-based research methods to promote and assess the IIR processes of students (e.g., Manor Braham & Ben-Zvi, 2017; Meletiou-Mavrotheris, & Paparistodemou, 2015). The study reported in this paper contributes to this body of literature through the design and implementation of a LT aimed at offering learning opportunities that progressively develop IIR in students early in their school training (Grades K–3).The designed LT on IIR for early grades proposes an exploration of data that excludes computational simulation and uses specific random artifacts (coins). In this way, students can explore the distribution of the data they obtain and make conjectures as to what would happen if it were repeated multiple times. The proposed learning trajectory includes accessible inferential concepts in a playful context that dispenses mathematical procedures and encourages children to use their IIR successfully.

3.1. A LEARNING TRAJECTORY AS AN INSTRUMENT OF DESIGN RESEARCH

The research presented in this paper was part of a broader study that aimed to develop a theoretical and empirically based LT to initiate an understanding of key concepts of ISI among students in the early grades. The LT is comprised of four stages that promote specific statistical knowledge related to IIR. In the theoretical design of the LT, hypotheses are posed regarding the learning of ISI, including the concepts of sampling, frequency distribution, randomness and sampling variation, which progress in steps (see Table 1). Table 1 summarizes the proposed progression in terms of tasks, teaching activities, student learning, the statistical concepts involved, and the expected IIR achieved in Steps 1, 2, 3 and 4 of the LT. These steps are concatenated and developed progressively. This table contents, based on van Dijke-Droogers et al. (2020, 2021), provides an overview of each of the steps in promoting IIR. Below, the four steps of the LT, the hypotheses that are considered and a description of their fulfilment are characterized.

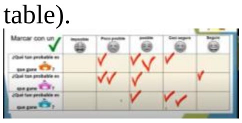
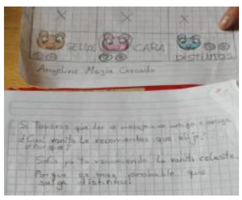
Step 1. Make a prediction that contrasts with the obtained data. Two hypotheses were considered: (1) the students will begin to develop a data sense as they obtain and record categorical data in which there is variation and uncertainty in a playful context; (2) after making a prediction, the students will obtain data that they will contrast with their predictions, allowing them to begin to identify regularity and to progressively assess the necessity of the data. Thus, in Step 1, the first hypothesis is that students will become aware of the data in a significant context and of the uncertainty engendered by sampling variation. By recording and obtaining the data for their sample, they will be able to see and compare their results and will understand that the behavior of their sampling data is not necessarily the same as that of other samples.

Step 2. Visualize and recognize sample variation. The first hypothesis suggests that (1) the students will become aware of the randomness of the experience based on repeated sampling and of their justifications for their predictions both in terms of the regularity and of the unpredictability of the results; consequently, they will develop awareness of sample variation. Through repeated sampling, students will visualize and recognize sample variation and variation among samples. When they present their samples to others, they will be able to see and compare their results, and the students will understand that the behavior of the data in their sample is not necessarily the same as that of their classmates' samples.

Step 3. Become aware of the regular behavior of the samples by assigning a level of confidence to each event. Two hypotheses are proposed: (1) students will begin to distinguish the frequency distribution model of their obtained results from a model of a repeated sampling frequency distribution, which allows for the investigation of variation and uncertainty; (2) qualitatively, students will assign a level of confidence to each event, determining its possibility of occurrence using language that conveys degrees of uncertainty. As the number of samples increases, students will notice a certain regularity in the behavior of the data and will understand that although there are changes in the data of each sample, there are some common patterns of sample variation, for example, the occurrence of an event that has the maximum frequency in most samples. This allows them to express different levels of confidence regarding the occurrence of each event in future samples.

Step 4. Generate statements beyond the available data as evidence using expressions of uncertainty. Two hypotheses are considered: (1) students will recognize the effect of repeated sampling, and through the use of more samples, they will come to understand the regular behavior of samples, interpreting the variation and uncertainty involved; and (2) students will be able to make a reasonable conclusion about which result has a qualitatively greater chance of being true based on the data samples. Students will be able to propose claims about an event that is most likely to occur in the future in a numerical or graphical conclusion that integrates the regularity detected in the behavior of the samples and the frequency distribution, making explicit their level of confidence in the generalization.

Table 1. Summary of the activities in each step of the LT for IIR

	Step 1	Step 2	Step 3	Step 4
Task	Which frog do you think will win? Let's play and record the results on the boards. What happened after we played?	Let's look at the boards. What happened in the four games? Between each game, did you change your prediction about which frog would win? Why?	After looking at the boards for the games that you played and those games that your classmates played, how likely is it for each frog to win?	Which frog do you think will win the next game?
Teaching activity	Play the game, recording the resulting data on the boards (sampling).	Display the individual samples (boards). Visualize the frequency distribution.	-Ask each student to estimate the possibilities in terms of levels of confidence (impossible, unlikely, possible, very likely or certain), accounting for the variation pattern.	-Encourage students to make inferences regarding a future outcome based on the samples with a certain level of confidence.
Concepts	Variable; randomness; sample; sampling variation; repeated sampling.	Randomness; increase in sample size; sampling variation; frequency distributions for repeated sampling.	Confidence levels; randomness; sampling variation; frequency distributions for repeated sampling.	Repeated sampling for inference.
IIR components expected	Data as evidence: experience the variation within the individual sample and repeated sampling.	Data as evidence: recognize the variation among samples and observe a pattern in that variation.	Data as evidence expressed with uncertainty: recognize the uncertainty of the situation and the pattern of variation in the samples; assign possibilities of occurrence.	Data as evidence expressed with uncertainty and claims beyond the data: express an informal statistical inference.
Student activity	Predict, play the game and record the resulting data for each category of the variable. 	Compare predictions with the data from the samples. Visualize the sample variation. 	Recognize regularity. Estimate possibilities by assigning a confidence level. Compare and visualize estimates with those of others (recorded in the table). 	Communicate a claim based on IIR. 

Central task. The playful context involves comparing predictions about a game called the *Frog Race* (see Figure 1). The features of the central task are described below. For each step, Table 2 shows corresponding task questions and the indicators of student learning behavior that support each hypothesis. This game uses a game board and an "X" to indicate advancement to the goal box, which results from the toss of two coins. The rules indicate that if there are two heads, the orange frog advances; if there are two tails, the pink frog advances; and if the two coins are different, the blue frog advances. The game ends when one of the frogs has five Xs (a completed game board constitutes a sample).

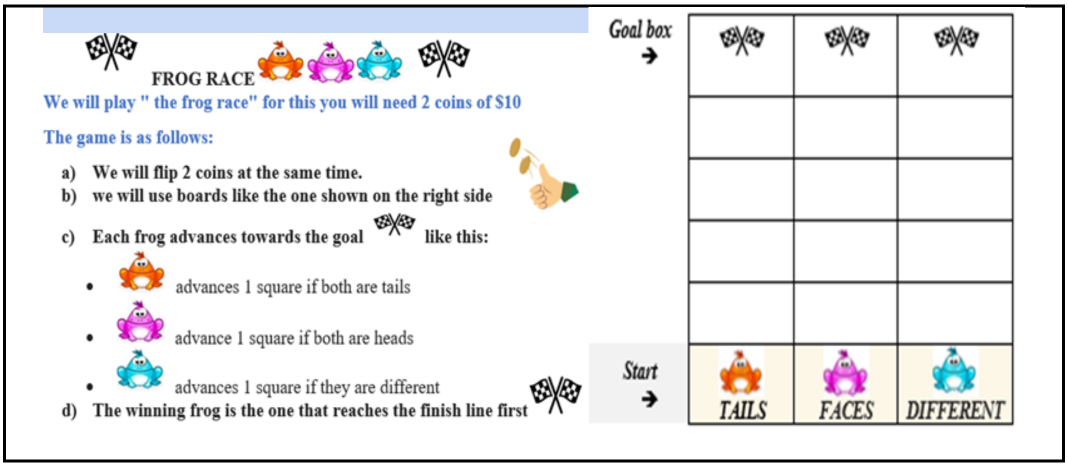


Figure 1. Game instructions and board for the "Frog Race" (Estrella et al., 2022)

Table 2. Indicators of student learning behavior that support the hypotheses

Step	Task	The indicators of student learning behavior that supports hypothesis
1. Data collection and sample	Which frog do you think will win? Let's play and record the results on the boards. What happened after we played?	1a, Recognizes the events that occurred in the random experience and chooses on of these (make a prediction after obtaining data). 1b. Performs an iterative action with specific objects and identifies the resulting event. 1c. Records the occurrence of each of the events (resulting data) until one of them reaches the set maximum frequency for a representation (data collection; sample). 1d. Determines the number of times or height of the occurrences of each event on the board (absolute frequencies). 1e. Contrasts whether the prediction was fulfilled by the data resulting from the random experience. 1f. Experiences the uncertainty present in the random experience (unpredictability).
2. Repeated sampling and sampling variation.	Let's look at the boards. What happened in the four games? Between each game, did I change my prediction about the winning frog? Why?	2a. Observes the different samples for each random experience (randomness; repeated sampling). 2b. Compares the frequencies of events in a sample (sample variation). 2c. Compares the frequencies of events after several repetitions (variation among samples).
3. Frequency distribution.	Based on the boards for the games that you played and those that your classmates played, how likely is each frog to win?	3a. Detects visually the event with the maximum frequency as the number of samples increases (regularity). 3b. Compares these frequencies to observe the regular behavior (pattern after several repetitions (sample variation, regularity). 3c. Assigns a level of confidence to each event that describes its possibility of occurrence (language with levels of uncertainty).

4: Inference with degrees of confidence	In an upcoming game, which frog do you think would win?	4a. Makes conclusions beyond the obtained data, that is, based on the random behavior of the samples. 4b. Expresses a conclusion about the future with a certain level of confidence. 4c. Expresses confidence about the occurrence of an event in the next sample space. 4d. Makes an argument based on the events with maximum frequency in the total of all samples. 4e. Makes an argument based on the height of the frequencies in the representation, expressing a certain level of confidence (frequency distribution, sample variation). 4f. Outlines the frequency distribution, imagining the behavior of the next sample (frequency distribution, sample variation: regularity.

3.2. PARTICIPANTS AND CONTEXT

In 2020, a lesson study group was formed (cf. Estrella et al., 2022) that consisted of six teachers (three primary school teachers, one preschool educator, two special educators) who worked in an urban school and three researchers with expertise in statistics education who collaboratively designed, implemented, evaluated and modified a learning sequence to promote IIR in Grade 3 students.

The lesson study group met online due to COVID-19 weekly for 2 hours over the course of 9 weeks. The activities that the teachers directed were delineated by the team of researchers and based on four principles: authentic experiences of informal inferential reasoning; planning of a learning sequence; implementation, revision, and adjustment of the lesson plan that promotes IIR (for further details, see Estrella et al., 2022); and analysis of and reflection on the process. After its realization, the in-service teachers were able to better understand the elements that were characterized by the research and pedagogical level of the ISI approach and integrate these elements of the LT proposal into the lesson plan as formulated, implemented, and evaluated.

To conduct the empirical verification of the hypotheses, the LT proposal was implemented in four 45-minute lessons over the course of two weeks, offered online due to COVID-19. The teacher in charge of each class, together with a special educator, implemented each of the tasks in the sequence of the LT into the context of a lesson study.

Specifically, this study reports the implementation of these lessons with 12 third-grade students (8 years of age). The students were inexperienced with ISI. The selection criterion for the inclusion of episodes considered the richness of the students' verbal and written productions during classroom interactions between teachers and students and among students.

3.3. DATA COLLECTION

For Steps 1 and 2, the data collected included the game boards completed by the students in the form of photographs of each board and a video recording of the entire lesson. For the other steps, we proceeded in a similar way; that is, all of the sessions were recorded. For example, in Step 4, the students wrote their conclusions and inferences in their notebooks, photographed the work and then submitted it to their teacher using WhatsApp. The meetings and the implemented lessons were held in 2020 online through the Zoom platform, which expedited video storage.

3.4. DATA ANALYSIS

At the beginning of the data analysis, the researchers (authors) transcribed and coded specific episodes from video data that were relevant to the research questions. The authors summarized in an overview the steps of the LT from a progressive perspective, which related the aspects of the ISI to the data sources (i.e., whole class discussion (WD), student game boards (SGB), teacher-student interaction (TSI), and student–student interaction (SSI)), through the indicators that support the hypotheses. The three authors ratified most of the indicators and reached consensus on the discrepancies. Then, the researchers chose the relevant episodes and analyzed them according to the indicators of student learning behavior.

Finally, the data analysis included verification among the researchers, in which the three authors reviewed the episodes as a unit of analysis to characterize each step of the LT for IIR. During the process the group of researchers looked for agreed-upon interpretations and examined contradictory interpretations to confirm the students' level of verbal and written informal inferences.

4. ANALYSIS AND RESULTS

The LT implemented increased progressively the complexity of the concepts and the connections the students encountered (see Table 3). This maintained a teaching dynamic that activated whole-class discussion and teacher–student interaction and promoted student–student interaction, as well as the educational use of an individual student game board. These elements provided the instances of data analysis that aligned with the indicators of student learning behavior (see Table 2).

Table 3. Summary of the indicators for each step of the LT and the key concepts of ISI

	LT Step 1: data collection and sample			LT Step 2: sample variation			LT Step 3: frequency distribution			LT Step 4: inference with degree of confidence		
Increased complexity	Introduction to prediction, unpredictability and variation			Visualization of variation and regularity			Interpretation of variation and uncertainty			Generalization with expressions of uncertainty		
Data source	WD	GB	TSI or SSI	WD	SGB	TSI or SSI	WD	SGB	TSI or SSI	WD	SGB	TSI or SSI
Sample	1a 1d 1e 1f	1a 1b 1c 1d 1e 1f										
Repeated sampling	1f	1b 1c	1a 1e 1d 1f	2a	2a	2a	3a	3a				
Sample variation				2a 2b 2c	2a 2b	2a 2c	3b		3a 3b	4e	4f	4c 4d 4e
Regularity							3b 3c	3b	3a 3c	4a	4f	4a 4b 4c
Randomness	1f	1f	1f	2a			3c		3c	4a		4a 4b
Frequency distribution				2a 2b	2a 2b 2c		3b	3b		4a 4e	4f	4b 4c 4d 4e
Use of visualization	1e 1f	1d		2a 2b		3a 3b				4a	4f	4a 4b 4e

Note: Whole-class discussion (WD); student game board (SGB); teacher-student interaction (TSI); student–student interaction (SSI).

The teacher oriented the questioning process and monitored student learning by asking questions that encouraged the verbalization of ideas and reasoning, asking for data-based justifications to support their claims, and inviting observations of the behavior of their own tabulated data (variation of the results) and that of others (frequency distribution).

In the following four sub-sections, for each step of the LT, an episode of the lesson is shown that reveals examples of the expected indicators of student learning behavior (see Table 2 & Table 3). In addition, Figures (2–5) present the students' game boards (the numbering of the game board proceeds from left to right) that show a circle that indicates the student's predictions before playing and frame rectangles (green indicates the winning results after playing, and gray indicates the non-winning results).

4.1. STEP 1: MAKE A PREDICTION THAT CONTRASTS WITH THE DATA COLLECTED

***Oscar*.** In this episode, the student predicted who would win before playing and obtaining data. He communicated his first prediction about the winning frog among the available events (with a circle indicating the event "tail and head." Oscar then analyzed the data for the sample obtained and contrasted his initial prediction with the results, indicating the correctness of his prediction (a tail and a head). This step entailed reviewing the data that constituted the first sample and observing the occurrences of each event on the board (absolute frequencies).

Teacher: This is the first game (...). Can you tell me which frog you chose in Game 1?
Oscar: The blue one.
Teacher: … And why did you choose that frog?
Oscar: Because sometimes a tail and a head happens.
Teacher: In that first game, which frog won?
Oscar: The blue one.
Teacher: So, what happened in that game of yours?
Oscar: I won!

Figure 2. First sample from a student, Oscar, with his predictions (circled) and the maximum frequency (framed with a green rectangle)

4.2. STEP 2: VISUALIZE AND RECOGNIZE SAMPLE VARIATION

***Oscar*.** This episode shows the point at which the student recognized the variation in the sample. After obtaining additional samples (see Figure 3), he compared the frequency distributions when the teacher asked questions that led him to observe and differentiate the frequencies of the events that occurred in one versus several of his samples.

Teacher: In that first game [Game Board 1], which frog won?
Oscar: The blue one.
Teacher Which frog won in this game [Game Board 2]?
Oscar: The blue one.

Teacher What is the difference between Game 1 and Game 2 [Game Boards 1 and 2]? What happened to the other frogs? [focuses on the student visualizing the sample variation]

Oscar: In [Game Board 2], for the pink one, I made 3, and for the orange one, I made 2. Instead, on [Game Board 1], [it was] pink 2 and orange 0.

Figure 3. Four samples from a student, Oscar, with his predictions (circled) and the maximum frequency (framed with a rectangle)

Valery. This episode shows how the student perceived sample variation by comparing the frequencies of events after several repetitions. It illustrates that Valery chose different events that she sometimes chose correctly. Her predictions were blue, pink, orange, blue, while the results were blue, orange, blue, blue (see Figure 4). and, finally, how she chose the event with the maximum frequency [she described aggregately how samples behave], which occurred twice previously (blue frog won in Game Boards 1 and 3), although she expressed the unpredictability of the events at the time when she expressed uncertainty about the blue frog winning.

Teacher: Considering all the games you played, did you change the color of the frog between each game [board]?

Valery: Yes.

Teacher: And why did you change colors?

Valery: Because it may be that the blue one will not win.

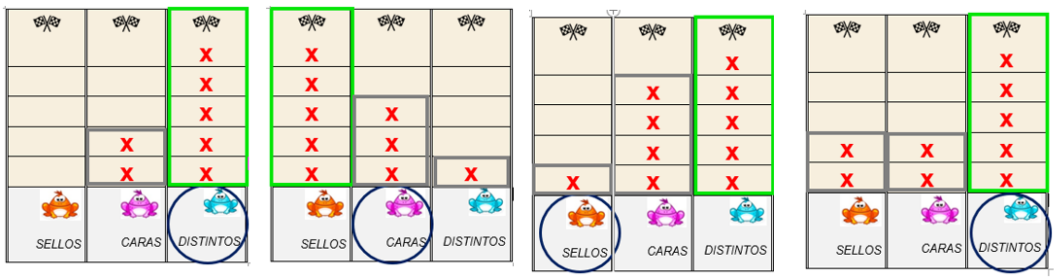


Figure 4. Four samples of a student, Valery, with her predictions (circled) and the maximum frequencies (framed with a rectangle)

4.3. STEP 3: SHOW AWARENESS OF THE REGULAR BEHAVIOR OF SAMPLES BY ASSIGNING A LEVEL OF CONFIDENCE TO EACH EVENT

Anne. The following episode shows the student identifying the event with the maximum frequency. As the number of samples increased, she persisted in her prediction [always chose blue] (see Figure 5), which demonstrated her level of confidence in the regularity of this outcome's possibility of occurring. As the game progressed, Anne compared the results with her predictions and adjusted her thinking accordingly.

Teacher: Did you always choose blue? Why?

Anne: Yes, because I thought it was more likely that heads and tails would land than others. However, I was wrong ... [contrasts her previous predictions].

Teacher: Where did you go wrong?

Anne: In the second game.
Teacher: What happened in the second game?
Anne: I chose the blue frog, and the pink frog won [contrasts her prediction and perceives the sample variation and unpredictability].
Teacher: So which frog had the least chance of winning in your games?
Anne: The orange frog [fewer possibilities].

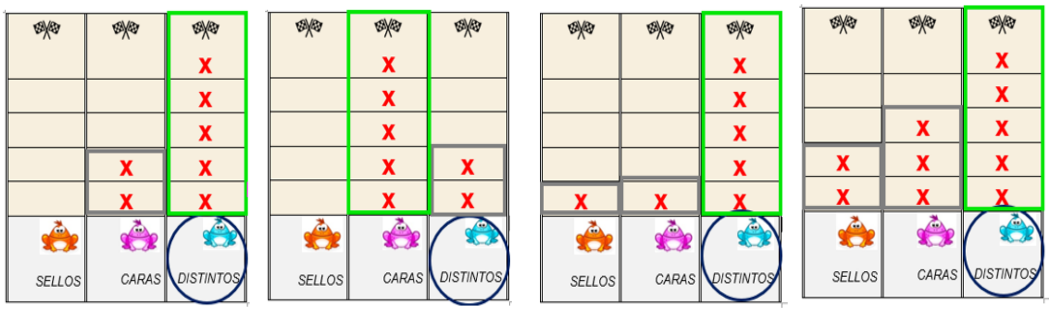


Figure 5. Four samples of a student, Anne, with her predictions (circled) and the maximum frequencies (framed with a rectangle)

Emanuel and Jessica. This episode shows the recognition of regularities in the behavior of the data in the samples when observing the frequency of events, specifically, that some events occur more often than others. The students assigned different levels of confidence to events using terms such as impossible, unlikely, possible, very likely and certain (each expression was accompanied by smileys; see Figure 6). This was a collective moment in which each student communicated his or her assessment of the possibilities of each event. During this activity, the teacher completed a table (see Figure 6) that allowed the whole class to consider all of the students' assessments.

Teacher: Which frog would you choose if you were to play again?
Jessica: Pink is my favorite color, and sometimes it may be the turn of this [pink frog].
Teacher: What does Emanuel think of what Jessica says?
Emanuel: It is possible that the pink [frog] will win.
Teacher: And why didn't you choose the pink [frog] then?
Emanuel: Because I chose the blue [frog] because it was more likely.

Marcar con un ✓	Imposible	Poco posible	posible	Casi seguro	Seguro
¿Qué tan probable es que gane [orange frog]?	✓	✓	✓✓	✓✓	
¿Qué tan probable es que gane [pink frog]?		✓✓✓✓	✓		✓
¿Qué tan probable es que gane [blue frog]?			✓	✓✓	✓✓✓

Figure 6. Table showing a group of six students' assessments of the possibilities of each event according to confidence levels

4.4. STEP 4: GENERATE STATEMENTS BEYOND THE AVAILABLE DATA USING EXPRESSIONS OF UNCERTAINTY

For this step, two student interactions with the teacher were chosen; the first included two students, and the second included one student.

Emanuel and Jessica. This episode involved an activity in which the students sketched an invented game board by imagining the behavior of a future sample. The students differed in their claims when choosing an event among the possible events: one based his choice on maximum frequency (since he had detected the regularity and the sample variation), while the other based her choice on a personal preference that did not consider the regularity in the samples.

Teacher: Which one wins on your game board? Why did you imagine the game like this?

Emmanuel: The blue [frog] won, and the others lost. My imagination said it's going to be like this … first, the blue wins; it has won a lot of games. It says, "It has already won a lot, then let it be the winner."

Teacher: What do you think of Jessica's [imagined] game board? Because in her games, the little pink frog won.

Emmanuel: Her imagination told her to do that, so it's okay.

Teacher: Emmanuel, do you think Jessica made the game thinking about the results we saw or something else?

Emmanuel: Ah! I say no ... [she did not think about the results]; she thought that pink would win.

Oscar. The following episode shows the student´s inference, in which he has recognized a regularity —"It is more likely to come out heads and tails more often"—and expressed confidence about the future occurrence of the event using language with uncertainty: "because it is more likely" (translated from Figure 7). In this way, he generalized the outcomes by drawing a conclusion that was beyond the data available but was based on the samples.

Teacher: If you played more times, which frog would you choose? Why?

Oscar: I would choose the blue one. There are more chances that it will come out heads and tails more often.

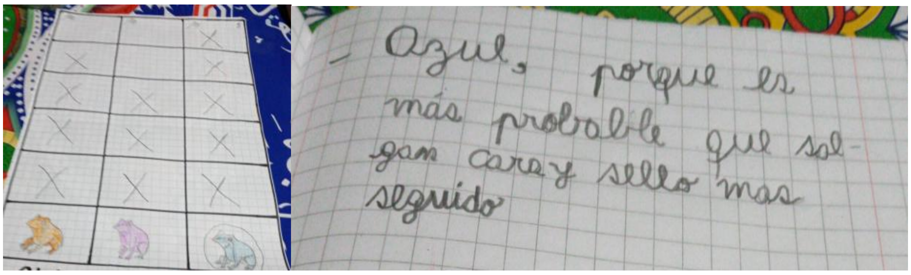


Figure 7. *Oscar's imagined board and written inference about the event that could have the maximum frequency and his assessment using confidence levels*

4.5 SUMMARY OF STEPS 1–4

The random experiment of tossing two coins provided these Grade 3 students with a learning experience that culminated with decision-making in a situation of uncertainty, which generated ISIs. Through this exercise, the students were able to experience repeated sampling and develop an informal understanding of the concepts of sampling, sampling variation, randomness and frequency distribution. The results empirically corroborate the theoretical design of the LT; although the result of each individual toss of two coins is uncertain, a regular pattern of behavior emerges after a large number of repetitions through repeated sampling.

In turn, the connections among the four sequential steps designed to promote the students' learning processes were analyzed in terms of their use of language regarding levels of uncertainty, data-based reasoning, perceptions of variation and unpredictability and explicit awareness of regular behavior. The students were able to identify the characteristics of a sample and interpret the frequency distribution as a description of the frequencies of all possible repetitions of a sample under the same conditions. Several students were able to distinguish the frequency distribution model for the obtained results from the model of a repeated sampling frequency distribution, which allowed them to approach the ideas of variation and uncertainty.

In each LT step, the students had playful experiences that allowed them to complete the gameboard with data generated by them and by their classmates. These teaching activities encouraged them to observe the behavior of the patterns in data and use key concepts such as sampling, variation, regularity, randomness, and visualization of frequency distribution, which underlay their informal inferences.

5. CONCLUSIONS

This article reports a study with a theoretical and empirical design based on an LT that introduces ISI in Grade 3 and allows probabilistic generalizations based on data. The research goal was to investigate an LT about ISI through the following research question: How can we promote young students' reasoning about fundamental inferential concepts (sampling, frequency distribution, randomness, and sampling variation) in third-grade students through a playful learning context, and how can these concepts be taught through the enactment of an LT such as the one designed and implemented in this study? The findings show that the enacted LT encouraged children to make predictions and subsequently contrast them with collected data, visualize and recognize the variation in the sample, become aware of the regular behavior of the samples and generate statements beyond the available data using expressions of uncertainty.

The playful context of the random experiment fostered great interest among students, who created and collected authentic data and had the opportunity to explore the data, thus making ISI accessible to these young learners. By prompting the students to question the data and the informal inferences that arose, the teacher fostered children's intellectual curiosity and made it possible for them to express their developing reasoning regarding the stochastic. According to Simon (2018), an essential aspect of the learning of statistical concepts occurs through the learner's activity, which is largely mental. The results of this teaching show that this LT promotes the integration of progressive IIR's components in the oral and written inferences of students.

The LT proposal has allowed us to interrelate and articulate key ideas of ISI: (1) students record the occurrence of each of the events (resulting data) until one of them reaches the set maximum frequency for a representation (data collection; sample), (2) students compare and contrast their predictions with the actual data resulting from random experiments, (3) students familiarize themselves with uncertainty experiences (unpredictability), (4) students observe the different samples for each random experience (randomness; repeated sampling), (5) students compare the frequencies of events in a sample (sample variation) and after several repetitions (variation among samples), (6) as the number of samples increases, students can visually detect the event with the maximum frequency, and compare frequencies to recognize key patterns in the data (sample variation; regularity), and (7) students assign a level of confidence to each event that describes its possibility of occurrence (a language with levels of uncertainty) to draw inferences beyond the obtained data by imagining the behavior of the next sample (frequency distribution; sampling variation; regularity).

Online participation did not seem to hinder students' oral expression of informal inferences related to statistical situations or the teacher-student interaction. The students exhibited a data sense as evidenced by the fact that when they solved the data-based problems, they understood how the data were obtained, they recognized the variation of data, they made predictions based on the behavior of the data, and they made judgements and informal statistical inferences on the basis of the data without performing calculations (cf. Estrella, 2018; Estrella et al., 2021).

One of the results shows that after the teaching activities, more than half of the students had difficulty attaching a level of certainty to their inferences. This highlights the need for more research on lesson designs that encourage dialogic conversations about sampling issues, such as variability and representativeness, and confront students with the impossibility of making inferences with absolute certainty.

Among the limitations of this LT is the lack of size increases for each sample, which could cause bias due to the law of small numbers. Therefore, a step should be incorporated that allows students to recognize the effect of sample size, namely, that repeated sampling with a larger sample size reduces the variation in estimates for a broader group and, therefore, yields a more accurate inference (e.g., Estrella et al., 2023; Manor Braham & Ben-Zvi, 2017).

Although there is growing interest in developing richer understanding of children's early statistical thinking (e.g., Frischemeier, 2020; Leavy et al., 2018), this area of research continues to be incipient in the early grades of school (e.g., Estrella et al., 2022). This research can contribute to teachers' understanding of ISI topics and how they can help their students to develop these ideas since the LT proposal for early IIR has allowed students to draw conclusions beyond the data obtained from their samples and express these conclusions qualitatively with certain levels of confidence that consider their reliability.

ACKNOWLEDGEMENTS

This research was conducted with financing from ANID/CONICYT FONDECYT 1200346 and National Doctorate Scholarship ANID 21210862.

The authors are grateful for the helpful comments of the anonymous reviewers of earlier versions of this manuscript.

REFERENCES

Arnold, P., Confrey, J., Jones, R. S., Lee, H., & Pfannkuch, M. (2018). Statistics learning trajectories. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 295–326). Springer. https://doi.org/10.1007/978-3-319-66195-7_9

Arnold, P., Pfannkuch, M., Budgett, S., Regan, M., & Wild, C. J. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education*, *19*(2). https://doi.org/10.1080/10691898.2011.11889609

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

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, *88*, 291-303. https://doi.org/10.1007/s10649-015-9593-3

English, L. D., & Watson, J. M. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education*, *47*(1), 28–62. https://doi.org/10.5951/jresematheduc.47.1.0028

Estrella, S. (2018). Data representations in early statistics: Data sense, meta-representational competence and transnumeration. In A. Leavy, A., M. Meletiou-Mavrotheris & E. Paparistodemou (Eds.). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking* (pp. 239–256). Springer. https://doi.org/10.1007/978-981-13-1044-7_14

Estrella, S., Méndez-Reina, M., Olfos, R., & Aguilera, J. (2022). Early statistics in kindergarten: Analysis of an educator's pedagogical content knowledge in lessons promoting informal inferential reasoning. *International Journal for Lesson and Learning Studies*, *11*(1), 1–13. https://doi.org/10.1108/IJLLS-07-2021-0061

Estrella, S., Méndez-Reina, M., Salinas, R., & Rojas, T. (2023). The mystery of the black box: An experience of informal inferential reasoning. In G. Burrill, L. De Oliveira & E. Reston, *Research on reasoning with data and statistical thinking: International* perspectives (pp. 191–210). Springer.

Estrella, S., Vergara, A., & González, O. (2021). Developing data sense: Making inferences from variability in tsunamis at primary school. [El desarrollo del sentido del dato: haciendo inferencias desde la variabilidad de los tsunamis en primaria]. *Statistics Education Research Journal*, *20*(2), 1–14. https://doi.org/10.52041/serj.v20i2.413

Frischemeier, D. (2020). Building statisticians at an early age: Statistical Projects exploring meaningful data in primary school. *Statistics Education Research Journal*, *19*(1), 39–56. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics*, *88*, 327–342. https://doi.org/10.1007/s10649-014-9541-7

Konold, C., Higgins, T., Khalil, K., & Russell, S. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*, 305–325. https://doi.org/10.1007/s10649-013-9529-8

Langrall, C., Nisbet, S., Jansem, S., & Mooney, E. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning*, *13*(1–2), 47–67. https://doi.org/10.1080/10986065.2011.538620

Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (Eds.). (2018). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking*. Springer. https://doi.org/10.1007/978-981-13-1044-7

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, *18*(1), 1–24. https://doi.org/10.1080/10986065.2016.1107820

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105. http://dx.doi.org/10.52041/serj.v8i1.457

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8

Maloney, A., Confrey, J., & Nguyen, K. (Eds.). (2014). *Learning over time: Learning trajectories in mathematics education*. Information Age Publishing. https://www.doi.org/10.1126/science.1204537

Manor Braham, 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. http://dx.doi.org/10.52041/serj.v16i2.187

Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics*, *88*(3), 385–404. https://doi.org/10.1007/s10649-014-9551-5

Moore, D. S. (2000). *The basic practice of statistics* (2nd ed.). W. H. Freeman.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education*. Proceedings of the seventh International Conference on Teaching Statistics (ICOTS7), Salvador, Bahia, Brazil. https://iase-web.org/documents/papers/icots7/6A2_PFAN.pdf?1402524965

Pfannkuch, M., Arnold, P., & Wild, C. J. (2015). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics*, *88*, 343–360. https://doi.org/10.1007/s10649-014-9539-1

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, *7*(2), 107–129. http://dx.doi.org/10.52041/serj.v7i2.472

Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5–19. https://doi.org/10.52041/serj.v7i2.467

Saldanha, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, *51*(3), 257–270. https://doi.org/10.1023/A:1023692604014

Shaughnessy, J. M. (2019). Recommendations about the big ideas in statistics education: A retrospective from curriculum and research. *Cuadernos de Investigación y Formación en Educación Matemática*, *18*, 44–58. http://funes.uniandes.edu.co/21613/1/Shaughnessy2019Recommendations.pdf

Silvestre, E., Sánchez, E. A. & Inzunza, S. (2022). The reasoning of high school students about repeated sampling and the empirical sampling distribution. *Educación Matemática*, *34*(1), 100–130. https://doi.org/10.24844/em3401.04

Simon, M. (1995). Reconstructing Mathematics Pedagogy from a Constructivist Perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145. http://dx.doi.org/10.2307/749205

Simon, M. (2018). An emerging methodology for studying mathematics concept learning and instructional design. *Journal of Mathematical Behavior*, *52*, 113–121. https://doi.org/10.1016/j.jmathb.2018.03.005

van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2020). Repeated sampling with a black box to make informal statistical inference accessible. *Mathematical Thinking and Learning*, *22*(2), 116–138. https://doi.org/10.1080/10986065.2019.1617025

van Dijke-Droogers, M., Drijvers, P. & Bakker, A. (2021). Introducing statistical inference: Design of a theoretically and empirically based learning trajectory. *International Journal of Science and Mathematics Education*, 1–24. https://doi.org/10.1007/s10763-021-10208-8

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

Zieffler, A., Garfield, J., DelMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40–58. https://doi.org/10.52041/serj.v7i2.469

SOLEDAD ESTRELLA
Pontificia Universidad Católica de Valparaíso
soledad.estrella@pucv.cl