REPEATED SAMPLING AS A STEP TOWARDS INFORMAL STATISTICAL INFERENCE

Marianne van Dijke, Paul Drijvers and Arthur Bakker
Utrecht University, The Netherlands
m.j.s.vandijke-droogers@uu.nl

Recent literature suggests that secondary school students should learn about informal statistical inference (ISI) as preparation for formal statistical inference. Not enough is known, however, how to realize this. The design study reported here aimed at developing a theoretically and empirically grounded learning trajectory (LT) on ISI for 9th-grade 14-year-old students. The LT focused on sampling, the frequency distribution of results from repeated samples, and sampling distribution. The current paper reports on the part of the LT on using the frequency distribution of repeated samples as tested with twenty students. The results showed that these students could imagine and sketch the frequency distribution of repeated samples, which indicates a step forward in their understanding of variation and uncertainty in drawing informal statistical inferences.

NEED FOR INFORMAL STATISTICAL INFERENCE

The use of sampling to draw inferences about an unknown population is at the heart of statistics, and therefore important to learn. Statistical inference includes both a generalization and an argument about the likelihood of the formulated generalization. Learning and teaching about inference is a key concern of statistics education (Pratt & Ainley, 2008). Too strong or too early an emphasis on procedural and formal knowledge, however, may result in students' inability to interpret results (Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007), or to understand statistical concepts such as uncertainty, probability, and sampling (Konold & Pollatsek, 2002).

To address this issue, statistics educators propose to focus on informal statistical inference (ISI) first. Makar and Ruben (2009) identify three key principles of ISI: (1) *generalize* beyond data; (2) *data as evidence* for these generalizations; and (3) *probabilistic reasoning* about the generalization. Recent research suggests that informal statistical inference activities at an early age may prepare for the understanding of more formal procedures (Zieffler, Garfield, delMas, & Reading, 2008). ISI involves using informal statistical knowledge to support inferences about an unknown population based on observed samples (Zieffler et al., 2008). Various studies have shown that ISI can be developed by young students (Makar, 2016; Paparistodemou & Meletiou-Mavrotheris, 2008). However, how these promising results can be translated into a learning trajectory (LT) for grade 9 (14 years old) remains largely unknown. Therefore, this study aims to develop a prototypical LT on ISI for grade 9.

THE DESIGN OF THE LEARNING TRAJECTORY

The trajectory as a whole

Because educational materials focusing on the development of statistical reasoning for grade 9 in the Netherlands hardly exist, we had to design them. In this design study, a hypothetical learning trajectory (Simon, 1995) was developed. Guidelines were identified through literature review, and the opportunities of educational software were explored. This resulted in the design of a LT. An overview of designed teaching activities for each LT step is given in Figure 1. Hypotheses about students' learning were formulated for each step of this LT. In this paper we focus on the LT's second step, in which students investigated what happens if a sample is repeated.

The second step: "What happens if this experiment is repeated many times?"

This LT step follows the physical "Black Box" experiment in step 1, in which students examined the proportion of yellow balls in a box by counting the number of visible balls in the viewing window. Performing and exchanging data allowed students to become aware of sampling variation, the effect of sample size and repeated sampling during step 1.

As a next step, students were invited to think about the most likely sample results if the physical experiment was repeated many times. The hypothesis for step 2 was that students would understand that most sample results will be close to the population proportion and that strong

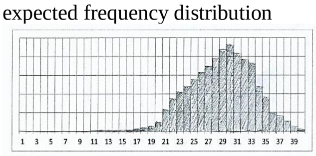
In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

deviations are unlikely. To conceptualize this idea, students were first asked to think about the question "What happens if this experiment is repeated?" This question is paramount in understanding statistical inference (Rossman, 2008). Also, students' involvement and statistical reasoning is expected to increase if they make predictions (Bakker, 2004). During a whole-class discussion, the students shared their expectations for the number of yellow balls in a sample of 40 from a box containing 750 yellow and 250 non-yellow balls, and discussed the boundaries of possible sample results. Subsequently, students were asked to sketch the expected frequency distribution of repeated samples if the experiment were repeated 100,000 times. As a follow-up, students were asked in the second step to determine the probability of a certain sample result based on their sketch of the frequency distribution. This approach meets various studies (Rossman, 2008; Watson & Chance, 2012), that indicate students can use the frequency distribution on repeated sampling to investigate and informally determine the probability of certain sample results.
This second step prepared for probabilistic reasoning with the sampling distribution in step 3, in which students simulate frequency distribution in many repeated samples (Chance, Ben-Zvi, Garfield & Medina, 2007; Watson & Chance, 2012).

Step	1 Conduct physical experiment (Black Box with small and large window)	2 Examine frequency-distribution >100.000 repetitions of physical experiment	3 Simulate sampling distribution of physical experiment *(with ICT)*	4 Test a claim with sampling distribution (Social Media) *(with ICT)*	5 Conduct physical experiment (Black Box with notes)	6 Examine sample size and repeated sampling *(with ICT)*	7 Test a claim with sampling distribution *(with ICT)*	8 Compare groups based on samples *(with ICT)*
Data	Ordinal – nominal level of measurement				Interval – ratio level		All levels of measurement	
Con-cept	Sampling variation	Repeated sampling + Frequency distribution	Sample size + Sampling distribution		Sampling variation	Sample size + Repeated sampling	Sampling distribution	
Proba-bilistic reason-ing	In words	In words + argument frequency distribution	In words + argument sampling distribution		In words	In words + argument sample size	In words + argument sampling distribution	

Step 1:
Estimation of content accompanied by probabilistic reasoning based on sampling variation

Step 2:
Determine most likely sample results accompanied by probabilistic reasoning with the expected frequency distribution



Step 3:
Probabilistic reasoning with the sampling distribution

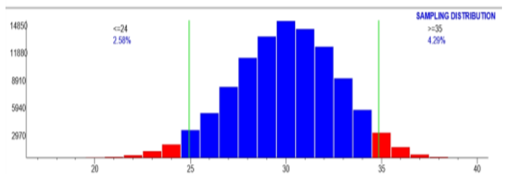


Figure 1: Overview of designed teaching activities for each LT step

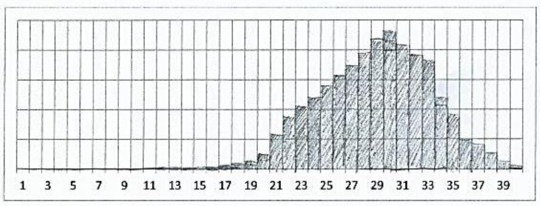
METHOD

The complete intervention comprised a series of ten 45-minute lessons and was piloted in one class, which was taught by the first author. The participants consisted of twenty 14-year-olds in grade 9 on pre-university level. The second step was carried out in lesson 2 and included one lesson. The collected data in this step were video data from classroom discussion, students' worksheets and teacher notes. The data analysis focused on identifying what students already knew and determining whether the intervention indeed stimulated ISI. With regard to the results, we focus on the analysis of students' work.

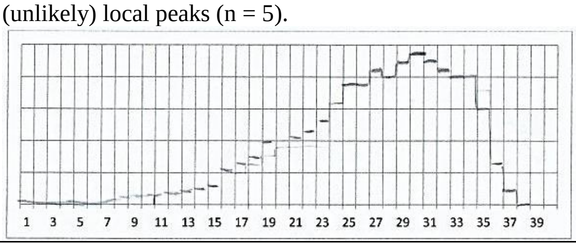
RESULTS

After the group discussion on the question "What happens if this experiment is repeated?" students were asked to sketch the expected frequency distribution of the number of yellow balls in a sample of 40, if the first-step physical experiment were repeated 100,000 times. The students were given a coordinate system with the values 0 to 40 along the horizontal axis and no values vertically. Students' work showed they were able to imagine and sketch the results of repeated sampling in a frequency distribution. In doing so, they demonstrated that samples vary, but that a sample result that resembles the population proportion (75%) will occur most frequently in more than 100,000 repetitions. No students placed values along the vertical axis. The drawings could be divided into the four types shown in Figure 2. Eleven students sketched a bar diagram with a top at 30 and a negative skew (type 1). Five students indicated that for so many repeated samples, the sample results will not increase/decrease monotonously, but local peaks may occur (type 2). They probably thought that chance plays a role in this and did not (yet) realize that the distribution of samples will stabilize on so many repeated samples (known as the law of large numbers, which students did not yet have to understand here). Two students sketched an almost linear course (type 3) and two others a smooth curve (type 4). However, the latter may be caused by the word *sketch* (rather than *draw*).

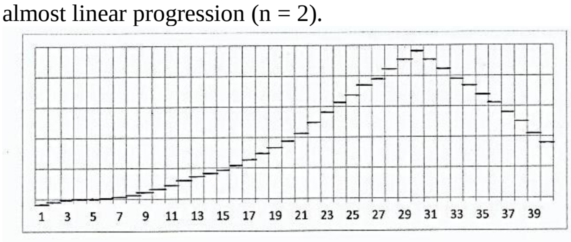
Type 1: Well-sketched out frequency distribution in the form of a bar graph with a top at 30 and a negative skew (n = 11).



Type 2: Well-sketched out global form of the frequency distribution in a bar graph with a top at 30, but with (unlikely) local peaks (n = 5).



Type 3: Well-sketched out global form of the frequency distribution in a bar graph with a top at 30, but with almost linear progression (n = 2).



Type 4: Well-sketched out global form of the frequency distribution with a top at 30, but with an unrealistic smooth line (continuous distribution) (n = 2).

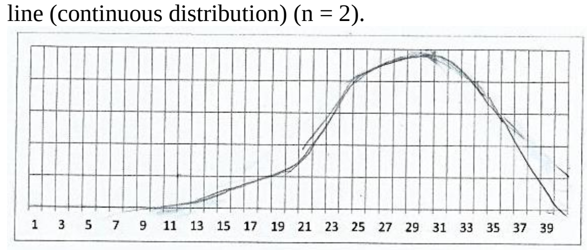


Figure 2: Four types of students' sketches (N=20) of the expected results of repeated sampling (100,000 repetitions) with sample size 40 in a frequency distribution

It is worth noting that when estimating the probability of a certain sample result, all students recorded their estimates as a percentage. It was expected they would describe this probability in words, but they apparently felt a need to quantify it (probably because this activity is part of the mathematics lesson) and chose to use percentages. All students estimated the probability of a sample result under 10 less than or equal to 10%, with only six out of twenty students estimating this probability close to zero (see Table 1). This indicates that students understood that the probability of a strong deviation is small. However, this probability was overestimated, as six students indicated that it would be 5% or higher. The frequency distributions sketched by the students did not always correspond to their estimations. Although the students wrote down a numerical value, it seemed to be based on their intuitive idea of probability and was not calculated using the frequency distribution. Some students mentioned explicitly that they calculated the probability: "I estimate the probability at 0.01% which is about 10 out of 100,000 times."

Table 1: Students' work (N=20)

Question	Probability	Probabilistic reasoning (examples of students 'work)
1. How do you estimate the probability of a sample result (amount of yellow balls) of less than 10 at a sample size of 40? *(population proportion 75%)*	(Almost) 0% (n = 6)	very small, but it is possible though
	1% (n = 6)	a very small probability actually, almost 1%, because there are simply many more yellow than orange balls
	5% (n = 3)	.. because there will always be a chance, only he gets less because the larger majority has that color
	10% (n = 3)	75 out of 100 balls are yellow
	Empty (n = 2)	*(they had no time to write an answer)*

CONCLUSION

In step 2, students were invited to think about the most likely sample results as the physical experiment is repeated a large number of times. Our hypothesis was that students would understand that most sample results will be close to the population proportion and that the probability of a strong deviating result is small. The intervention data of students' work from lesson 2 show that:

- students were aware of sampling variability;
- students could imagine the frequency distribution of 100,000 repeated samples;
- students understood that strong deviating sample results were less common, but …
- students overestimated the probability of these deviating sampling results;
- the estimated probability of a particular sample result (for example less than 10 yellow balls) did not always correspond to its sketched frequency distribution.

Key elements in the LT that promote this step seem to be (1) the strong connection between the physical experiment in step 1 and the frequency distribution in step 2; (2) sharing expectations and discussing the boundaries of possible sample results during a classroom discussion; (3) estimating the probability of certain sample results based on the outlined frequency distribution. A point of attention was probabilistic reasoning with the frequency distribution. Students tended to overestimate the probability of a strongly deviating sample result and seemed to be intuitively estimating probability instead of using the frequency distribution. Our advice is to link students' prior knowledge on ratios to reasoning with the frequency distribution. A way to achieve this is by calculating proportions, for example 100 out of the 100,000 samples is 0.1%.

Our preliminary theory is that strengthening probabilistic reasoning with frequency distribution of a physical experiment is an essential step towards reasoning with sampling distribution. The results showed that these students could imagine and sketch a frequency distribution of repeated samples with most sample results close to the population proportion and in which strong deviations hardly occur. Consequently, students' understanding of the role of variation and uncertainty in drawing informal statistical inferences has been strengthened. As such, this intervention is a promising support of the designed LT for ISI for 9th-grade students.

REFERENCES

Bakker, A. (2004). *Design research in statistics education. On symbolizing and computer tools; A study in statistics education*. Utrecht: Universiteit Utrecht.

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98–113.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal*, *1*(1), 1–23.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 259–289.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, *18*(1), 1–24.

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

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

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

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

Watson, J., & Chance, B. (2012). Building intuitions about statistical inference based on resampling. *Australian Senior Mathematics Journal*, *26*(1), 6–18.

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