BRIDGING FREQUENTIST AND CLASSICAL PROBABILITY THROUGH DESIGN

Amanda Provost
Montclair State University
provosta1@montclair.edu

Su San Lim
Montclair State University
lims1@montclair.edu

Toni York
Montclair State University
yorka1@montclair.edu

Nicole Panorkou
Montclair State University
panorkoun@montclair.edu

The frequentist and classical models of probability provide students with different lenses through which they can view probability. Prior research showed that students may bridge these two lenses through instructional designs that begin with a clear connection between the two, such as coin tossing. Considering that this connection is not always clear in our life experiences, we aimed to examine how an instructional design that begins with a scientific scenario that does not naturally connect to theoretical probability, such as the weather, may support students' bridging of these two models. In this paper, we present data from a design experiment in a sixth-grade classroom to discuss how students' shifts of reasoning as they engaged with such a design supported their construction of bridges between the two probability models.

Keywords: Probability, Design Experiments, Middle School Education, Integrated STEM.

There are distinct views of probability in the literature which differ not only in the way they define probability but also in the nature of their emerging solutions to the solving of problems (Batanero et al., 2016). The classical view of probability is one of the earliest approaches and is connected to chance in games. From this view, probability is considered as a fraction of the number of favorable cases divided by the number of all possible cases. This definition was created on the assumption that "all possible elementary events were equiprobable" (Batanero et al. 2016, p. 5), which is applicable to most situations in games. However, this view has been criticized because the idea of equiprobable outcomes is not always valid in natural phenomena.

The other most common approach, the frequentist view, sees probability as a convergence of relative frequency when a random experiment is repeated infinitely many times. According to Cosmides and Tooby (1996), humans have more experience with encountered frequencies in their observation of the world, therefore students would be more receptive to frequentist probability where data is collected through experience. In contrast to the connection to real-life experiences, the frequentist ideas of probability are not appropriate when discussing a single event or when the experiment cannot be repeated multiple times under the same conditions. This frequentist view is only an estimation of probability that results from a series of repetitions.

Considering the above, Lee et al. (2010) claim that since most everyday probabilistic situations (such as weather forecasting) "do not allow for the classical approach to probability" (p. 91), it is necessary for students to examine situations where it is possible to both calculate classical theoretical probability and make the connection to the frequentist experimentally collected empirical probability. Similar to Lee et al. (2010), many researchers have claimed that developing a connection between the classical and frequentist views is helpful for students to fully grasp the concept of probability (e.g., Henry & Parzysz, 2011; Ireland & Watson, 2009). During the learning process, students need to distinguish between the two models and understand when each one can be used for solving problems (Batanero et al., 2016).

Prodromou (2012) argued that while the two models differ, they are complementary and not mutually exclusive. She engaged pre-service teachers (PTs) in dice rolling tasks related to both theoretical and experimental probability and examined how they developed a bidirectional relationship between the two, explaining that "PTs perceived the theoretical probability as the intended outcome and the experimental probability as the actual outcome. In the opposite direction, PTs considered the theoretical probability as the target towards which the experimental probability is directed" (p. 866). This notion was also highlighted by Stohl and Tarr (2002), who reported that the tasks that engaged students in bidirectionality of probability helped students to develop robust inferences about probability. To engineer this bidirectionality, they started with coin tossing and dice rolling designs, and then used three different computer simulation tasks, two of them using experimental probability as a tool to evaluate theoretical probability and the other using theoretical probability as a tool to anticipate the result of the experiment.

Many supported that the connection between classical and frequentist probability is grounded in the Law of Large Numbers (LLN) (e.g., Drier, 2000; Prodromou, 2012; Stohl & Tarr, 2002), which states that larger numbers of trials performed for a given event lead the relative frequency for that event to approach the theoretical probability. A study by Aspinwall and Tarr (2001), examined how sixth graders' understanding of experimental probability related to sample size and the LLN using designs with flipping chips, spinners, and dice. They demonstrated that probability simulations can challenge students' preexisting conceptions. However, they also noted that there were challenges in developing the understanding of the role that sample size played in determining experimental probability through the LLN. Aspinwall and Tarr (2001) claim this theorem has been shown to be challenging and nonintuitive for students. Students may not recognize when to use the LLN (Fischbein & Schnarch, 1997), or they may believe that the LLN applies to small numbers as well (Tversky & Kahneman, 1971).

Researchers found that offering students multiple representations of data supports their understanding of the connection between theoretical and experimental probability. Stohl and Tarr (2002) reported that the use of graphs and tables enables students to see bidirectionality of probability by involving them in representing and analyzing data with different forms. Similarly, Ireland and Watson (2009) claimed that multiple representations in computer simulations can help students perceive how theoretical probability is connected to experimental probability. In their experiment with digital mixers and spinners, students made a connection between theoretical and experimental probability by comparing data in various representations.

Using computer simulations as a mode of representation provides students with the opportunity to collect large amounts of data in shorter periods of time, which addresses the limitations of time constraints and resources that Biehler (1991) argued students encounter as they explore the concept of probability. Computer simulations also support students' connection-making between classical and frequentist approaches of probability (Ireland & Watson, 2009; Prodromou, 2012). Abrahamson and Wilensky (2005) used modified coin flipping computer simulations to help students develop their understanding of experimental probability by bridging the gap of theoretical probability with simulating and collecting large amounts of experimental data. Paparistodemou (2005) conducted a study where students were able to manipulate aspects of the computer environment to affect the generation of random events. The students manipulated the simulation in multiple ways to make use of the LLN to achieve the target probability.

While many studies have examined how students construct bridges between the two models of probability, their designs involved beginning with a clear connection between the two, such as coin tossing or dice rolling scenarios (e.g., Abrahamson & Wilensky, 2005; Prodromou, 2012). However, our real-life experiences of probability, such as predicting the weather, do not always have clear theoretical probabilities that can be calculated. Consequently, we started with the conjecture that it is possible to support students' bridging of these two models by first engaging them with probability scenarios situated in the context of science. We hoped that this would make use of students' realistic experiences to construct more meaningful connections between the two models. Specifically, we explored: *(a)* What kind of design that starts with the frequentist perspective without a clear connection to the theoretical probability and moves to the classical perspective would support students' bridging between the two models? *(b)* How may students' reasoning progress as they shift between the frequentist and classical perspectives?

Methods

In this paper we report on the results of a whole-class design experiment (Cobb et al., 2003) in a sixth-grade classroom in the Northeastern U.S. The class met in ten 15- to 50-minute sessions via Google Meet due to COVID-19 restrictions. To test our conjecture we designed two simulations, one based on the frequentist and one on the classical model, along with investigation and interview questions. We chose a scientific context because our previous work (e.g., Panorkou & Germia, 2021; Panorkou & York, 2020) showed that sixth-grade students can engage meaningfully with mathematical concepts in a setting integrated with science concepts. We focused on the topic of weather because research showed that students consider weather events as a relevant context for probability discussions (Chick & Baker, 2005).

The Weather Forecast simulation (Figure 1) is based on the frequentist model and generates the results of an imaginary weather forecast. The chosen Data Set determines the forecasted percentages of days that are expected to be rainy and sunny. The Run Size determines how many times the simulation runs the experiment. The larger the chosen Run Size the more accurate the resulting forecast will be for the chosen Data Set. After exploring this simulation, students were asked to gather data about the results of different Run Sizes in a table and graph these values on a log-scale plot of Percentage of Rainy Days versus Run Size to observe how the results tend towards certain probabilities for each Data Set as the Run Size increases.

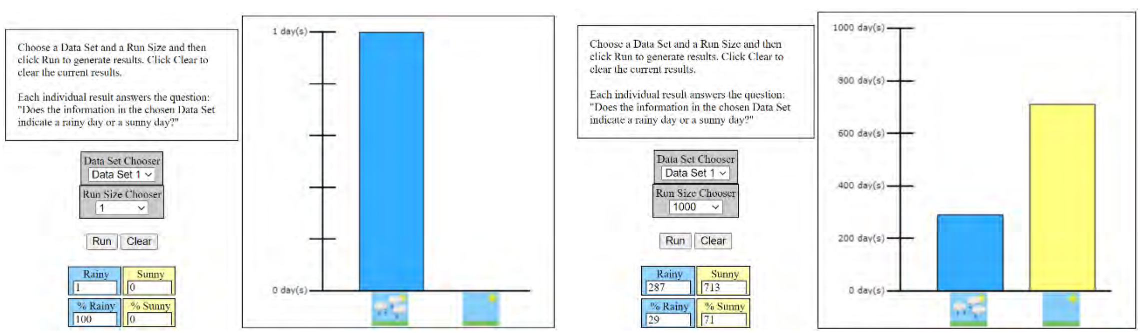

Figure 1: The Weather Forecast Simulation Showing Different Run Sizes for Data Set 1

The Chance of Rain simulation (Figure 2) is based on the classical model and represents weather data for every day during the month of June for 20 years in a certain location. The data is not based on a real location but was instead chosen to provide certain ratios of the different outcomes. Each of the 600 individual data points indicates if that day was sunny, cloudy, rainy, or stormy. The student can view a random day, a specific day, an entire specific month, or a bar chart summarizing all 600 days. As in the classical model, the probability of each weather outcome for any given day can be described by a fraction with the total number of days in the denominator. After exploring the simulation, the students were asked to gather data by clicking on the View Random Day button 100 times, recording what the weather was for each of these random days, and then comparing this to the probability fractions for each type of weather.

Figure 2: The Chance of Rain Simulation Showing the Different Data View Screens

In this paper, we focus on the analysis of one pair of students, Violet and Anne, to describe a chronological account of the progression of their reasoning about probability and the design decisions that the researcher made to support the development of their reasoning further. Of the three pairs of students recorded, this pair was chosen for an initial analysis due to completeness of data and their work with the research team member who was also part of the design team.

Findings

At the beginning of the design experiment, Violet and Anne were asked to state if they knew anything about weather reports or the concept of chance in predicting rainy or sunny weather. Violet made a connection to a recent snow day and mentioned weather forecasters talking about the chance of snow, saying, “There’s a 50% of snow, 50% chance that snow might come. Well, when they say that they're not completely sure if it’s going to come though.” Her reasoning shows a classical understanding of probability as a percent.

Next, students were asked to explore the Weather Forecast simulation. First, they examined Run Size 1 for Data Set 1 (30% rainy, 70% sunny) and identified that it was sometimes rainy and sometimes sunny. As Violet said, “it only got rainy two times so it’s not always rainy” showing the classical perspective of considering two rainy times out of the total number of runs. They then tried Run Size 5 and identified that it was now more likely for the predictions to be sunny than rainy. As Violet stated, this is because “it mostly shows that it’s going to be more percent chances of being sunny than rain,” illustrating that she was starting to notice a pattern in the data. Subsequently, students were asked to explore the larger Run Sizes (10, 50, 100, 500, 1000, 5000, 10000, 50000, 100000) and observed that there are usually more sunny than rainy days. However, they were not yet ready to bridge the connection to a specific theoretical probability.

In the next task, students were asked to use the simulation to record data in a table with the percent rainy and percent sunny for a single run for each size (Figure 3). When asked if they saw a pattern, Violet replied “I got 30 and 70, 3 times. … But I’m a little surprised I got the same one three times because I thought I was gonna get different more but then I got the same.” This collection of data was challenging Violet’s preconceptions about the expected probability.

They then graphed the data from their tables and compared their graphs to each other (Figure 3). Violet identified that their graphs were similar, but not the same: “When I got to 100 my number, her number, like the Run Size started to get a little different from 100 to 1000. But then when we got into the even bigger numbers, we had the same again.” Violet was beginning to identify patterns in the Run Sizes, but the limitations of collecting random data was impeding her ability to make a claim that larger Run Sizes are more precise than smaller Run Sizes.

Run Size	% Rainy	% Sunny
1	0%	100%
10	30%	70%
100	28%	72%
1,000	31%	69%
10,000	30%	70%
100,000	30%	70%

Run Size	% Rainy	% Sunny
1	0%	100%
10	30%	70%
100	29%	71%
1,000	28%	72%
10,000	30%	70%
100,000	30%	70%

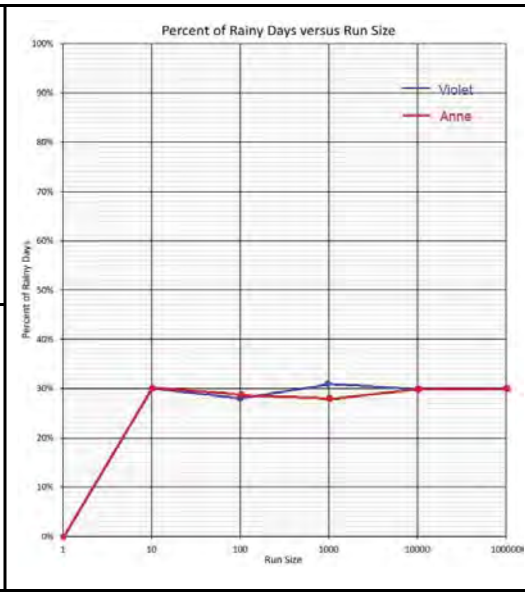

Figure 3: Recreation of Tables and Graphs of Data Set 1

Run Size	% Rainy	% Sunny
1	100%	0%
10	90%	10%
100	93%	7%
1,000	90%	10%
10,000	90%	10%
100,000	90%	10%

Run Size	% Rainy	% Sunny
1	100%	0%
10	100%	0%
100	88%	12%
1,000	90%	10%
10,000	90%	10%
100,000	90%	10%

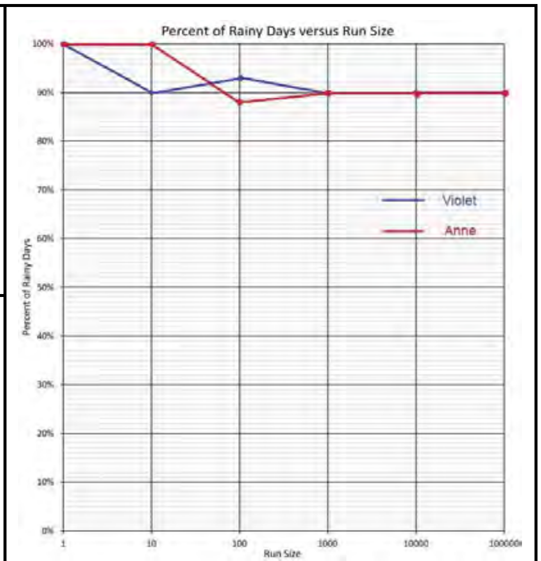

Figure 4: Recreation of Tables and Graphs of Data Set 2

The students went through the same process for Data Set 2 (90% rainy, 10% sunny). As Figure 4 shows, Anne got 100% rainy for Run Size 1 and 10, and 90% rainy for Run Sizes 1,000 to 100,000. When asked to make a numerical prediction about the chance of rain Violet replies, “there is 100% chance in the beginning of the day in the graph. But then it gets lower. So, I’ll put it between, like 99 and 100%, it’s gonna rain.” Her reasoning shows that she erroneously considered the x-axis to represent time of day rather than Run Size and this led her to struggle in identifying what the different probabilities represented. The researcher then intervened to clarify this and asked her to explain what happens when the Run Size increases:

> Violet: I think you should run the simulator, seven times. Or a lot of times, because the number you get the most is probably the percentage of rain or the chances of rain, you're gonna get that day. Because if you run it one time, you're probably not going to get the right answer. Cause it could say that it's going to be 100% sunny when you run it one time, but it's probably going to be rainy that day. But when you run it like seven or more times, you're gonna get probably the same number like two or three times. But if you get it more than that, it means that it's probably going to be that weather that day.

Violet identified that there was a "right answer" or target theoretical probability that the experimental data should approach. She was also able to observe the limitations of small Run Sizes. However, for her a Run Size of "seven" or "a lot of times" was large enough to make a claim about the effect of the increased Run Size.

At this point the researcher modified the original design and discussed coin flipping and dice rolling aiming to examine whether the focus on the classical approach would help them in developing their frequentist understanding. Violet was able to bridge the connection between flipping a coin and its theoretical probability by saying there was "a 50% chance because there's like half a chance you're going to land on tails." They struggled with understanding the theoretical probability of the dice, but Violet used her understanding of experimental probability to bridge the connection to the theoretical probability by explaining the chances of rolling a one:

> Violet: Yeah, six possibilities, you can get it, but most of the time, you're not gonna get it. It's like, 10 out of 100, that you're probably going to get one. When you roll the dice, you're not going to get the number, most of the time, you're not going to get the number you wanted.

After this introduction to classical probability, they engaged with the Chance of Rain simulation and used the bar graph (Figure 2) to calculate the theoretical probability for each weather type in June (Figure 5, left). The students then collected 100 data points using the View Random Day button (Figure 5, right).

Weather	Days	Fraction of Total Days
Sunny	300	300/600 = 1/2
Cloudy	150	150/600 = 1/4
Rainy	75	75/600 = 1/8
Stormy	75	75/600 = 1/8

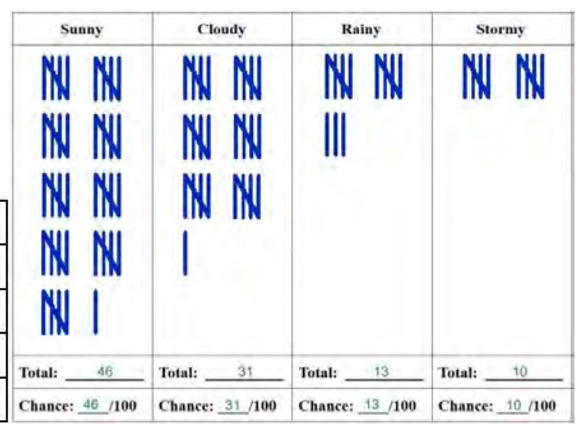

Figure 5: Theoretical and Experimental Probability for Chance of Rain Investigation

Working with the probabilities in fraction form required that they use strategies to compare fractions, which made connecting the probabilities more difficult. They discussed rainy and stormy which each had a theoretical of "one-eighth" and then created 12.5/100 as an equivalent fraction to compare it to their experimental 10/100 and 13/100. As Violet stated, "I tried simplifying some of them to see if I got closer to the same number." They were able to connect the experimental to the theoretical probabilities. As Anne stated, "they're all in the same thing cause like 10, if it was 11 it would have been more closer, like 11, 12, 13. But I think they're like, equal." They were then asked why they collected 100 data points, and not more or less:

Anne: So that you could get a good amount of like, sunny, stormy, rainy, and cloudy. And, it won't have to only be sunny, cause maybe if you did 10, it would maybe be like 30 sunny maybe like two stormy, three rainy.

Violet: I do prefer 100 more than 10. But I tried it out. I clicked the view Random Day button 10 times. And I actually got rainy, stormy and cloudy. But I only got like two chances of sun.

They both saw the limitations of a small sample size and Violet even used experimental data to help support her point. At this point the researcher brought up the example of coin flipping again in another modification of the experimental design, to help illustrate what happens as the sample size increases. They discussed how it was more believable that they would get heads repeatedly with small sample sizes, but Violet explained:

Violet: When the number gets bigger, the more unbelievable it gets that you're gonna get tails, or heads like that much. I mean, it could be possible to get heads or tails, like three or two times the same. … Because if I take a few minutes to flip my small watch 100 times, I'm most likely going to get tails and heads at the same time. Because like I said, there's two sides of a coin.

Violet showed some understanding of the LLN and the advantages of larger sample sizes by using her understanding of experimentally flipping a coin and connecting it to the target theoretical probability of flipping a coin, bridging the frequentist and classical perspectives. Anne then showed similar reasoning when talking about the possibility of a computer simulation flipping heads 100 times in a row, saying, "there might be a 50% chance that the computer might just get like heads 100 in a row and then there's like the other 50% chance," showing that with large samples, she would expect the results to approach the target theoretical probability.

At this point they were asked to return to the Weather Forecast to continue their discussion on the LLN in a modification of the design. The researcher bridged the transition by continuing the discussion on the 100 data points collected and what they thought would happen if they collected more data. Both students agreed that collecting more than 100 would have been better with Violet saying, "if we did 200, the higher we go, I think it might be the easiest to compare them," and Anne saying, "maybe if we did more than 100 maybe it would be like, close to each other. A little bit." They had both explained the limitations of smaller sample sizes and agreed that larger sample sizes could make it easier to compare to the theoretical target.

However, when asked to draw conclusions about what happens as the Run Size gets larger from the graphs of the Data Sets, they still had difficulty connecting those thoughts to their analysis of the graphs. As Violet stated, "sometimes it goes higher, and then lower, and then it's gonna go higher," and Anne said, "it was one at the bottom, then at the top, then at the center, then at the top again." Here they were focused on the variation of the data points and had not developed the covariational reasoning to connect their data to the Run Size. The researcher decided to prompt them to discuss the 'gaps' between their data points (Figure 4), which was a productive design modification:

Violet: [Looking at the right of the graph of Data Set 2, Fig. 4] Well, for the first three, there's not as big a gap because they're all 90. [Looks at the left of the graph] That's a big gap. Because for the last one here, it's 89. But this one's 100. That's sort of a big gap.

Anne: [Referring to Data Set 1, Fig. 3] Because you were comparing like there was a big gap or if they were closer to each other. Most of them were closer to each other, especially the 30 and the 29 and stuff.

This discussion was productive in helping them distinguish the pattern with smaller Run Sizes compared to larger Run Sizes. Violet was able to summarize this:

Violet: Because as larger as it gets, it shows how likely the number you picked is going to be the answer. Like in my graph as the numbers got bigger, it showed 30, two times. So that means it's gonna be 30% chance of rain. So as larger as the number is, is how it's going to be or how it may be. … Because the run number did get larger, and it did show why it's gonna be that percentage.

Violet's explanation shows her understanding of the LLN and how it can be used to approach the target theoretical probability or "answer." Anne mostly agreed with Violet's thinking but questioned what amount of data was sufficiently large. Despite this continuing limitation, Violet and Anne developed their understanding of experimental probability to predict a theoretical probability using the LLN, through their bidirectional engagement with the two simulations and the use of multiple representations of probability.

Discussion

Our results show that our design that started with a frequentist perspective without a clear connection to theoretical probability and moved to the classical perspective supported students' construction of a bidirectional link (Prodromou, 2012) between the two probabilities. While engaging in the simulations, the students' learning process was not linear but rather it was their transitions between the different representations that helped them develop bridges of the two models. Through these transitions, students developed an understanding of the LLN by emphasizing the strength of large numbers and the limitations of smaller Run Sizes.

The scientific context of the weather showed to be productive in illustrating the utility of probability for understanding phenomena in the real world. Bringing in examples of flipping coins and tossing dice helped bridge the gap for students in connecting the two probability models. Interaction with data by engaging with multiple modes of representations, including simulations, tables, and graphs, also fostered the students' understanding of this connection.

This experiment was limited as we only analyzed the reasoning of one pair of students. Thus, for future research, we would like to investigate other pairs to see how their connection of theoretical and experimental probability would develop. We would also plan to continue revising our design by exploring further how we can support students' struggles related to reading graphs, comparing fractions, and understanding what Run Size is sufficiently large.

Acknowledgments

This research was supported by a grant from the National Science Foundation (#1742125). The data presented, statements made, and views expressed are solely the responsibility of the authors. We would like to thank Jay Singh, Archana Prasad, Anthony Cuviello, and Richard Tashjian for their help in the design of the simulations and activities.

References

Abrahamson, D., & Wilensky, U. (2005). ProbLab goes to school: Design, teaching, and learning of probability with multi-agent interactive computer models. In D. Pratt, R. Biehler, M. B. Ottaviani, & M. Meletiou (Eds.), *The Proceedings of the Fourth Conference of the European Society for Research in Mathematics Education* (pp. 570-579).

Aspinwall, L., & Tarr, J. E. (2001). Middle school students' understanding of the role sample size plays in experimental probability. *The Journal of Mathematical Behavior*, *20*(2), 229-245.

Bakker, A., Hahn, C., Kazak, S., & Pratt, D. (2018). Research on probability and statistics education in ERME: Trends and directions. *Developing Research in Mathematics Education: Twenty Años of Communication, Cooperation and Collaboration in Europe* (pp. 1-15).

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). Introduction. In C. Batanero, E.J. Chernoff, J. Engel, H.S. Lee, & E. Sánchez (Eds.), *Research on Teaching and Learning Probability* (pp. 1-33).

Biehler, R. (1991). Computers in probability education. *Chance encounters: Probability in education*, 169-211.

Chick, H. L., & Baker, M. (2005). Teaching elementary probability: Not leaving it to chance. In P.C. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Building connections: Theory, research and practice: Proceedings of the 28th annual conference of the mathematics education research group of Australasia* (pp. 233-240). MERGA.

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

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1-73.

Drier, H. S. (2000). Children's meaning-making activity with dynamic multiple representations in a probability microworld. In M.L. Fernandez (Ed.), *Proceedings of the 22nd annual meeting of the North American Chapter of the International Group for Psychology of Mathematics Education*.

Fischbein, E., & Schnarch, D. (1997). Brief report: The evolution with age of probabilistic, intuitively based misconceptions. *Journal for research in mathematics education*, *28*(1), 96-105.

Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability—a psychological and pedagogical review. *Educational Studies in Mathematics*, *15*(4), 349-377.

Henry, M., & Parzysz, B. (2011). Carrying out, modelling and simulating random experiments in the classroom. In M. Pytlak, T. Rowland, E. Swboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 864-874).

Ireland, S., & Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education, 4*(3), 339-370.

Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, *9*(1), 68-96.

Panorkou, N., & Germia, E. F. (2021). Integrating math and science content through covariational reasoning: the case of gravity. *Mathematical Thinking and Learning*, *23*(4), 318-343.

Panorkou, N., & York, T. (2020). Designing for an integrated STEM+ C experience. In Sacristán, A. I., Cortés-Zavala, J. C., & Ruiz-Arias, P. M. (Eds.), *Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 2233-2237).

Paparistodemou, E. (2005). Young children's expressions for the law of large numbers. In D. Pratt, R. Biehler, M. B. Ottaviani, & M. Meletiou (Eds.), *The Proceedings of the Fourth Conference of the European Society for Research in Mathematics Education* (pp. 570-579).

Prodromou, T. (2012). Connecting experimental probability and theoretical probability. *ZDM, 44*(7), 855-868.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *The Journal of Mathematical Behavior*, *21*(3), 319-337.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, *76*(2), 105.