PRE-SERVICE MATHEMATICS TEACHERS' USE OF PROBABILITY MODELS IN MAKING INFORMAL INFERENCES ABOUT A CHANCE GAME

SIBEL KAZAK
Pamukkale University, Turkey
skazak@pau.edu.tr

DAVE PRATT
University College London, Institute of Education, UK
d.pratt@ioe.ac.uk

ABSTRACT

This study considers probability models as tools for both making informal statistical inferences and building stronger conceptual connections between data and chance topics in teaching statistics. In this paper, we aim to explore pre-service mathematics teachers' use of probability models for a chance game, where the sum of two dice matters in winning the game. We report on an interview with a group of three pre-service teachers as they engaged in predicting and conducting experiments and computer simulations as an attempt to develop a winning strategy. This paper focuses on how the participants came to use the theoretical model of the sum of two dice as they tried to coordinate the combinatorial analysis and the use of data as evidence in their predictions.

Keywords: *Statistics education research; Probability model; Informal statistical inference; Technology*

1. INTRODUCTION

We think that the [blue counters] have a wider spectrum to cross therefore they have more of an advantage. But according to average, the [yellow counters] have a more chance of winning because the average 7.2 and is a high chance of winning. It is a 50-50 chance of them both winning. (Chris, Jacob, Harry, 26.03.2013, Exeter, UK)

The above quote is from three 11–year–old boys when they wrote their initial hypothesis about the River Crossing Game (Canada & Goering, 2008) that they were about to explore: "Two players take turns rolling a pair of dice. If either player has any counters on the space corresponding to the sum of the dice, one of these counters can 'cross the river' and be removed from the board. The first player to remove all 12 counters on his or her side wins the game, but tied games are sometimes possible. Which arrangement do you think would win most of the time? Explain". The two given arrangements are displayed in Figure 1. In the quote above we see two competing ways of reasoning used when students made a prediction. On the one hand, the children considered the spread of counters as relevant to win the game and thought this gave the blue counters an advantage. On the other hand, they rolled the dice ten times to collect data and recognised that the average sum of the two dice (7.2 in their case) suggested the yellow counters had an advantage. They were unable to coordinate these conflicting pieces of evidence and so called it "50-50". In fact, a winning strategy for the game is not clear and the boys' intuitive approach raises a question about what the best arrangement would be. Another strategy might be to start with

an arrangement that resembles the theoretical distribution of the sum of two dice as closely as possible. According to Goering and Canada (2007), the best arrangement may not necessarily be the one that resembles the probability model. A large number of computer simulations conducted by Goering and Canada showed that arrangements with more counters piled in the middle sums (6, 7 and 8) consistently beat the probability distribution of the two dice sums. For example, they experimentally found that the number of rolls needed for arrangement X={3,4^3,5^4,6^6,7^8,8^6,9^4,10^3,11}[1] to win was about 70 while that for the probability distribution arrangement Y={2,3^2,4^3,5^4,6^5,7^6,8^5,9^4,10^3,11^2,12} to win was around 81 (p. 5). Hence, there is a complexity to the game that suggests studying solution strategies might reveal subjects' informal inferential reasoning about probabilistic models.
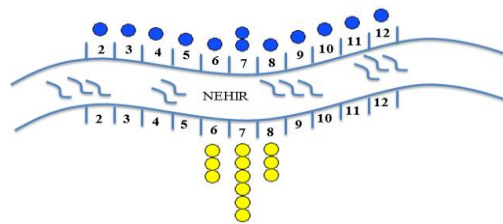

Figure 1. Initial arrangement for the River Crossing Game (blue counters at the top and yellow counters at the bottom)

In this study, we aim at researching how pre-service mathematics teachers use the probability model of the sum of two dice and what other resources they use as evidence when making informal statistical inferences in the context of the River Crossing Game. During the task, two dice and graph sheets were made available for use in the material game while *TinkerPlots* software (Konold & Miller, 2011) was on hand for creating probability models and executing them as computer simulations. Our assumption was that reasoning with models would also be influenced by the nature of the data, and the modelling and simulation capabilities of *TinkerPlots*.

2. EMPIRICAL BACKGROUND

The importance of developing students' informal statistical inference has been increasingly recognized in the statistics education community. The concept of informal statistical inference is typically used for "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (Zieffler, Garfield, delMas, & Reading, 2008, p. 44). It entails a process called informal inferential reasoning with data, which involves three components (Makar & Rubin, 2009): (1) making a generalization beyond the given data; (2) using the data as evidence to support the generalization; and (3) using probabilistic language to acknowledge uncertainty in describing the generalization. This implies that developing the language and understanding of probability is essential in informal inferential reasoning.

The role of probability in statistics is however not limited to the use of probabilistic language in reasoning about uncertainty. It is also used in more mathematical situations. As pointed out by Franklin et al. (2007), probability is a key tool in mathematical modelling. For instance, when we are asked, "How many heads are we likely to get if we flip a 'fair' coin four times?", it is a mathematical probability question, which leads to a

---

[1] As in Goering and Canada (2007), this set notation is used to show how the counters are arranged on the sums of two dice: "3" means one counter on sum 3, "4^3" means three counters on sum 4, "5^4" means four counters on sum 5 and so on.

probability model. "You take a coin; is it a fair coin?" is, however, a statistical question in nature and one can use the probability model obtained from solving the previous problem as a tool for arriving at an answer by comparing the experimental results with the expected outcomes from the mathematical model for a fair coin (Franklin et al., 2007). Rossman (2008) also pointed out that a probability model is needed in making a statistical inference. Thus, it is important to recognize the role of probability models in developing informal inferential reasoning as seen in previous studies with young students (Fielding-Wells & Makar, 2015) as well as teachers (Zapata-Cardona, 2015).

When the Exploratory Data Analysis [EDA] approach (Tukey, 1977) became the basis for the data strand in mathematics curricula (Biehler, 1986; Shaughnessy, Garfield, & Greer, 1996), the emphasis in instruction tended to be on "open-ended data exploration by students, aided by basic concepts of descriptive statistics, foregrounding data and making the mathematical model, probability, subsidiary" (Ainley & Pratt, 2001, p. 7). In this EDA approach, basic statistical ideas can be used without emphasising probability. Consequently probability and statistics are superficially treated as separate and independent topics in school curricula using the EDA approach in which the notion of probability is typically removed.

One approach to integrate these two connected topics involves both encouraging students to make informal conclusions based on data (Moore, 1990) and focusing on modelling simple random outcomes and using the models to interpret the distribution of empirical outcomes (Konold & Kazak, 2008). From this perspective, *modelling* can be viewed as a process in which "a student selects a probability generator whose sample space outcomes and their probabilities can be matched with the corresponding outcomes and probabilities of the contextual problem" (Benson & Jones, 1999, p. 2). Then a *model* is a representation of objects and functional relations observed in the real problem situation (Doerr & Pratt, 2008). When analysing random events, a *simulation* can be referred to as a "pseudoconcrete model" (Batanero, Henry, & Parzysz, 2005, p. 31) that provides a bridge between the probability model and reality. More specifically, this idea of pseudoconcrete model is considered as an educational tool that "can serve to improve students' probabilistic intuition, to teach them the different steps in the work of modeling, and to help them discriminate between model and reality" (Batanero, et al., 2005, p. 31).

Several previous studies have emphasized the role of models in developing young learners' fundamental probabilistic and statistical ideas. For instance, Horvath and Lehrer (1998) studied children's development of model-based reasoning about chance events. The study suggests the use of a sample space model as an ideal against which students can interpret the empirical results from experiments with dice. For example, constructing a model of sample space for the sum of two dice and recording their experimental results using bar graph notations encouraged second and fourth/fifth grade students to see the relationship between the number of possible outcomes and the distribution of actual outcomes. With the support of using notational representations and discussing the sample space model in class, students, in particular older ones, began to base their predictions on a model, rather than their previous experiences that they tended to use initially.

In Nilsson (2007), 12-13 year-olds were given pairs of unusual dice with only two different numbers on all sides, such as "111222 and 111222", "222444 and 333555", "111122 and 111122" or "222244 and 333355". The game task required them to place 24 markers (based on the sum of two dice) on the game board with numbers from 1 to 12 in line before they played the game with these pairs of dice. In the game, a marker was to be removed if it was on the sum rolled and the players needed to remove all their markers to win the game. The study showed that students tended to base their decision about the distribution of markers on what they thought the sample space was. When they ignored the

order of outcomes in rolling two dice, their incomplete sample space configuration however provided a limited model for students' predictions. The resulting distribution of outcomes after playing each game provided students with additional information to revise their initial model.

Konold and Kazak (2008) identified four main ideas used to form their activities connecting data and chance topics: model fit, distribution, signal-noise, and the law of large numbers. Model fit was one of the underpinning ideas for supporting middle school students' informal inference in the tasks involving explorations of data in chance through the use of modelling capabilities built in *TinkerPlots*. To make claims based on data, it is common to look at data with a model in mind, including our prediction or (personal or theoretical) expectation, and evaluate them with respect to that model (Konold & Kazak, 2008). *TinkerPlots*' probability simulation/modelling tool facilitates this model fit approach by enabling students to build models of random phenomena using a variety of devices and to carry out a large number of trials quickly (Konold & Kazak, 2008). As noted by Hawkins (1997), it is essential for students to construct probability models and be able to use them to understand the conceptual link between the physical experience and its formal symbolic version.

In order to implement a modelling approach to teaching of data and chance topics as seen in previous studies with children, teachers need to have strong content knowledge as well as knowledge of specific teaching strategies and effective use of technology tools. Although research on statistical knowledge of teachers has recently been growing (Batanero, Burrill, & Reading, 2011; Eichler & Zapata-Cardona, 2016), there is still not enough information about whether teachers are adequately equipped in these areas. Hannigan, Gill, and Leavy's (2013) study with 134 prospective secondary mathematics teachers showed that the prospective teachers tended to have poor conceptual understanding of certain statistical topics relevant to inference, such as sampling, population, randomization, and extrapolating from a regression model, on a standard test. Studying 15 elementary and 28 secondary school teachers' responses on a profile instrument, Watson (2001) reported teachers' difficulties and lack of confidence about teaching topics related to probability in particular and mostly at the elementary school level. Watson's findings indicated that some secondary school teachers also were not inclined to plan activity-based chance and data lessons including simulation or actual sampling. According to Steinbring (1991), however, in the teaching of probability theoretical knowledge, including signs and representations, should be mediated by the use of activities, such as experimentation and simulation. Therefore, teachers need to be familiar with developing and implementing such tasks that focus on the relationship between empirical situation and its mathematical model.

While our study with pre-service middle school mathematics teachers adds to our understanding of teacher knowledge in statistics and probability, it also has various aspects that address this issue raised by Steinbring. The task used in the study involves an experimental rather than abstract approach to probability. The chance game given in the task is challenging in a way that it engages pre-service teachers in a meaningful problem-solving situation to come up with a winning strategy. It also brings a modelling aspect and the idea of distribution for thinking about probability into the context of making informal inferences.

3. METHOD

3.1. RESEARCH CONTEXT AND PARTICIPANTS

Within the empirical background presented above, we report on a study investigating pre-service middle school mathematics teachers' spontaneous inferences based on their emerging model. Our main aim in this study was to explore how a group of pre-service mathematics teachers began to use a probability model when making informal inferences about winning strategies in a chance game, namely the River Crossing Game described in the Introduction section. The research questions were as follows:

1. In what ways does the group of pre-service teachers use the probability model of the sum of two dice for their winning strategy?
2. What other strategies do they develop as they explore and play the game with various numbers of counters on each side of the river?
3. To what extent does conducting a simulation of rolling two dice in *TinkerPlots* help them to develop a winning strategy in the game?

To address these research questions, we conducted a case study with a sample of pre-service mathematics teachers (6 females and 6 males) who were undergraduates in their third year in a four-year mathematics education programme at Pamukkale University, Turkey. The participants had already completed an introductory statistics course, which was compulsory in the degree programme and taught four hours a week during the fall semester. The course, taught by the first author, was tailored for pre-service mathematics teachers to support conceptual understanding and to develop an integrated set of probabilistic and statistical ideas in the context of model-based reasoning and making statistical inferences through the use of *TinkerPlots*. The idea of distribution was an overarching notion throughout this course. Students began with sorting and describing different (one variable) distributions by their shapes. This was a novel experience for these students because in school mathematics they only learned procedures and computations with small data sets. Statistical concepts and tools, including measures of centre and variability (mean, median, mode, range, mean absolute deviation, standard deviation, inter quartile range), graphical representations (dot plot, histogram, boxplot, scatter plot), regression (line of best fit), and sample and sampling, were introduced as ways of collecting, organising, describing, modelling and comparing distributions in the context of informal statistical inference. Probability topics included the idea of randomness, random variable, probability distributions (binomial, geometric and normal) and modelling random events. Students worked with various authentic data sets, some of which were collected in class, using *Fathom* software (Finzer, 2001) and conducted simulations of random events (i.e. coin flipping, dice rolling, and situations involving binomial and geometric probability distributions) using *TinkerPlots* software. They were also encouraged to work in small groups during in-class activities and in given assignments and projects.

3.2. TASK AND PROCEDURES

In this exploratory study, four groups of three students[2] were interviewed while engaging in the adapted version of the River Crossing Game (Canada & Goering, 2008). Each session lasted about 100–120 minutes. The interviewer (the course instructor) first showed the sheet that had the initial arrangements of 12 yellow and 12 blue counters on each side of the river (as seen in Figure 1) and began to describe the task orally:

12 counters are placed on each side of a river by two players and one of these can 'cross the river' (i.e. be removed from the board) if there is any counter on the space corresponding to the sum of the two dice rolled by the player. The first player to remove

---

[2] In the rest of the paper we use 'students' to refer to 'pre-service mathematics teachers' who are the participants of this study.

all his or her counters wins the game. What would be the best arrangement to win the game? Do you have any question about the game?

The game describes a situation in which the probability model of the sum of two dice can be used to make a decision under uncertainty in connection with the empirical results.

After describing the game, students were asked to decide together which arrangement in Figure 1, blue or yellow, they would expect to win the game and to explain their reasons. Then they were introduced to a scale from 0 ["not sure at all"] to 10 ["definitely sure"] on which they could mark the level of their confidence in their choice on the given sheet. After the predictions, students played the game with two dice and made a dot plot of the results on the graph sheet provided. They were again asked to evaluate the level of their confidence in their choice after playing Game 1 with an explanation. During the interview the students were expected to discuss and make decisions together while working on the task as a group, similar to what they had done in group work during the course described above. The role of the interviewer was to give them the questions to discuss and ask for explanations and some clarifications if needed. She did not interrupt or help them while they were discussing together.

The same procedure was repeated for Game 2. Before Game 3, they were asked to come up with a better arrangement of 12 counters to be more confident about winning against the other arrangement given (blue or yellow). In this phase, the interviewer asked the following question: "Can you built a model for the game using *TinkerPlots* to see which arrangement will win in the next game?" Then each group decided to model rolling two dice in *TinkerPlots* and simulated the game results (i.e., sum of two dice rolled) for their choice of number of trials without any guidance. Again they evaluated the level of confidence before and after playing the game. They were also asked what they would do to be "more sure" after Game 3. Later, they were given a game sheet as seen in Figure 1 but with no counters on either side of the river and asked to suggest the best arrangement if they played the game with 36 counters. Once they decided about their arrangement on one side of the river on the given sheet, an alternative was presented on the opposite side by the interviewer. One alternative was a distribution of 36 counters based on the probability model of dice sums, if not suggested. Otherwise, another possible arrangement was $\{3,4^3,5^4,6^6,7^8,8^6,9^4,10^3,11\}$, which won 60% of the games against the probability model according to the large number of simulations by Canada and Goering (2008). Game 3 had required the students to state their level of confidence, first after making a prediction and then after playing the game. The same procedure was followed for Game 4 and Game 5. After Game 4, students were again allowed to revise their initial arrangement (for the 36 counters).

Finally, a follow-up question for arranging 360 counters and 360000 counters was asked to elaborate more on the participants' reasoning about the probability model without actually playing the game. They were only expected to give the number of counters they would place on each sum to win the game most of the time. The computer simulations of the number of rolls required for different arrangements of 36 or less counters to win the game are somewhat misleading. Goering and Canada's (2007) simulations showed that placing any counters on sums 2 and 12 was actually not the best strategy for games involving 36 or less counters. For example, when they conducted a simulation of 10000 games for arrangement $A=\{2,3^2,4^3,5^4,6^5,7^6,8^5,9^4,10^3,11^2,12\}$ vs. arrangement $B=\{3,4^3,5^4,6^6,7^8,8^6,9^4,10^3,11\}$, they found that arrangement A (proportional to the probability distribution of sum of two dice) did not have the smallest expected duration for the number of rolls and arrangement B was superior in terms of the number of games won. Moreover, they experimentally found a unique optimal arrangement for each case involving the number of counters (*n*) greater than or equal to 1 and less than or equal to 12,

such as {7}, {6,7}, {5,6,7,8}, {5,6,7,8,9} and so on for n=1, 2, 3 and 4 respectively. For n=12, arrangement $\{4,5^2,6^2,7^3,8^2,9,10\}$ was the optimal one without any counters on sums 2, 3, 11 and 12. Therefore, one needs to adopt a different perspective involving reasoning about the probability model as the number of trials goes to infinity because the law of large numbers suggests that the empirical distribution of the sum of two dice will converge to the theoretical distribution as n gets larger and larger. However, this part is left as a question for the readers in Goering and Canada (2007, p.11). They did not explore what would happen with more counters since it was not possible to conduct simulations for large values of n in their computer programme. Thus we wanted to further explore students' reasoning about the cases of 360 counters and 360000 counters in this study.

3.3. DATA COLLECTION AND ANALYSIS

Data collection included the video recording of interview sessions with each group of students working together on the task described above. Other data for the task involved computer simulations and written artefacts (game sheets and graphs) generated by the students during the interview. Students' work on the computer in all sessions was recorded by *Camtasia* software (www.techsmith.com).

The initial data analysis of each group work involved identifying key moments that might provide insight into students' ways of informal inferential reasoning and building probabilistic models of the results from a game. The first author watched the video recordings of each group interview with written artefacts collected at hand and took notes (in English) detailing students' responses, group discussions, arguments, explanations and game results obtained during each phase, i.e., predicting, stating confidence level, playing game (simulation) and stating confidence level again. Each author analysed these documents to identify any key ideas regarding probabilistic reasoning, modelling, informal inference, and strategies used by the participants. Then, a selection of episodes transcribed by the first author and student artefacts provided the basis for further analysis. Independent content analyses by the authors for each case, i.e., each group of three students working together on the task, centred around the following themes: participants' construction of a model with regard to which empirical data were evaluated, reasoning about the probability model based on sample space, and the way they used this model to make inferences about the best arrangement of counters to win the game. These detailed analyses for each case were then compared by searching for commonalities and disparities, i.e. What was common across cases? What was special about a particular case? The two authors shared their interpretations to reach a common understanding of the data.

4. RESULTS

After our qualitative analyses of data described above, we saw several commonalities as well as some disparities in the ways the four groups of participants came to use a probability model when making informal inferences about winning strategies in the River Crossing game. The groups began to list the possible configurations of two dice in the sample space at the very beginning and used this list to make their predictions about the arrangements of 12 counters. All groups tended to use this list as their model in conjunction with the empirical data obtained from the actual games while making their decisions throughout the task. Moreover, building a model of rolling two dice in *TinkerPlots* and watching the resulting distribution of the sum of two dice for a large number of trials helped them justify their predictions. While three groups simulated 1000 rolls at most, one group (group 4) did 10000 repetitions in *TinkerPlots*. It seemed that simulating 10000 trials was

more effective since it gave a "perfect" triangular shape while the other groups had to deal with the variability in the outcomes of 1000 trials. In addition to the use of empirical data and the probability model to make predictions, groups used various other strategies to win the game. For example, three of the groups tended to avoid placing counters on the extreme sums, such as 2 and 12. Two of these also used a strategy of balancing the number of counters on each side of the river, e.g., H: "If we don't put any on these [*sums of 6 and 8*], when six is rolled, this guy [*blue arrangement*] will be ahead again. So we need to put one or two counters on six and eight" (Group 2, predicting arrangement 2 (12 counters) after Game 2). On the other hand, the fourth group had a unique strategy that we called "minimising wasted rolls", which then shifted to a "minimising waiting time" strategy to win the game. Based on these analyses, we decided to focus on a single case that stood out from among the four groups in order to give more in-depth analysis of this particular group. We focus on the analyses of strategies used in the game task by the group comprised of Beril, Ezgi and Filiz. They used a unique approach for understanding the role of the probability model in developing a winning strategy for a complex game situation.

4.1. ARRANGEMENTS FOR 12 COUNTERS

Initial prediction for arrangement with 12 counters The group began to discuss whether they should choose blue or yellow counters in arrangement 1 given to them initially (see Figure 1) to win the game. First Ezgi suggested that they could remove the blue counters faster because whatever sum they would roll, they could remove a blue counter. After Beril challenged her idea by saying "Here [*blue counters*] there is one on 6. The other rolls of 6 will be wasted", and Ezgi seemed to agree with her. Then Filiz pointed out the likelihood of each sum of two dice: "Here [*for sum 12 on the game board*] only one-one can be rolled, there is no other possibility. But it seems like this middle part there are more [*ways*]. For example for [*sum of*] six, there can be three-three, two-four, I mean it has more probability". This led students to write down the possibilities for each sum (see Figure 2). As seen in Figure 2, they did not however consider different orders, such as 1-3 and 3-1, and thus their list showed 21 possible outcomes rather than 36.

				3-3	3-4	4-4				
			2-2	2-3	2-4	2-5	3-5	4-5	6-4	
1-1	1-2	1-3	1-4	1-5	1-6	2-6	3-6	5-5	6-5	6-6
2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**

Figure 2. Group's initial list of (incomplete) sample space (n=21)

By focussing on the number of ways to roll sums 6, 7 and 8 in their incomplete list, Ezgi, Filiz and Beril began to talk about "wasted rolls" for both blue and yellow counters as arranged in Figure 1. They debated between choosing blue or yellow counters.

Beril: For example if one-one is rolled, it will be wasted here [*yellow counter*]. If we count the wasted rolls here,
Filiz: Four. And four more here. Eight of them would be wasted, I mean these [*showing the sums where there is no yellow counter, such as 2, 3, 4, 5, 9, 10, 11, 12*]
Beril: If we roll more, more will be wasted
Filiz: Eight or more
Ezgi: But in the above [*blue counters*] we can remove at least one in every roll.
Beril: Yes, I think the same way, but would at least one of each be rolled? None of them might be rolled also.

Ezgi: Yes, alternatively none of them might occur. In order for the one below [*the yellow counters*] to win, seven, those whose sum is seven, should be rolled several times, or, you see, one [*sum of*] six, one [*sum of*] seven should be rolled so that we can remove sevens. Otherwise, it will be wasted. But in the above one [*the blue counters*], for example if anything, their sum anything from two to twelve, is rolled, we will remove that in the above one. But there is a problem there too. For example, [*a sum of*] two is rolled and we remove the two [*the counter on sum 2*]. When the next [*sum of*] two is rolled, it will be wasted. There is that too.

Filiz: In any case, it would be wasted in this one [*the yellow counters*] too.

It seemed that students did not use the likelihood of rolling specific sums from their theoretical model (Figure 2). Instead they focussed their attention on whether they would waste more rolls for the blue or the yellow counters by counting the number of unique configurations in their sample space. In the rest of the discussion, Ezgi pointed out "Nine would not be wasted. We will have twenty-one rolls. In other words twelve of those will be wasted. Because we have twenty-one possibilities, we have counters for nine of these [*counting possible outcomes to get sums of 6, 7, 8*]". More specifically, if they were to choose the yellow counters, they thought that they would be more likely to waste rolls (12 rolls according to their incomplete sample space). However, if they were to choose blue counters, then they thought only 5 rolls (counting two extra possible ways of getting sum 6, one extra possible way of getting sum 7 and two extra possible ways of getting sum 8 in their list, see Figure 2) would be wasted in total. After that, they chose to base their decision for blue counters on their idea of "minimising wasted rolls". Before playing the actual game, the level of their confidence was recorded as 6 because they thought that the chances of the blue or the yellow counters winning were close to each other but the blue counters had some advantage.

Playing games and predictions for arrangement with 12 counters After a total of 28 throws, the game ended when all the yellow counters had been removed and there remained blue counters on totals of 3, 10 and 11. After the group, who had favoured the blue counters, lost in the first game, they were asked to analyse the game results seen in Figure 3. They paid attention to the most frequent outcomes (sums 5, 7, 8, 9) in the graph. Beril was surprised that all the yellow counters on 7 were removed because they thought that each sum of 6, 7 and 8 had an equal number of possibilities according to their initial model (Figure 2). In other words, since there were more counters initially on 7 than either 6 or 8, it was reasonable for her, given her supposition about the distribution, to think that removing the counters on 7 would take longer. Yet, this result did not lead to any change in their theoretical model. When they made their next prediction, they chose the yellow counters because they thought it was more likely to roll the sums 6, 7 and 8 based on the game results.
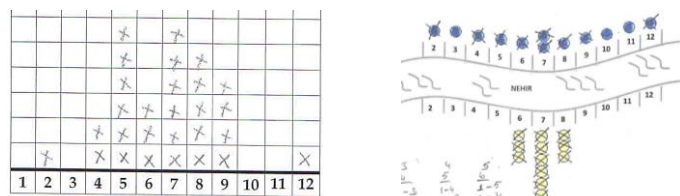

Figure 3. The graph of results after playing the game physically (on the left) and the removed counters marked on the game sheet (on the right)

After this analysis, they were asked to mark their level of confidence before playing the game again. Their discussion led to a different insight about their winning strategy.

Intv: Before you play the game, if you choose yellow [*counters*], how confident are you that the yellow will win? Can you mark it here?

Filiz: This time it is a bit higher because there were several repeated sums [*Figure 3*] but sevens were rolled more quickly.

Ezgi: What do we say is the probability of yellow winning?

Filiz: There were many instances where the rolls were wasted in both, but when we look at the wasted rolls, we waited for these more [*showing the smaller, such as 2, 3, 4, and bigger, such as 10, 11, 12, sums for the blue counters*], we needed to roll the dice more.

Beril: Also in the end we couldn't roll the three sums [*3, 10, 11*] in this one [*the blue counters*].

Filiz: Exactly.

In this exchange between students, there appeared to be a shift from a "minimising wasted rolls" perspective to a "minimising waiting time" perspective. While playing the game, they realised that the sums on the lower and higher ends of the totals from 2 to 12 were less likely to be rolled because they had to wait longer to remove the counters on those sums. Hence, the waiting time became more salient to them after the actual game. As a result they seemed to feel more confident about their prediction, rating their level of confidence as 8.

In the second game, after a total of 45 throws, the yellow counters had all been removed whereas there remained a blue counter on 12 (see Figure 4). After these results, the students began to discuss the probability of winning if they were to choose the yellow counters. They agreed that the sums 6, 7 and 8 had more chance to be rolled and confirmed this by emphasizing the piling up on 6, 7 and 8 in the results graph. Beril also noted that there was one yellow counter on the total 7 and one blue counter on the total 12 before the game was over. They came to a conclusion that it was more difficult to roll 12 because there was only one way to roll a sum of 12 (six and six) and that 7 would occur more easily. Then they changed their confidence level from 8 to 9, but not 10 because they thought there could be still a chance to roll 12 before 7.

Figure 4. The graph of results after playing the game a second time (on the left) and the removed counters marked on the game sheet (on the right)

TP simulation and revised theoretical model After game 2, the students were encouraged to model rolling two dice and simulate the game results using *TinkerPlots*. As they were used to making this kind of simulation in *TinkerPlots* during their statistics course, they quickly built the model of rolling two dice and ran 1000 trials (Figure 5). Beril noticed that the frequencies for 4 and 5 were about equal (see graph on the left in Figure

6). They wanted to increase the number of trials to see if the resulting distribution would change and did another 10000 trials. When they saw the graph on the right in Figure 6, Ezgi called this an "exact normal distribution". This led them to rethink their initial theoretical model for the sum of two dice, which had only 21 possible outcomes because they did not consider the order.
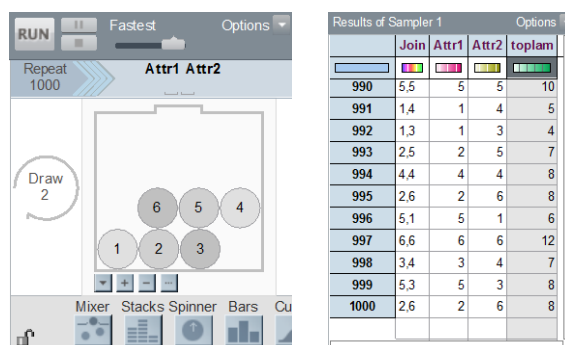

Figure 5. Computer model of rolling two dice built by Group 4 in TinkerPlots and its outcomes, including the sum ("toplam" in Turkish) variable in the table
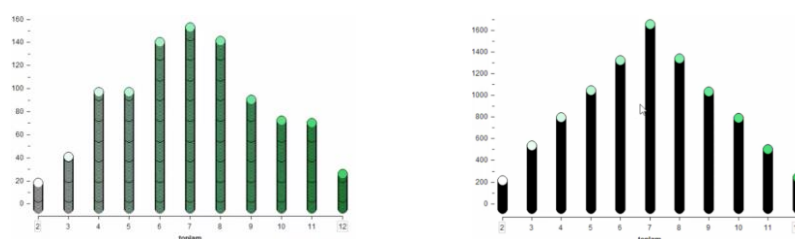
Figure 6. TinkerPlots simulation results for N=1000 (on the left) and for N=10000 (on the right)

They first focussed on the middle sums (6, 7 and 8). These were still the most likely outcomes according to the simulation results, yet Beril pointed out that they had found them equally likely when they had listed all possible outcomes for each sum (Figure 2). Ezgi then added "but when we increased the number of trials, seven has occurred more often". They all were surprised by this observation that led them to notice that they had not counted different orders in their initial list of outcomes. Once they revised their model to the triangular distribution with 36 possible outcomes, the simulation results made more sense to them. Their new model also helped them realise why the sum of 7 was rolled more often (or "easily" in their terms). Their level of confidence had already reached 9 out of 10 earlier but the students now felt even more confident because of the theoretical explanation of why rolling a 7 is more likely, including recognising the importance of order, as seen in the following excerpt:

Beril: Actually after this one [*TinkerPlots simulation result for n=10000*] we became more confident. After we realised why seven was easily rolled.
Filiz: We are more sure now. We also realised the importance of order.

Even though they realised there was some possibility of losing, they expressed their increased confidence as 100%, which we take to mean as even higher than 9 out of 10.

4.2. ARRANGEMENTS FOR 36 COUNTERS

In this phase, students were asked to suggest the best arrangement for winning if the game was played with 36 counters. Initially they tended to pile the counters in the middle (see Figure 7, arrangement on the left) because of the theoretical distribution so that they could remove those more quickly. At the same time they did not want to waste any roll so they placed one on each less likely sum on both ends. Beril thought that they would need to roll the dice more because they had 36 counters to remove and thus it would be likely to get the sums 2, 3, 4, 10, 11 and 12. Once they all decided on their arrangement of 36 counters, the interviewer arranged the counters on the opposite side based on the theoretical distribution (the triangular distribution). When they were asked to evaluate the level of their confidence about winning against the theoretical model, they agreed on 8. They seemed to be pretty confident about winning with their arrangement based on an emerging idea from the tension between the two strategies: "minimising the wasted rolls" and "minimising the waiting time".
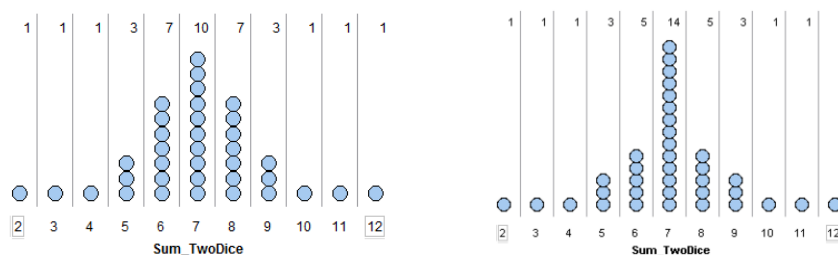

Figure 7. Group's 1st arrangement of 36 counters (on the left) and 2nd arrangement of 36 counters (on the right)

They played the game with two dice and won. Only one counter on the sum 10 was left in the opposing side (theoretical distribution). Looking at their results they recorded during dice rolling, they noted that the sum 7 occurred the most and a lot more than 6 and 8. This led them to think that even though they had more counters on 7, they could still win the game. They also mentioned that they had to wait for the sum 8 for a long time. Hence, the experience of playing the actual game helped them increase their confidence level to 9 this time.

After analysing the game results, students were asked whether they would change their initial arrangement of 36 counters in any way to become more confident about winning against the theoretical distribution. While revising their arrangement, they seemed to use the data from the game results as the basis for their decisions. For example, since they had to wait for rolling the sum 8 in the last game, they decided to move some of the counters on 8 to 7. Eventually, they all agreed on having more counters on the sum 7 and less on the sums 6 and 8 and kept the symmetry (see the distribution on the right in Figure 7). They kept the counters on the tails as they were before. Hence, once more their strategy emerged from the tension between "minimising the wasted rolls" and "minimising the waiting time". Their level of confidence remained unchanged even though they made some revisions to their original arrangement.

In the next game, the students again won against the theoretical distribution. This time one counter was left on the sum 11. After the game, the students moved to analysing the

game results. They noticed that the resulting graph was changed slightly, but the sum 7 had still the highest frequency. Also, Filiz shared an observation that before they rolled any sum of 8, they still had many counters to remove from the sum 7, but they could remove those counters faster than removing the counters on 8. So their focus began to shift to the most likely outcomes in terms of minimizing the wait time for any particular sum. Perhaps because there was no unexpected insight from playing the game the second time, their level of confidence did not change.

4.3. ARRANGEMENTS FOR 360 COUNTERS

Like all the other groups, Beril, Filiz and Ezgi did not spontaneously consider the theoretical distribution as the optimum arrangement for 36 counters to win the game. In order to test whether their emerging strategy in the previous games would ever need to be reconsidered when the conditions changed, such as the number of counters used in the game, we asked the students how they would arrange the counters if they were to play with 360 of them. They began with the last game results and Filiz pointed out the relevance of number of rolls needed: "here we had 87 rolls but before we did 10000 [*referring to their TinkerPlots simulations earlier*] and that looked like normal distribution. As the number of rolls increases, it is more like normal". Filiz first noted the number of rolls (87) they needed to remove all 36 counters in their arrangement (second arrangement in Figure 7) in the last game against the theoretical distribution. Others agreed with her idea since they thought they would need to roll the dice a lot more to remove all 360 counters than to remove the 36 counters they had previously experienced. Since their simulation was only about rolling two dice and showing the sums, students needed to relate the number of rolls to generate a large sample for the outcomes of two dice (i.e., $n=10000$) to the number of rolls to win the game when the number of counters involved in the game was increased. Using the idea of the law of large numbers (discussed in the statistics course that they had taken), they came up with a distribution of 360 counters proportional to the theoretical distribution: $\{2^{10}, 3^{20}, 4^{30}, 5^{40}, 6^{50}, 7^{60}, 8^{50}, 9^{40}, 10^{30}, 11^{20}, 12^{10}\}$. Hence they were able to coordinate the combinatorial analysis of sum of two dice and the use of simulation data in their last prediction with a confidence level of 9.5 with a very small level of uncertainty.

5. DISCUSSION AND CONCLUSION

The paper began with a vignette of three 11–year–old boys as they were deciding which counters, blue or yellow, in Figure 1 to choose in order to win the River Crossing Game. Since there was no easy way to see a winning strategy for the game, the children had an intuitive approach that led them to consider both the spread of the counters and the expectation for the most likely outcome based on data by rolling two dice. However, they were not able to coordinate these two ideas that were in conflict with each other, nor did they consider the theoretical distribution of the sum of two dice as a model. Therefore, we aimed to explore how students begin to use a probability model when making inferences about winning strategies in this particular chance game. To do so, we worked with pre-service mathematics teachers who had completed an introductory statistics course for teachers. In the study, we focussed on the following research questions:

1. In what ways does the group of pre-service teachers use the probability model of the sum of two dice for their winning strategy?
2. What other strategies do they develop as they explore and play the game with various numbers of counters on each side of the river?

3. To what extent does conducting a simulation of rolling two dice in *TinkerPlots* help them to develop a winning strategy in the game?

As in children's case, the students in our study were also initially struck by the spread and the expected outcomes in the given arrangements; but they began to use a model that they thought represented the distribution of sum of two dice. In particular, their initial model was based on an incomplete sample space. Instead of 36 possible outcomes, they generated 21 possibilities (Figure 2) when they ignored the order of outcomes of rolling two dice for getting the sums. Students' difficulties in considering the order important in rolling two dice seemed to be persistent across age levels as seen in other studies with school students (e.g., Fischbein, Nello, & Marino, 1991; Horvath & Lehrer, 1998; Nilsson, 2007). In fact, in their study with 9-14 year olds Fischbein et al. (1991) concluded that it does not naturally occur to students that possible simple outcomes constituting the compound results should be distinguished and counted separately when constructing the sample space.

By generating 21 possible outcomes for the sum of two dice initially, the students looked at the given arrangements of blue and yellow counters with this model to make their first prediction about which counters to choose in order to win the game. Instead of using the likelihood of each sum occurring based on this sample space (e.g., it is more likely to roll totals 6, 7 and 8 because there are more possible ways to get these sums), they focussed on "wasted rolls" in each arrangement by counting the possible outcomes for the sums only where there was no counter. Hence, they found that there were 12 possibilities to waste if they were to choose yellow counters. In the case of blue counters, however, they counted any additional possible pair of dice for the sums with respect to the number of counters on each sum and thought that only 9 rolls would be wasted. We called this the "minimising wasted rolls" strategy. At this stage they did not seem to note the relationship between the number of possible outcomes for each sum and the distribution of empirical results in the sense that young students were encouraged to make predictions about the outcomes of rolling two dice in Horvath and Lehrer (1998) study. In other words, they did not yet appear to have any expectation for the empirical distribution of totals in rolling two dice. The complex nature of the problem could have made that connection difficult because students tried to use the probability distribution for the random variable "X= The sum of two dice" to decide the arrangements of counters while their strategy to win the game focussed mainly on a very complex random variable "Y= Waiting time until all sums have been rolled to remove all counters" which they did/could not simulate in *TinkerPlots*.

Our analysis showed that playing the actual game physically and examining the distribution of empirical data helped students to revise their strategy to win the game. From the results of playing the first game the students made some changes in their strategy based on data and their observations during the game. The graph of results (Figure 3) led them to think that the sums 6, 7 and 8 were more likely to be rolled because of their high frequency of occurrence. Rolling the actual dice to play the game helped students realise how long they had to wait until they rolled a certain sum while their *TinkerPlots* model for very large number of rolls was set to generate all 10000 rolls at once very fast. For example, since they had to wait for the sum 8 for a long time in game 3, they decided to move some of their counters on 8 to 7 before the next game. They also took the observation of rolling 7 "more quickly" than the others into consideration. Hence, they began to shift their perspective from "minimising wasted rolls" to "minimising waiting time". They realised that the smaller (2, 3, 4) and bigger (10, 11, 12) sums were less likely to occur since it took longer time to remove the counters on these sums. At this point, they still did not relate the likelihood of outcomes to the number of possibilities in the sample space. Neither did they realise the limitation of their model based on an incomplete sample space until they ran

large amount of trials using the *TinkerPlots* simulation tool and saw the triangular shape, which they called the "normal distribution" (Figure 6). As suggested by previous findings the need to consider all possible equally-likely outcomes that make up an event is not intuitive and can pose a challenge for the students before evaluating the likelihood of events (e.g., Fischbein, et al., 1991; Horvath & Lehrer, 1998; Nilsson, 2007; Zapata-Cardona, 2015). Our analysis of this case study shows that even though the students considered possible outcomes initially, in their predictions they did not link these possibilities with the probability of obtaining these sums.

The students in our study tended to use the data they collected through playing actual games as evidence not only to revise their model used to make predictions similarly to findings in Nilsson (2007), but also to express their level of confidence. The level of the uncertainty in their predictions increased according to their belief about how much support the data provided. It seems as if through the analysis the students were heavily influenced by the data that they observed in practice. When they were asked to make a prediction about the arrangement of 360 counters in the last interview question, there were no data and theory came to the fore. Through the use of the law of large numbers they could relate the probability distribution to the expected outcomes of the game. Making this link was a breakthrough for the students and supported by the use of simulation tool to gather more data as seen in previous studies using *TinkerPlots* (Konold & Kazak, 2008).

In conclusion, the chance context described in this paper provided opportunities for the students to make inferences based on data in a task context that placed emphasis on an underlying model. As a result, model-based reasoning was facilitated. This process was also supported by the use of technology tools for computer-based simulations and data visualisation. The existing studies focusing on models and modelling in probability situations (e.g., Horvath & Lehrer, 1998; Nilsson, 2003) do not necessarily take an inferential reasoning perspective, which is key to bridging data and chance concepts. In terms of students' engagement in the task, we saw an emerging strategy based on the tension between seeking to have sufficient spread to avoid throws which missed all remaining counters and having enough counters at the centre which occurred most frequently. This strategy emerged from their attempt to coordinate the combinatorial analysis and the use of data as evidence in their predictions. Eventually, acknowledgement of the law of large numbers seemed to be the key to their reasoning about the probability model for winning the game in the long run.

The findings from our study add to the body of research on pre-service mathematics teachers' knowledge about chance and data topics. While other studies (e.g., Hannigan et al., 2013; Watson, 2001) pointed out the difficulties that teachers had in relation to probability and statistics topics, this study presented insights about engagement of a group of pre-service teachers, who had prior experience with both experimental and modelling approaches to probability supported with data explorations and computer simulations, in a complex problem-solving situation involving uncertainty. The given task provided them with opportunities to use both their theoretical knowledge about probability and empirical results from the actual games and simulations in developing strategies to win the game. Their focus on distribution shape (emphasised in the course) seemed also to be critical in their thinking about probabilities of chance outcomes (i.e., seeing the distribution of sum of two dice as a whole as the number of rolls increased). In addition, this study contributes to the issue of addressing the theoretical nature of probability in teaching mathematics as Steinbring (1991) suggested by showing the importance of future teachers' familiarity with developing and implementing tasks focussing on the relationship between empirical situation and its mathematical model. Moreover, developing, using, and evaluating probability models in exploring chance processes are becoming a part of mainstream school

curriculum. For example, the seventh grade Statistics and Probability domain in the Common Core State Standards for Mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) prescribes developing a probability model for finding probability of a chance event and comparing probabilities based on a model with those from observed frequencies. In Turkey and other countries, teacher education courses emphasise procedures, formulas and methods, rather than providing opportunities for informal inference in learning statistics and probability. In this study, we illustrate how informal inference can be facilitated through experimentation in a situation where model-based reasoning is encouraged.

REFERENCES

Ainley, J. & Pratt, D. (2001). Introducing a special issue on constructing meanings from data. *Educational Studies in Mathematics, 45*(1), 1-8.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 15-37). New York: Springer.

Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study*. Netherlands: Springer.

Benson, C.T., & Jones, G.A. (1999). Assessing students' thinking in modelling probability contexts. *The Mathematics Educator, 4*(2), 1-21.

Biehler, R. (1986). Exploratory data analysis and the secondary stochastics curriculum. In R. Davidson and J. Swift (Eds.) *Proceedings of the second international conference on teaching statistics* (pp. 79-85). University of Victoria Conference Services, Victoria, British Columbia, Canada.

Canada, D. & Goering, D. (2008). Deep thoughts on the river crossing game. *Mathematics Teacher, 101*(9), 632-639.

Doerr, H. & Pratt, D. (2008). The learning of mathematics and mathematical modelling. In M. K. Heid & G. W. Blume (Eds.), *Research on technology in the teaching and learning of mathematics: Syntheses and perspectives: Mathematics, teaching and policy*, (Vol. 1, pp. 259-285). Charlotte, NC: Information Age.

Eichler, A. & Zapata-Cardona, L. (2016). *Empirical research in statistics education ICME-13 topical surveys*. DOI 10.1007/978-3-319-38968-4_1

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

Finzer, W. (2001). *Fathom dynamic statistics*. Emeryville, CA: Key Curriculum Press.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics, 22*(6), 523–549.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre- K-12 curriculum framework*. Alexandra, VA: American Statistical Association.

Goering, D. & Canada, D. (2007). The river crossing game. *Mathematics Magazine, 80*(1), 3-15.

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education, 16*(6), 427–449.

Hawkins, A. (1997). Myth-conceptions. In J. B. Garfield and G. Burrill (Ed.), *Research on the role of technology in teaching and learning statistics* (pp. 1-14). Voorburg, The Netherlands: International Statistical Institute.

Horvath, J. & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. LaJoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12* (pp. 121-148). Mahwah, NJ: Lawrence Erlbaum.

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

Konold, C. & Miller, C. D. (2011). *TinkerPlots 2.0: Dynamic data exploration*. Emeryville, CA: Key Curriculum.

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

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common Core State Standards (Mathematics Standards)*. Washington D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Nilsson, P. (2003). *Experimentation as a tool for discovering mathematical concepts of probability*. Paper presented at the *Third Congress of European Research in Mathematics Education* (CERME3), Bellaria, Italy. Retrieved from http://www.mathematik.unidortmund.de/~erme/CERME3/

Nilsson, P. (2007). Different ways in which students handle chance encounters in the explorative setting of a dice game. *Educational Studies in Mathematics, 66*(3), 293–315.

Rossman, A. J. (2008). Reasoning about informal statistical inference: one statistician's view. *Statistical Education Research Journal, 7*(2), 5-19. [Online: http://iase-web.org/documents/SERJ/SERJ7(2)_Rossman.pdf]

Shaughnessy, J. M., Garfield, J. & Greer, B. (1996). Data handling. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education* (pp. 205-237). Dordrecht: Kluwer Academic Press.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 135-168). Dordrecht, The Netherlands: Kluwer.

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

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education, 4*(4), 305–337.

Zapata-Cardona, L. (2015). Exploring teachers' ideas of uncertainty. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 163-181). Minneapolis, MN: Catalyst Press.

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. [Online: https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf]

SIBEL KAZAK
Pamukkale University
Eğitim Fakültesi
Matematik ve Fen Bilimleri Eğitimi Bölümü
Denizli 20070, TURKEY