ATTENDING TO STUDENTS' REASONING ABOUT PROBABILITY CONCEPTS FOR BUILDING STATISTICAL LITERACY

Hamid Sanei and Hollylynne Lee
North Carolina State University, United States
hrsanei@ncsu.edu

This paper investigates two specific probabilistic biases which middle graders usually exhibit when reasoning about probability and randomness on assessment items. We discuss how students' reasoning about key probability concepts undergirds statistics literacy related to randomness, independence, and the likelihood of future events based on past results. We examine factors evoking misconceptions and students' (in)consistency in exhibiting them. Findings indicate that misconceptions can be evoked based on three types of factors including (1) students' particular understandings of probability and randomness, (2) general item characteristics, and (3) aspects of probability in items. Moreover, possession of a specific misconception will most likely result in exhibiting the bias again on other occasions including the same evoking factors (consistency).

INTRODUCTION

The *Diagnostic Inventories for Cognition in Education (DICE)* project aims to develop a freely available, web-based assessment system that provides teachers with timely, accurate, and actionable feedback about student cognition about several probability concepts. Due to the plethora of misconceptions in probabilistic reasoning (e.g., Batanero et al., 2016; Jones et al., 2007) and the difficulty with probabilistic reasoning that many teachers share (Stohl, 2005), there is a strong need to create assessments that can identify students' probabilistic conceptions, including misconceptions, and effectively communicate that information to teachers to inform instruction. As part of this project, we conducted cognitive interviews with students to think aloud as they worked through a set of assessment items and to answer questions about their rationale and justification for their choices. This paper aims to present results from a detailed analysis of eight middle school students who were interviewed about a set of items assessing students' reasoning about concepts related to independence and the probability of events when a sample space is given.

Though the term misconception can be perceived as having a "fix and replace" connotation, we use the term broadly to reflect that most misconceptions are a mix of both flawed and productive thinking (Schoenfeld et al., 1993; Smith et al., 1994). While the term can be interpreted to have a negative connotation associated with an outdated "fix and replace" instructional approach, our broader use of the term implies having a misconception. Students' (mis)conceptions can reflect a degree of sophistication in reasoning, are often logical to the student and influenced by a reasoning situation, and can be useful building blocks towards more robust conceptual understanding (Smith et al., 1994). This paper focuses on students' reasoning related to independence and equiprobability. A misconception about independence is an example of Representative Bias (Tversky & Kahneman, 1974) and stems from a lack of understanding about independent events. Students with this misconception ignore the independence of events (II) and might assume that the probability of an event can be influenced by a pattern of outcomes of recent events. For example, if a fair coin flip results in three Heads, a student believes it is more likely for the next flip to be Tails, as this will even out the previous flips (negative recency). Conversely, a student might believe a Head is more likely because the coin is on a streak of Heads (positive recency). A misconception about all outcomes being equally likely is commonly termed the Equiprobability Bias (Lecoutre, 1992) and involves assigning equally likely (EL) probabilities to events that are not equally likely to occur. For example, if a cooler contains two Red and one Blue can, students will expect that a randomly selected can have equal chances of being Red or Blue as there are two colors.

Developers of concept inventories hold working theories that misconceptions exist as traits to be measured and that a student with a misconception will likely select options that correspond to the misconception across a set of items (Bradshaw & Templin, 2014). Thus, items are written such that incorrect response options are mapped to reasoning consistent with a targeted misconception. The research questions guiding this study are: (1) What seem to be factors that evoke students' reasoning with EL and II misconceptions and what possible factors inhibit the emergence of such biases? (2) How consistent are students who possess a misconception in exhibiting the bias in a set of items?

METHOD

This paper only focuses on eight of the students interviewed in cognitive lab interviews (April-May 2019). Each interview began with a series of questions asking about how they would describe probability and randomness, and whether they could give an example of something with a high and low probability. All interviewed students responded to an item, called Raffle (Figure 1) targeting both misconceptions, and had an opportunity in their interview to respond to several additional items measuring one or both EL and II biases, as well as additional items measuring other concepts that are not the focus of this paper. Eight students were selected for careful analysis based on the large number of items they worked on within the interview that measured EL and II misconceptions. Each of the eight students saw seven to eleven items. The eight students included three boys (two African Americans, and one Asian) all in 6th grade, and five girls (two African Americans, one White, one Hispanic, and one who did not answer) which three of them were 6th and two were 8th graders. We focused on describing students' reasoning across all items which they saw in the interviews by analyzing data coming from different sources such as interview audio recordings and transcripts, the output of students' final response on each item, and field notes of interviewers. Students' responses were coded to determine the extent to which the option selected was reflective of their understanding as expressed during the interview. All coding was reviewed by the two researchers until an agreement was reached.

Your school's drama club is holding a raffle for teachers after the school play on Friday and Saturday nights. Here are the rules of the raffle:

- Each teacher can buy only 1 ticket each night.
- The drama club sells 20 raffle tickets each night.
- The principal randomly draws 1 winning ticket each night.
- A ticket can only be used on the night that it was bought.

What does "randomly draw" mean in this question?
Each raffle ticket is put in a hat. The hat is shaken to mix up the tickets. The principal draws 1 ticket without looking.

Mrs. Vail won the raffle on Friday!
Mrs. Vail bought a ticket for the Saturday raffle.
With which statement do you MOST agree?
a. Mrs. Vail will either win or lose, so she has a 50% chance of winning on Saturday.
b. Mrs. Vail has a better chance of winning on Saturday than she had on Friday.
c. Mrs. Vail has a lower chance of winning on Saturday than she had on Friday.
d. Mrs. Vail's chance of winning on Saturday has not changed.

Figure 1. Raffle item used in the cognitive interviews

RESULTS

Factors Contributing to Biased Reasoning

Different factors recognized in this study seemed to play an influential role in evoking the misconceptions. These factors were categorized into three main domains: (1) *General Understandings of Probability and Randomness*. (2) *General Item Characteristics* including item context, call-out boxes defining randomness, and graphics demonstrated in items. (3) *Aspects of Probability in Items* including the list of outcomes, and the ability to reduce outcomes to binary options or binary results.

General Understandings of Probability and Randomness. This was the most common factor observed in all students' misconception responses working as the key factor in evoking both biases. A lack of understanding about probability and randomness was observed in students' responses to introductory questions asking them to define probability and randomness, and in diagnostic items measuring the misconceptions. Various ideas were expressed by students to define probability such as *"possibility"*, *"percentage"*, *"how often something occurs"*, *"something out of a whole"*, *"how to understand something"*, *"the chance"*, and *"the ability to be right or wrong"*. Likewise, none of the students was able to define randomness while almost all of them considered random events as *"low probability"*, *"unpredictable"*, or *"rare events"*. Moreover, other ideas such as *"unknown results"*, *"random number generators"*, *"unusual phenomena"*, *"patternless processes"*, and *"unexpected events"* were expressed when explaining random events.

Such a weakness caused students to apply insubstantial reasonings to justify their misconception responses. For instance, some EL reasoners made up factitious likelihoods (unreal similarities) to justify their biased reasoning. Knowledge of probability concepts including the second axiom of probability (P[S] = 1) could inhibit the exhibition of biases, especially EL bias, in many students who applied the misconceptions. Many of the EL reasoners in this study assigned 50% as an equal chance to each of the outcomes in the sample space ignoring the fact that the sum of probabilities exceeds 1. For example, in the FidgetSpinners item (Figure 2), an EL reasoner who lacked general knowledge of probability argued that *"each color fidget spinner has 50% chance of being randomly drawn as each of them will either be picked or not."* Likewise, inadequate understanding of probability and randomness was influential in exhibiting II misconception as most of the II reasoners could not distinguish the independence of outcomes.

You won the ring toss game at the school carnival! Now you get to choose a fidget spinner from the prize bag. The prize bag has 3 green spinners, 5 red spinners, and 4 white spinners. All of the spinners are the same size.
You randomly choose 1 spinner from the bag.
Which color fidget spinner are you MOST likely to draw?
a. Green
b. Red
c. White
d. You are equally likely to draw a green, red, or white spinner.

What does "randomly choose" mean in this question?
You shake the bag to mix up the spinners. Then you choose 1 spinner without looking.

Figure 2. FidgetSpinners item used in the cognitive interviews

General Item Characteristics

(1) Item Context: Students who previously experienced the context of an item in real life mostly relied on their experiences and ignored their correct intuitions and/or quantifications. This happened to both EL and II reasoners who ignored verbal, numeric, and graphical information provided in items and merely applied their real-life experiences. For instance, most of the students who saw the Raffle item (Figure 1) applied their understandings and interpretations of real-life raffles/lotteries when responding to this item. EL reasoners argued that *"the chances are equal (50%) as you will either win or lose each night the raffle is held."* Also, II-Negative Recency reasoners explained that *"Mrs. Vail's chances of winning on Saturday night will be lower as winning the lottery has a very low chance."*

(2) Call-Out Boxes: The Call-Out Boxes were designed to define randomness in items whose context included phrases such as *random drawing*, *tossing a fair die*, and *flipping a fair coin*. However, these boxes often confused students and caused them to mistakenly consider the chance of an element equal to the chance of the category to which the element belongs. For example, in the FidgetSpinners item (Figure 2), students who picked the equally likely option confused the equal chance of each spinner mentioned in the Call-Out Box with the chance of each color spinner asked by the question. Students who picked equally likely options based on their misinterpretation of the concept of randomness usually argued: *"equal chances… since it is random and random means equal chances."* Similarly, EL reasoners picked equally likely options because they believed in the unpredictability of outcomes of random events and mentioned *"there is no control over outcomes because it is random… and random means anything can happen. So, chances are equal… ."*

(3) Graphics: In some items, the context included jars, boxes, or spinners, and therefore, we added graphics to make it easier for students to make sense of the context and be able to extract information easier. Meanwhile, some students had difficulties interpreting graphical information. For example, in the Spinner4a item (Figure 3), some EL reasoners argued that "*landing on each color depends on the roughness of the spin and the initial location of the arrow… Thus, the arrow will or will not land on each color and the chances are 50%"* The same type of reasoning might be seen in II reasoners' arguments. For example, if an item illustrates a box consisting of 10 red and 10 yellow balls where the red balls are depicted all over the box while yellow balls are concentrated in a corner, then a sequence of outcomes such as Red, Red, Red, Red, [?] looks normal to an II reasoner and they will argue that *"Red ball will be drawn next since Red balls are everywhere in the box."*

Aspects of Probability in Items.

(1) Presentation of Outcomes: How outcomes were presented in an item was a key factor in

evoking II bias. Most items designed to measure II bias presented a series of outcomes related to a random event and asked students to predict the next possible outcome. Outcome presentations included: (a) An outcome which is occurred in a row like T, T, T, T, T, [?]. (b) An outcome that is repeated frequently in a series of outcomes like T, T, T, H, H, T, [?]. (c) An easily recognizable pattern in a series of outcomes like H, H, T, H, H, T, [?]. Students with II-Positive Recency tendencies predict T in the first two sequences as T is the trend in the first sequence and repeated more than H in the second sequence. Though an II-Negative Recency reasoner predicts H as the next possible outcome in both first and second sequences and argues that "*there have been a lot of tails showing up so far, and therefore, it is time to have some heads."* On the other hand, a student who looks for patterns picks H in the third sequence since two Heads appear right after a Tail occurs. In all of the above cases, students focused heavily on the past results and ignored that the coin is fair and the chance for each side to show up in every coin flip will be 1/2.

The spinner has 4 equally sized sections.
When you spin the arrow, it can land on red, yellow, green, or blue.
Jan says, "The probability of landing on green on the 1st spin is 0.50 (50%) because it either lands on green or it doesn't."
Do you agree or disagree with Jan?
a. Agree
b. Disagree

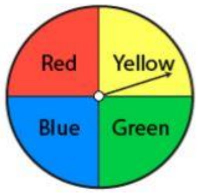

Figure 3. Spinner4a item used in the cognitive interviews

(2) Binary Outcomes: Items in which presented experiments with binary outcomes such as win/lose, or heads/tails evoked EL misconception most of the t. EL reasoners usually argued that *"the chances are equal since there are only two possible outcomes/results."* For instance, if a spinner consists of four equally sized colored sections, then an EL reasoner argues that *"the chance of landing on each color is the same (50%) since the arrow will or will not land on each color."*

(3) Samples Divided into Two Categories: In items which the elements were divided into two categories such as two colors, two boxes, two genders, etc., EL misconception was evoked. For example, if a box included 10 colored balls which seven are Red and three are Yellow, then an EL reasoner argues that *"the chance of randomly drawing each color ball will be the same (50%) since there are only two colors."* Similarly, if an item mentions that there are two boxes (Box A and Box B) which there are three red and yellow balls in box A, and two Red and eight yellow balls in box B, then a student with EL tendency argues that *"the chance of randomly drawing a Red or Yellow ball is the same (50%) since there are only two boxes."* Likewise, if an item mentions that there are six Boys and eight Girls in a class and the teacher wants to randomly draw one person to receive a prize, an EL reasoner concludes that *"the chances are the same (50%) for Boys and Girls as there are only two genders."*

(4) Proximity in the Frequency of Outcomes: This was an essential factor evoking EL bias in items presenting outcomes with similar frequencies. For instance, when the frequencies of three types of outcomes (black, brown, and white) were three, four, and three, an EL reasoner argued *"I think that they are the same. Because the colors are similar – the number of colors is the same – there's 3, 4, 3."*

Nevertheless, we observed some factors inhibiting the emergence of misconceptions including (1) Adequate knowledge of probability, (2) Enough quantitative skills, (3) *Higher Frequency Means Higher Chance* intuition which the student sorts out the frequency of outcomes in the sample and then assigns the highest probability to the outcome with the highest frequency. This intuition was applied more than any other factor to inhibit the exhibition of EL misconception. (4) Counterexamples applied to reject options including EL or II statements. Counterexamples were mostly used by students who exhibited lower EL exhibition rates.

Consistency

As Table 1 indicates, 38 opportunities measured EL bias while four students exhibited EL bias in their reasoning in 13 opportunities. Also, 13 opportunities measured only II bias which four students exhibited the misconception in seven of them. Further, 23 opportunities included both misconceptions in which EL bias was exhibited in eight opportunities by five students, though II bias was exhibited in four opportunities by only two students. Consequently, the exhibition rate for II bias in opportunities only measuring II bias was 7/13 (53.8%) and the exhibition rate for EL bias in opportunities only measuring EL bias was 13/38 (34%). Though EL bias was exhibited more than II bias in opportunities presenting both misconceptions at the same time (8/23 vs. 4/23).

Moreover, Table 1 shows that there was a total of 61 opportunities measuring EL bias (38 included only EL bias and 23 included both EL and II biases) while the bias was exhibited by six students in 21 opportunities. On the other hand, there were a total of 36 opportunities measuring II bias (13 included only II bias and 23 included both II and EL biases) while the bias was exhibited by four students in 11 opportunities. Therefore, the overall exhibition rates for EL and II biases were 21/61 (34.4%) and 11/36 (30.5%) respectively. Such close rates might indicate that both EL and II misconceptions were exhibited almost similarly in students' reasonings through a set of items (opportunities) as the difference between the two overall exhibition rates is only 3.9%. Nevertheless, the exhibition rate differences between the two biases were 19.8% $(53.8\% - 34\%)$ and 17.4% $(34.8\% - 17.4\%)$ in opportunities measuring only one of the biases and in opportunities including both misconceptions, respectively.

In general, it seems that EL bias was exhibited more consistently and with almost a similar exhibition rate throughout the different types of opportunities measuring the bias as the exhibition rates for the bias were 34% in items included only EL, 34.8% in items included EL and II, and 34.4% in items included either EL or EL and II. However, the exhibition rates for II bias were not as consistent as EL bias since the exhibition rates were 53.8% in items included only II, 17.4% in items included II and EL, and 30.5% in items included either II or II and EL.

Table 1. Bias Exhibition Rates and Students' Performances

	Grade Level	EL Rate in Items only Measured EL	II Rate in Items only Measured II	EL Rate in Items Measured EL and II	II Rate in Items Measured EL and II	Overall EL Rate	Overall II Rate
Student1	6th	1/3 (.33)	3/4 (.75)	0/2 (0.0)	1/2 (.5)	1/5 (.2)	4/6 (.67)
Student2	6th	2/5 (.4)	1/1 (1.0)	1/1 (1.0)	0/1 (0.0)	3/6 (.5)	1/2 (.5)
Student3	6th	6/6 (1.0)	0/1 (0.0)	2/2 (1.0)	0/2 (0.0)	8/8 (1.0)	0/3 (0.0)
Student4	6th	0/4 (0.0)	2/2 (1.0)	1/4 (.25)	3/4 (.75)	1/8 (.13)	5/6 (.83)
Student5	8th	0/2 (0.0)	0/1 (0.1)	2/5 (.4)	0/5 (0.0)	2/7 (.29)	0/6 (0.0)
Student6	8th	0/4 (0.0)	1/2 (.5)	0/4 (0.0)	0/4 (0.0)	0/8 (0.0)	1/6 (.17)
Student7	6th	0/7 (0.0)	0/1 (0.0)	0/2 (0.0)	0/2 (0.0)	0/9 (0.0)	0/3 (0.0)
Student8	6th	4/7 (.57)	0/1 (0.0)	2/3 (.67)	0/3 (0.0)	6/10 (.6)	0/4 (0.0)
Total Rate		13/38 (.34)	7/13 (.54)	8/23 (.35)	4/23 (.17)	21/61 (.34)	11/36 (.31)

Each row of Table 1 shows individual performances in three different types of opportunities measuring either only one or both biases. In general, we might categorize each student's performance regarding their bias tendencies into four classes by looking at their overall bias rates presented in the last two columns: (1) Bias-Free Reasoner which possesses neither of the two misconceptions, (2) EL Reasoner which only possesses EL bias, (3) II Reasoner which only possesses II bias, (4) EL and II Reasoner which possesses both biases. As Table 1 demonstrates, Student1, Student4, and Student6 have higher II exhibition rates compared to EL bias in all three types of opportunities. Thus, we would categorize them as II reasoners. However, the EL exhibition rate was higher for Student3, Student5, and Student8 and we would categorize them as EL reasoners. Meanwhile, Student2 was categorized in class 4 as the overall exhibition rates for both EL and II biases were 50% (3/6 and 1/2). On the other hand, Student7 was categorized in class 1 as they exhibited neither of the two biases in their reasoning. In summary, six students were either in class 2 or 3 (three II reasoners and three EL reasoners), while one student was in class 4, and one was in class 1. Overall, seven of the eight students exhibited at least one misconception in their responses.

Additionally, analyzing the responses given by students in each of the four classes revealed a strong relationship that exists between bias exhibition and evoking factors presented in items. Indeed, consistency in the exhibition of a specific bias occurred about presented evoking factors that directly influence the exhibition of that misconception. For instance, if a student is categorized in class 2, then it means that they consistently exhibited EL bias in a set of items, while all of the items in which the bias was expressed were similar in providing common factors evoking EL misconception.

DISCUSSION

In this study, we investigated the characteristics of tasks that might influence students' reasoning about probability concepts, and whether the students exhibit consistent reasoning when given a series of tasks designed to evoke specific probability biases. Findings from this study reveal that various factors can affect students' reasoning. Though some factors evoke only one of the biases, others are common in evoking both misconceptions. Furthermore, findings indicate that assessment of students' reasoning should happen based on a set of items instead of using one or two items. This way, we might be able to check for the consistency in bias exhibition and also to categorize a student in one of the four mentioned classes of reasoners by checking the misconception exhibition rates. Meanwhile, we should consider that consistency in exhibiting a specific bias highly depends on evoking factors presented in a set of items as consistency and evoking factors are interwoven concepts and should be studied together to help students improve their probabilistic reasoning.

Finally, though the number of students interviewed in this study was not large enough to generalize the findings, the results indicate that the overall misconception exhibition rates for EL and II biases were not as different as expected. Consequently, we could not conclude that the participants in this study exhibited one misconception more than the other. However, we observed that in items that measured both misconceptions, EL bias was exhibited with a higher rate. Further, outcomes of this study might be used to support the teaching and learning of foundational concepts in middle-grade probability which can improve data and statistical literacy at secondary-level and in diverse contexts. It is because data and statistical literacy for good citizenship rely on reasoning about uncertainty and considering the likelihood of events. Our results have implications for the types of tasks and assessment items that can be used to assist students in improving their probabilistic reasoning.

ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170441. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). *Research on teaching and learning probability*. Cham: Springer.

Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika*, *79*(3), 403-425.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In D. A. Grouws (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Charlotte NC: Information Age Publishing.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

LeCoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, *23*(6), 557-568.

Schoenfeld, A.H., Smith, J.P., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject-matter. In R. Glaser (Ed.), *Advances in Instructional Psychology, Vol. 4, 55-175*. Hillsdale, NJ: Erlbaum.

Smith, J. P., diSessa, A.A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of Learning Sciences, 3*(2), 115-163.

Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.). *Exploring probability in schools: Challenges for teaching and learning* (pp. 345-366). New York: Springer