STUDENT-GENERATED CONNECTIONS IN LEARNING ABOUT COMPOUND PROBABILITY AND THEIR EMERGENCE DURING INSTRUCTION

RANDALL E. GROTH
Salisbury University
regroth@salisbury.edu

MEGAN RICKARDS
Salisbury University
mrickards1@gulls.salisbury.edu

ELIZABETH ROEHM
University of Notre Dame
eroehm@nd.edu

ABSTRACT

In this report, we analyze students' learning of compound probability by describing connections they generated while engaged with tasks involving two independent events. Several of their connections were compatible with the development of expertise, such as recognizing the need to determine sample spaces across a variety of situations and noting structural similarities among tasks, even when their task solutions were incomplete from a normative standpoint. Students reasoned about dimensions of context, variation, mathematical structure, sample space, and probability quantification. We describe the extent to which they coordinated these dimensions. We also describe teaching moves, such as posing idealized situations and shifting to structurally similar tasks, that prompted students to attend to multiple relevant task dimensions.

Keywords: *Statistics education research; Compound probability; Connections; Qualitative research*

1. INTRODUCTION

The ability to see connections among tasks in a given domain is commonly considered a hallmark of expertise (Kimball & Holyoak, 2000). For example, suppose a student has learned to solve the problem shown in Figure 1, which involves flipping two coins. Also suppose the student is later given the assessment item shown in Figure 2, which is analogous but involves two spinners instead of two coins (Zawojewski & Shaughnessy, 2000). A student who solves the two-spinner task using knowledge related to the two-coin task exhibits a degree of transfer, in the traditional sense of the word (Bransford & Schwartz, 1999; Chow & Van Haneghan, 2016), by recognizing that ideas used to solve the first task are also pertinent to the second one.

> Two fair coins are part of a carnival game. A player wins a prize only when both coins come up heads after each coin has been flipped once.
>
> James thinks he has a 50-50 chance of winning. Do you agree? Justify your answer.

Figure 1. Two-coin task involving compound probability and fairness.

The two fair spinners shown above are part of a carnival game. A player wins a prize only when both arrows land on black after each spinner has been spun once.

James thinks he has a 50-50 chance of winning. Do you agree? Justify your answer.

NAEP Question ID: 1996-12M12 #9 M070501

Figure 2. Two-spinner task to assess knowledge of compound probability and fairness (Item source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP, 1996 Mathematics Assessment).

From a teaching and learning standpoint, it is important to note that a student who does not yet recognize all the structural similarities between the tasks shown in Figures 1 and 2 may nonetheless bring some smaller, yet important, connections to bear in approaching both. For instance, a student might recognize the general principle that there is a need to determine the sample space in each task even if they are not able to determine sample spaces precisely. There is value in recognizing when students make such incremental connections because they reveal smaller changes in learning that can be leveraged to help students develop expertise (Pratt, 2000; Wagner, 2010). In the present study, we examine students' generation of such connections in the domain of compound probability. The specific research questions we address are: (i) What connections does a small group of students generate when reasoning about compound probability tasks involving two independent events? and (ii) Under what instructional circumstances do these connections emerge? In posing these research questions, we sought to construct a detailed localized theory of teaching and learning from an in-depth qualitative examination of a single case.

2. THEORETICAL PERSPECTIVE

Many traditional studies of transfer rely upon students' responses to paired tasks such as those shown in Figures 1 and 2 to assess connections students make across situations (Wagner, 2010). Although there is value in such an approach, investigating only students' final responses to such tasks leaves aspects of their thinking unexplored. For example, students who are not completely successful solving both tasks might still recognize the need to examine the mathematical structures of each situation or to determine the sample spaces in each case. Such small-scale connections can remain hidden unless a suitable theoretical lens is brought to bear.

We used an actor-oriented transfer (AOT) perspective to study small and large scale connections students generated when solving compound probability tasks. Lobato and Siebert (2002) wrote, "Actor-oriented transfer is defined as the personal construction of relations of similarity between activities, or how 'actors' see situations as similar" (p. 89). From this perspective, student-generated connections are central objects of study (Lockwood, 2011). Connection building can be conceived of as "the expansion of instructional or everyday experiences beyond the conditions of initial learning" (Lobato, 2012, p. 232) and requires studying "the influence of a learner's prior activities on her activity in novel situations" (Lobato, 2012, p. 233). Hence, to detect student-constructed connections, we attended to students' discourse as they engaged with tasks selected to introduce new contexts and mathematical structures in a gradual, incremental manner.

The AOT perspective emphasizes examining similarities individuals (or "actors") see among situations rather than just assessing whether they see similarities apparent to experts, because many transfer situations "include a dimension of complexity that is hidden from the view of an expert until one investigates students' understanding of the transfer situations more closely" (Lobato, 2012, p. 241). AOT entails examining the connections learners generate among instructional tasks rather than just whether or not they have attained expertise (Lockwood, 2011). Assessing whether learners have attained expertise, as in many traditional studies of transfer, is still an important goal; the AOT perspective provides a complementary portrait of smaller connections generated along the way (Lobato, 2012).

The AOT perspective entails a fundamental metaphor of transfer as construction rather than application of knowledge (Lobato & Siebert, 2002). In an AOT framework, "one assumes that learners are making connections between situations nearly all the time, guided by aspects of the situation that they find personally salient" (Lobato, 2003, p. 19). It is axiomatic that individuals do not all come to see similarities between situations in the same way (Wagner, 2010). These differences among learners' pathways make it problematic to assume that knowledge of one task is simply applied to another (Lockwood, 2011). Hence, AOT studies aim "to understand the interpretative nature of the connections that people construct between learning and transfer situations, as well as the socially situated processes that give rise to those connections" (Lobato, 2012, p. 239).

3. STUDENT-GENERATED CONNECTIONS IN LITERATURE ABOUT LEARNING COMPOUND PROBABILITY

In the present study, we sought to understand the nature of student-generated connections and aspects of the classroom environment that would foster active making of connections. As we began to design an instructional environment, we searched the literature to anticipate the reasoning patterns students exhibit (Stein et al., 2008) when trying to construct connections across compound probability tasks. We were particularly interested in helping students construct connections that would help them reason about compound probability tasks involving two independent events that can be modeled concretely. Our attention was drawn to this type of task because of the difficulties it has caused students. When the spinner task shown in Figure 2 was posed to a large, representative sample, only 8% of students in their final year of compulsory schooling were able to provide a correct response and justification; approximately half simply agreed that there was a 50-50 chance of winning (Shaughnessy, 2007). In the present study, we sought to better understand reasons that may underlie this type of performance and to develop strategies to help students move toward expertise on this type of task. In this section, we summarize some existing literature that helped us anticipate some of the reasoning students might exhibit during instruction focused on helping them learn to solve tasks like those shown in Figures 1 and 2.

Given our AOT theoretical perspective, we sought to anticipate both normative and non-normative student-generated connections (Lockwood, 2011). We use the word "normative" to describe disciplinary thinking commonly accepted among experts (Shaughnessy 2007). We found three salient themes in the literature about teaching and learning with two-stage tasks like those shown in Figure 1 and 2 that helped us anticipate students' normative and non-normative connections: (i) students may approach compound probability problems with strategies appropriate for simple probability problems, (ii) attending to order when relevant to determining sample space is a complex cognitive process, and (iii) students may use visual models to solve compound probability problems when prompted but not employ them otherwise.

One prevalent theme in research is that students often appear to approach compound probability situations as simple probability ones (Iversen & Nilsson, 2019; Lysoe, 2008). For example, Shaughnessy and Ciancetta (2002) had students work with a single spinner in one task and later with two spinners, as shown in Figure 2. In the two-spinner task, students were to decide if there was a 50-50 chance of both landing on black. Most students looked at the total amount of black area across the two spinners, saw it was 50%, and decided there was a 50-50 chance of both landing on black. Watson and Kelly (2004) reported similar findings. Iversen and Nilsson (2019) interpreted such findings to indicate that students dealt with the compound probability situation "by applying methods working in a simple stochastic environment" (p. 3). This student strategy is prevalent and well-documented in the literature (Fischbein & Snarch, 1997; Pratt, 2000).

One might characterize students' non-normative approaches to the two-spinner task as indicative of over-generalization and negative transfer (Chen & Daehler, 1989). From another perspective, one might acknowledge that students have not developed expertise for this type of problem; yet, one might also recognize that students have constructed the generalization of needing to account for area in some manner when determining probability in such cases. Recognizing such a student-generated connection is useful from a teaching perspective because it brings to light a cognitive asset to leverage as teachers design sequences of activities to help students continue to work toward expertise.

In many compound probability tasks with two independent events, the sample space can be represented as a collection of ordered pairs. For example, the set of outcomes for tossing two fair six-sided dice consists of 36 ordered pairs. Learning to account for order in such situations is non-trivial for students (Alston & Maher, 2003; English & Watson, 2016; Iversen & Nilsson, 2019). Vidakovic et al. (1998) noted that some students considered the outcomes (5, 4) and (4, 5) to be the same because in each case the sum is 9. One could hypothesize that such thinking occurred, in part, because students used the commutative property of addition in a situation where it was not applicable (Alston & Maher, 2003). One might characterize this as another instance of negative transfer, because a generalization useful for a variety of situations (in this case, the commutative property) was applied in an unintended, perhaps unanticipated, way to a new situation.

From another perspective, characterizations of student thinking about the two-dice situation other than negative transfer might be made. Although the students did not count outcomes as intended, they did recognize the need to systematically count them. Arguably, students who know when to attempt to count outcomes are further along the path to expertise than those who rely solely on carrying out game trials or non-mathematical considerations in determining probabilities of outcomes. Students may also have constructed an internally coherent generalization for what it means for outcomes to be the "same" (i.e., yielding the same sum) even though it differs from how the word might be used in normative discourse.

Visual representations such as tree diagrams (Konold, 1996) and area models (Ron et al., 2017) are frequently used to help students count outcomes and map two-dimensional sample spaces. Ron et al. found that students successfully used visual models when they were included in task statements but had less success when no such prompts were included. Others have documented a similar phenomenon (Iversen & Nilsson, 2019). One might characterize this phenomenon as indicating no transfer of learning; from another perspective, however, underlying reasons for it are of interest. Deeper analyses can reveal similarities and differences students see among task characteristics and explain why they believe certain tools to be relevant to one situation but not another. Such insights can then be used to design subsequent instruction. Situating research in classroom contexts can allow researchers to gain such insights about students' thinking along with how and when students generate normative and non-normative connections during instruction.

English and Watson (2016) mapped a research-based pathway that can be used to design instruction that fosters children's learning of compound probability. The pathway included phases of doing concrete experimentation to produce data related to a probability situation, organizing and representing the data produced, and interpreting the results in relation to the given task (Batanero et al., 2005). They used the pathway to support and study children's learning during a task that involved flipping one coin and a subsequent task that involved flipping two. Children's initial expectations, expressed as predictions about the outcomes of each coin task, were elicited at the start of instruction. These predictions were revisited as students gathered, represented, and analyzed more and more data from coin-flipping trials. These activities were used as a basis for helping students relate observed relative frequencies to theoretical probabilities. As they progressed from experimental relative frequencies to theoretical probabilities in this manner, students were prompted to reflect on connections among their intuitions, expectations, relative frequencies, and observed random variation. Through the process, English and Watson found that students "developed a deeper understanding of the relationship between relative frequency of outcomes and theoretical probability as well as their respective associations with variation and expectation" (p. 28). Hence, having students make predictions about outcomes of compound probability tasks, test and refine the predictions by gathering and analyzing data, and then link the results to formal theoretical probabilities has the potential to enhance student learning.

4. METHOD

4.1. DESIGN-BASED RESEARCH PERSPECTIVE

In the present study, we investigated students' generation of connections related to compound probability in a classroom context. We worked from a design-based research perspective, which emphasizes attending carefully to students' reasoning patterns during instruction in order to make ongoing conjectures about how to design and sequence learning experiences for them (Bakker & van Eerde, 2015; Cobb et al., 2017). This perspective resonates with the goal of creating a learning environment to support students in gradually making connections that are useful across a variety of similar contexts involving compound probabilities. In the present study, gathering qualitative classroom data from such an environment allowed us to detect several connections students generated over the course of instruction as we made teaching adjustments intended to optimize students' learning.

In design-based research, an overarching hypothetical learning trajectory is designed at the outset and then revised in response to classroom data as the study is carried out. Hence, at the outset of the study, we constructed a macro-level sequence to outline a beginning hypothetical learning trajectory that was gradually revised as a result of micro-level sequences taking shape in response to observed student thinking. In this report, we use "macro level" to describe a sequence of instructional scenarios that progressively build upon one another over multiple lessons and "micro level" to refer to sequences of learning experiences within one given instructional scenario.

4.2. PARTICIPANTS

We studied a small group of students in anticipation of time-intensive, detailed qualitative analyses of reasoning over multiple lessons and interviews. Four children, two boys and two girls, participated in the study: Tom, Laura, Aiden, and Emilia (pseudonyms). Three of them were students of color: Tom and Laura were Asian, Emilia was Hispanic, and Aiden was Caucasian. The four participants formed a self-contained group and were the only children in the classroom for the study. The parents of all four participants had submitted applications for their children to participate in summer mathematics instruction at the authors' home campus. Other children brought to campus for summer mathematics instruction formed their own self-contained groups, focusing on different, non-probability related subject matter at various grade levels in different classrooms with different instructors.

Groups for summer mathematics instruction were formed with the intent of balancing the numbers of boys and girls, having students from different school settings, and having a diverse range of learning needs represented. To attain diversity along these dimensions, we had all parents applying for the program complete a questionnaire about student characteristics and needs. Questionnaire responses indicated that all participants were preparing to enter seventh grade at their respective schools. Tom was 12 years old at the time of the study and attended a public school. Laura was also 12 years old; she attended a different public school. Aiden turned from 11 to 12 years old midway through the study. He was homeschooled by his parents. Emilia was 12 years old and attended a private school. Regarding student characteristics and learning needs, Aiden's parents reported that he had been diagnosed with dysgraphia but also had strong skills in mental mathematics. The other parents did not identify specific learning disabilities or strengths, but they all did express goals for the program. Emilia's mother wrote that she hoped the program would help Emilia develop a more positive attitude toward mathematics. Laura's mother wrote that she would like to see Laura's problem-solving skills develop. Tom's mother wanted him to be introduced to "other fun ways for math."

4.3. PROCEDURE

Participants' parents agreed to bring them to nine consecutive weekly sessions. The first session was used for 30-minute individual pre-interviews, the next seven sessions were one-hour lessons involving the entire participant group, and the final session was used for 30-minute individual post-interviews. Parents received a $20 stipend for each session their child attended and a $50 bonus for perfect attendance. All participants attended all sessions except Emilia, who missed one lesson. All sessions were video recorded and transcribed, and students' written work was retained for analysis. Next, we describe the interviews and lessons in detail.

At the start of each pre- and post-interview, students were given the task shown in Figure 3 so we could obtain baseline information about their knowledge of simple probability. The compound probability tasks shown in Figures 1 and 2 were then administered. We hypothesized that participants would not initially exhibit expertise on the interview tasks because their school curriculum (Common Core State Standards Initiative, 2010) would not include probability until the fall semester immediately following the summer sessions. The two-spinner task (Figure 2) was selected to create continuity with previous studies (e.g., Shaughnessy & Ciancetta, 2002; Watson & Kelly, 2004). The two-coin task (Figure 1) was designed to be analogous to the two-spinner task. We were interested in comparing students' approaches to the tasks before and after instruction in order to investigate the extent to which they might make connections between the two.

> Imagine you are playing a coin-tossing game against a friend. You take turns tossing a coin. If it is heads, you win a point. If it is tails, your friend wins a point. The person with the most points at the end of the game wins. Each person gets the same number of turns. Is this a fair game? Why or why not?

Figure 3. Task administered at the start of each pre- and post-interview.

During pre-interviews, we gave each participant the two-coin task (Figure 1) before the two-spinner task (Figure 2). During post-interviews, we changed the order of the two tasks because we had students work with situations similar to the two-coin task during instruction, but we intentionally avoided introducing spinners. We placed the two-spinner task first during post-interviews to avoid implicitly suggesting that strategies they had learned for the two-coin task should be used on the two-spinner task. After posing each task, we asked follow-up probing questions about students' strategies (Moyer & Milewicz, 2002) to learn as much as possible about their thinking. Some of the follow-ups were general "how" and "why" probes, and others were specific questions about strategies students used (contextualized examples of both types of probes are provided in section 5.1 of this article). Students' responses were captured on video, and the written work they produced was also retained for analysis.

Our first step in planning for instruction was to form an a priori hypothetical learning trajectory to construct a macro-level sequence of instructional scenarios. We aimed to help students compare similar scenarios that became progressively more complex in order to facilitate structural reasoning and connection generation (Barnett & Ceci, 2002). In accord with English and Watson (2016), we initially decided to have students make predictions about outcomes of probability tasks we posed, gather data to test and refine the predictions, and then use these experiences as a basis for learning about formal theoretical probabilities. We conjectured that these activities would put students in position to describe sample spaces for compound probability situations using organized lists and tree diagrams. These sample space representations could then be used to determine theoretical probabilities for outcomes of interest.

All lessons were planned collaboratively by the three authors of this article. The second and third authors taught each lesson while the first author observed. Lessons involved a combination of individual work, pair work, and whole-group work, with whole-group work being the predominant mode of instruction. During whole-group discourse, students were encouraged to share their ideas, think aloud with others, and assess one another's reasoning, because such exploratory talk also promotes transfer and making connections (Webb et al., 2017). Being able to hear and record students' exploratory talk also helped us attain our design-based research goals of constructing a localized theory of student learning and making incremental changes to the learning trajectory in response to the connections students verbalized. Our initial macro-level hypotheses were revised as we gathered and analyzed micro-level student data from each lesson. The instructional decisions made at the micro level and how they were influenced by our interpretations data are described in the results section of this report. A macro-level view of the four instructional scenarios that gradually took shape as a result of our ongoing analyses appears in Figure 4. Key similarities and differences among the scenarios are also summarized in Figure 4.

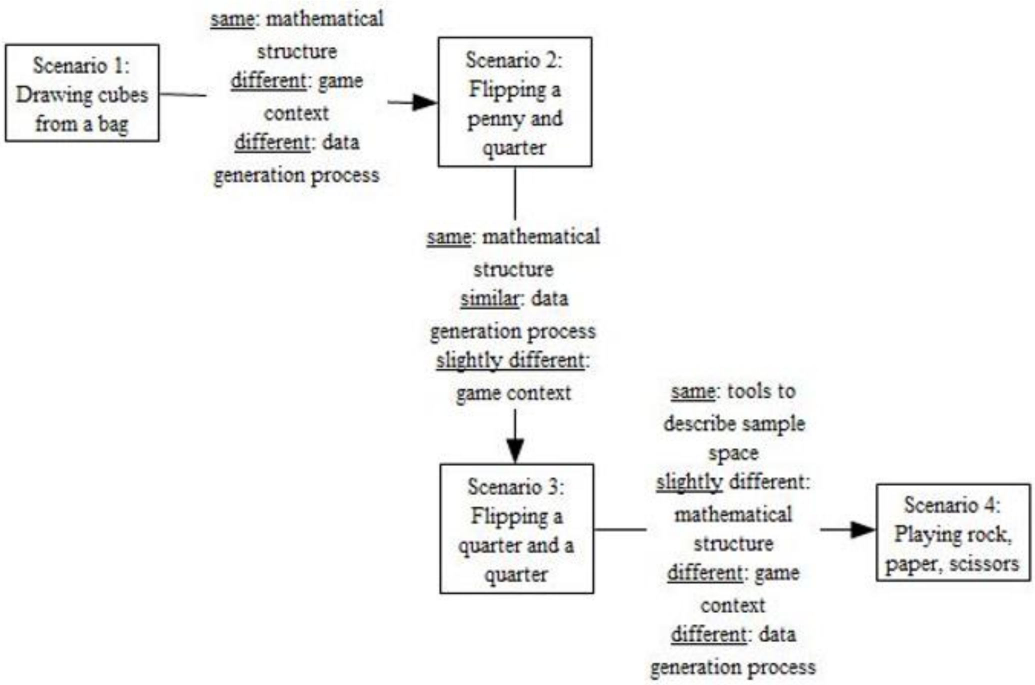


Figure 4. A macro-level chronological view of the four instructional scenarios for the study and their key similarities and differences in context, data generation processes, and mathematical structure.

Each instructional scenario shown in Figure 4 spanned 1–3 lessons. The scenarios were designed to be similar to one another yet gradually introduced new contexts, data generation mechanisms, and mathematical structures as students' reasoning suggested they would benefit from such shifts. In the first scenario (Lessons 1–3), students drew cubes from a bag containing five blue and five red. Early in the scenario, they drew one cube with replacement to study simple probability. Later in the scenario, they drew one cube from the bag, replaced it, drew another, replaced it, and recorded how many blues were obtained in each pair they drew. In the second scenario (Lessons 4–5), they flipped a quarter and a penny simultaneously. Students drew pictures of possible outcomes and conducted trials to generate data. They then moved to a third scenario (Lesson 6), which dealt with flipping two quarters simultaneously. Students produced organized lists to account for all possible outcomes. Structural similarities of the sample spaces for key tasks related to these three initial scenarios are included in the "D4" portion of Figure 15 later in this report. The fourth scenario (Lesson 7) involved generating data by playing "Rock, Paper, Scissors" (Nelson & Williams, 2008) and using and interpreting organized lists and tree diagrams representing the outcomes. The fourth scenario entailed a slight shift in underlying mathematical structure with an increase in the number of elements in the sample space as compared to the first three scenarios. We explain our design and sequencing decisions in more detail throughout the results section of this report to describe how our instructional decisions were responsive to the thinking students exhibited during specific lessons.

4.4. DATA ANALYSIS

Design-based research requires analyzing data on a continuous basis throughout a study in order to make grounded conjectures about optimal next steps in teaching (Bakker & van Eerde, 2015; Cobb et al., 2017). Accordingly, data analysis meetings involving all authors occurred after each interview and lesson. During each data analysis meeting, we read the session transcript while watching the accompanying video and reviewing participants' written work. Individual student responses to tasks were our units of analysis, but these generally occurred in the context of whole-class discourse, so watching the classroom video while coding participants' contributions helped us contextualize our findings and relate them to specific parts of lessons. Relating participants' responses to specific lesson tasks also helped us make conjectures about how to improve student learning in each subsequent lesson.

We began data analysis with open coding (Corbin & Strauss, 2008; Miles et al., 2020) of individual student responses to tasks. Given our AOT framework, we coded student utterances and written work that provided evidence of students' "personal construction of relations of similarity between activities" (Lobato & Siebert, 2002, p. 89) to identify connections they made during the study. We focused on student connections between their prior and new experiences, knowledge, and reasoning strategies. These connections were at times compatible with normative reasoning, such as when students mapped concrete experiences from game trials in class to formal statistical representations using technology. They were also sometimes incompatible with normative reasoning, such as when students conceptualized the fairness of a game only in terms of the number of turns taken. Such non-normative connections reflected the use of prior experiences (e.g., with games of chance) to approach a new classroom task, yet they were not complete from a normative perspective. As we identified student-generated connections, we created short descriptive codes for each one (e.g., students' mapping of concrete game trials to representations with technology was coded "MCT" and students' focus solely on the number of turns in a game when deciding on its fairness was coded "FT"). We continuously discussed codes as we created them to reach consensus on our characterizations of the data. As we collaboratively generated codes, they were recorded in a code book (DeCuir-Gunby et al., 2011) along with illustrative data excerpts to help us decide if an existing code should be used for a given portion of data or if a new one was needed.

Coding was carried out segment-by-segment for each interview and lesson. Boundaries of segments were defined using the notion of activity re-direction (Lineback, 2015). Such re-directions occur when teachers change the course of the lesson by posing a new task or question for students to consider. Hence, each interview question was considered its own segment, and each new task posed during instruction marked the beginning of a segment. During retrospective analysis, the codes were organized using time-ordered matrices (Miles et al., 2020) summarizing segments to facilitate the construction of a chronological account of student-generated connections over the course of the study.

Node-link diagrams (Nesbit & Adesope, 2006; Wheeldon & Åhlberg, 2011) were used to situate student-generated connections within the contexts of interviews and lessons. The nodes in each diagram were labeled with the incrementally more complex connections we hoped students would construct. Directional inks were inserted between nodes to indicate embedded opportunities to observe and facilitate student-generated connections. The node-link diagrams provided working outlines for our descriptions of students' reasoning in response to tasks. Visual diagramming software, *Inspiration*™ (2006), allowed for efficient revision of the diagrams as they were compared against the data. A sample node-link diagram for an instructional scenario from our study appears in Figure 5.

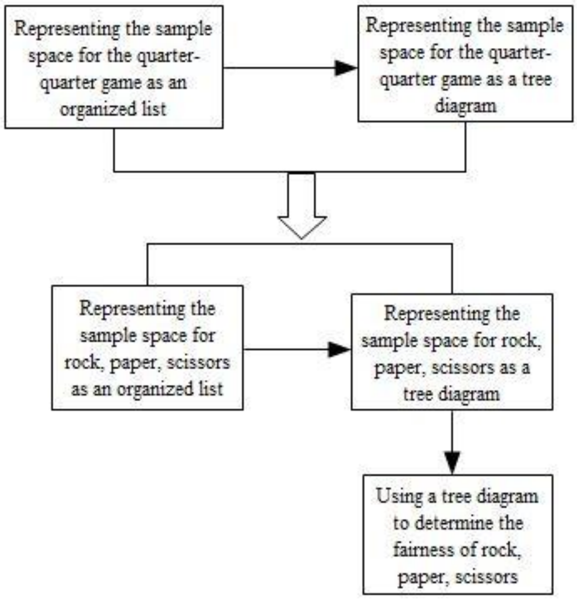


Figure 5. A sample node-link diagram used to characterize goals of an instructional scenario a game of "rock, paper, scissors".

5. RESULTS

Results are arranged chronologically to preserve the original flow of events during interviews and lessons. We begin with pre-interviews, continue with the four instructional scenarios summarized in Figure 4, and conclude with post-interview results. A summary of the student-generated connections we observed during the study is provided in Figure 6A. The connections were categorized according to the perceived relationships they indicated between prior and new experiences, knowledge, and reasoning strategies. The top three rectangles (A–I) contain connections that build toward normative expertise. The bottom rectangle (J–L) contains connections traditionally considered to indicate over-generalization and negative transfer.

Making connections between mathematically isomorphic tasks set in different everyday contexts

A: Using same learned strategy to determine the fairness of games with the same mathematical structure set in different everyday contexts.
B: Using same strategy to generate partial lists of outcomes for games with the same mathematical structure set in different everyday contexts.
C: Explicitly noting similarities in mathematical structure between games set in different everyday contexts

Making connections between mathematically isomorphic tasks set in similar everyday contexts

D: Linking activity of determining sample space to activity of determining probabilities
E: Linking activity of conducting empirical trials to calculating theoretical probabilities
F: Linking activity of conducting game trials to the activity of determining sample space
G: Using just one element of experience playing games to determine fairness of games (attending solely to the number of turns or mentioning "chance" of winning without quantification)

Connecting data and context to relevant mathematical representations

H: Using knowledge of possible game outcomes to read a tree diagram
I: Linking activities of conducting and tabulating game trials to activities of graph reading and construction

Connections incompatible with normative thinking

J: Using simple probability strategy to determine compound probability
K: Using deterministic thinking to answer a stochastic question (outcome approach or equiprobability bias)
L: Using informal reasoning incompatible with normative stochastic analysis (often preoccupation with what experts would consider "surface features" of tasks)

Figure 6A. Student-constructed connections observed over the course of the study.

A timeline representation of when each connection was evident for each student appears in Figure 6B. The lighter shading in a quadrant indicates evidence of incomplete or inconsistent use of a given connection by a student during an instructional scenario. The quadrants are arranged to match the spatial positioning of students around the shared classroom table during lessons, with the top of the circle corresponding to the front of the room. The spatial positioning convention is maintained for consistency for pre- and post-interviews, though interviews were individual rather than group sessions. Rows above the horizontal timeline represent connections building toward normative expertise rows below the timeline represent connections traditionally considered indicative of negative transfer. The letters on the vertical axis correspond to the student-generated connections identified in Figure 6A.

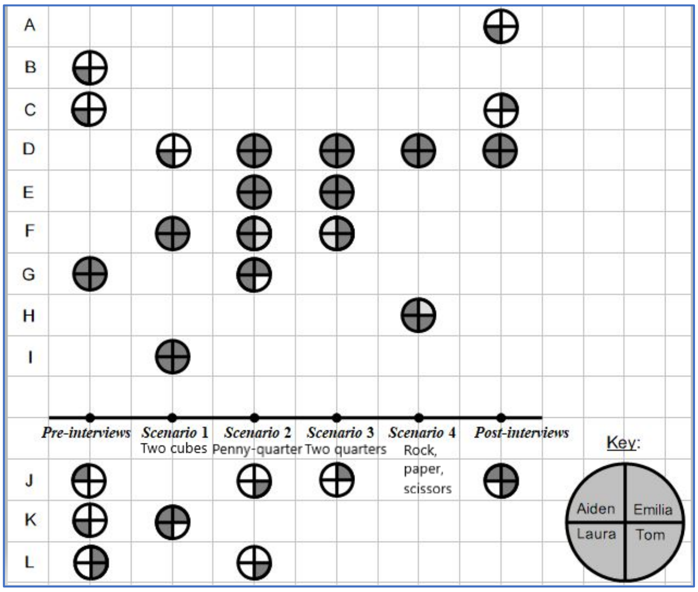


Figure 6B. The figure shows when different student-generated connections (shown in Figure 6A) were observed.

5.1. PRE-INTERVIEWS

During pre-interviews, we obtained data suggesting student-generated connections related to: (i) assessing fairness in a colloquial sense and mathematical sense, (ii) identifying analogous aspects of the two-coin task (Figure 1) and the two-spinner task (Figure 2), and (iii) determining the probabilities of winning the two-coin game and the two-spinner game.

Participants appeared to draw mainly upon colloquial generalizations about fairness in responding to the first interview task about a game that involved flipping one coin (Figure 3). All four students focused solely on the number of turns each player received during the game in judging its fairness. During his pre-interview, Tom, for example, responded, "It is fair because you get the same amount of turns." In their own individual interviews, the other three students similarly judged the game to be fair because each player received the same number of turns. As we planned the first lesson for the group, we did not want to discourage them from considering the number of turns in assessing fairness, because such considerations are relevant, but we aimed to expand their conceptions of fairness to encompass other relevant connections. Specifically, noticing that none of the students quantified each player's chance of winning the game, we aimed to help them recognize that quantification of probability is also relevant to determining fairness in such situations.

We observed greater variation in students' recognition of analogous elements of the two-coin and two-spinner tasks. Laura was the only one of the four students to explicitly note similarities between the two tasks. The following exchange took place when she was given the two-spinner task after the two-coin task:

Laura:	I'm kind of getting confused because these are both basically the same question.
Intv:	How are they similar?
Laura:	Because, um, it is, so, because basically heads would be either black or white, and tails would be black or white.
Intv:	OK.
Laura:	And if, since, it says both of them need to land on black after it wins, it's like kind of similar.

None of the other students explicitly spoke of the two-coin and two-spinner tasks as being similar during pre-interviews. Evidence that Tom and Emilia saw the two-coin and two-spinner tasks as distinct from one another could be seen in their approaches to them. The two concentrated on physical aspects of carrying out trials only in the two-spinner context. Tom, for example, wrote, "You have to have the perfect hit to land it on the black." When asked to explain, Tom stated, "It's on the white and like you have to hit it like lightly (makes flicking motion) or like medium to get it on the black; but someone might, they have to hit it really hard." Similarly, Emilia reasoned, "If you both spin them, they'll probably land on the white side, or it really depends on how hard you spin it, right?"

Students' attention to the physical aspects of spinning created a dilemma about the appropriate role of spinner contexts in our lessons. We took Tom and Emilia's reasoning to indicate they saw spinner outcomes as being within the control of the user rather than being stochastic in nature (Pratt, 2000). Ultimately, we decided to set the spinner context aside for our short seven-lesson sequence to observe whether connections students constructed in other compound probability contexts would help them with the spinner context when we returned to it during post-interviews.

Along with explicitly noting similarities between the two compound probability games (Figure 1 and 2), Laura used the same approach to decide if there was a 50-50 chance of winning each one. She named three possible outcomes in the two-coin game: heads on both, tails on both, and heads on one and tails on the other. For the two-spinner game she named black on both, white on both, and one black and one white as possible outcomes. Although she had apparently constructed the generalization that there is a need to determine the sample space in such situations, she did not yet account for order in either case. She also did not yet use her partial lists of outcomes to quantify probabilities. Instead, she stated, "I think he would either have a 100% or 0% winning because if just one arrow lands on white, he would lose the whole game, but if both landed on black, he would have a 100% chance of winning." This type of thinking is resonant with the outcome approach (Konold, 1995); Laura appeared to believe she was to give a deterministic prediction about the outcome of one trial rather than quantifying its probability. We took Laura's responses to indicate that the task of connecting sample space elements to their associated probabilities would need explicit attention during instruction.

Aiden and Tom diverged in their thinking about the probability of winning each game. Even though Aiden did not explicitly note similarities between the two-coin and two-spinner games, he used the same approach to state the chance of winning each one. Aiden looked at the total amount of black and white on the two spinners (Figure 2) and said there was a 50-50 chance of winning "because it's half black and half white," as many students in previous studies have done (e.g., Shaughnessy & Ciancetta, 2002; Watson & Kelly, 2004). For the two-coin task, Aiden similarly reasoned that half the coin sides were heads and half were tails. Tom's approaches to the two tasks differed from one another. He concluded there was a 50-50 chance of winning the two-coin game. He did not list sample space elements but assigned a value of 25 to each coin and added the amounts, apparently attending to monetary value rather than probability. For the spinner game, as noted earlier, Tom focused solely on the physical process of spinning rather than on area. He said, "You have to have the perfect hit to land on the black." When asked how often a "perfect hit" would occur, he offered a subjective estimate: "like, it has like a 10% chance, I think." Hence, during classroom discourse, we aimed to draw out and develop Aiden's implicit perception of mathematical similarities between such tasks, in part to help Tom recognize the need to attend to mathematical structure and area in such situations.

5.2. INSTRUCTIONAL SCENARIO 1: DRAWING CUBES

For the first instructional scenario, we used the context of drawing colored cubes from a bag. Just as English and Watson (2016) progressed from a one-coin task to a two-coin task with students, we progressed from one cube draw to two independent draws. Along with forming a viable instructional sequence, this progression allowed us to examine which student-generated connections used for structurally similar pre-interview tasks might be used to reason in a new context.

Our first lesson with students focused on simple probability in the context of drawing cubes from a bag. Because all four students focused only on the number of turns per player during pre-interviews, we built the first lesson to help them expand their ideas of fairness and begin to quantify probabilities. We had students take the same number of turns drawing one cube from a bag, with replacement. The bag contained 2 blue cubes and 8 red, and each player scored a point when their assigned color was drawn. After taking several turns and recording results, all participants agreed the game was not fair, even though each player was given the same number of turns. Laura explained, "Because I kept getting reds, I said to you that I didn't think there was any blue in it [the bag]; so, I think there's more red than blue." We then showed students the contents of the bag to emphasize the importance of taking each player's probability of winning into consideration, and not just number of turns, when analyzing games.

During the first lesson, students at times claimed the cube game was unfair because of the players' actions. This was particularly apparent after the first round; when Tom was asked if he considered the game to be fair, he said it was not, admitting, "I looked in the bag." So, to focus participants' attention on mathematical considerations, during subsequent lessons, we at times had them simulate games with a *TinkerPlots*™ (Konold & Miller, 2011) document we provided, emphasizing the computer had no vested interest in who won. As with the other instructional activities, we actively elicited students' thinking about simulated results so this information could be captured in our classroom video data (researchers seeking to delve more deeply into this aspect may consider augmenting data collection with screen-capture software recordings). Detailed descriptions of students' engagement with *TinkerPlots*™ are provided later in this section of the article.

In lesson 2, we had students draw two cubes with replacement. They took turns drawing one cube, replacing it, drawing another, replacing it, and then recording the colors of each pair of cubes drawn. Each bag contained the same number of red and blue cubes. As students gathered and represented data from game play, they identified the possible outcomes as drawing two blue, one of each color, and no blue. So, initially, we had them sort observed results into those three categories in a data display (e.g., as in Figure 7). As they worked with these data displays, we aimed to have them read the data, read between the data, and read beyond the data (Friel et al., 2001). Reading the data involved reading values from displays used to track results of trials. Reading between the data involved comparing frequencies of results within displays to one another. Reading beyond the data was required to compare trends and characteristics across different data displays. We also asked students to read behind the data (Shaughnessy, 2007) to explain why graphs took shape as they did, hoping to motivate interest in determining theoretical probabilities to explain the data display characteristics they observed. In particular, we aimed to have them think about why the "one of each color" category they suggested (Figure 7) tended to contain more observations than the other two categories to help motivate a more careful focus on mathematical structure and sample space.

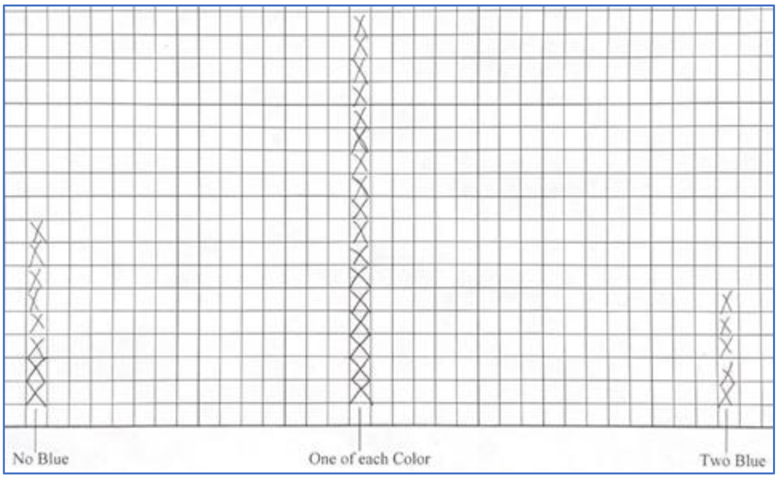


Figure 7. Student-generated line plot to represent the number of blue cubes obtained during random draws of pairs of cubes from a bag containing an equal number of red and blue cubes.

In accord with previous research (Friel et al., 2001), the graph comprehension activity of reading the data caused the least difficulty for students. Students were generally able to map their experiences with concrete play to *TinkerPlots*™—simulated play and read values from representations of results they generated through concrete and simulated play. For example, when questioned, students were readily able to say what each of the marks shown in graphs such as the one in Figure 7 represented. One exception to this trend occurred after students first read a line plot containing the results of a *TinkerPlots*™ simulation of drawing pairs of cubes with replacement. The class had tracked the results of the simulation using a line plot with three stacks: one to show how many times no blues were drawn, a second to show how many times one of each color was drawn, and the third to show how often two blues were drawn (as in Figure 7). Laura and Tom temporarily lost track of what the marks in the line plot represented, but such problems reading the data did not persist. Reading between the data, which mainly involved comparing the heights of stacks in the line plot to one another, also did not cause any observable difficulties; as we questioned students about the most frequent outcome, they consistently identified it.

At the outset of the third lesson, students did *TinkerPlots*™ simulations on their own computers, graphed the data in line plots like the one shown in Figure 7, and then compared results. Tom, Aiden, and Laura produced graphs in which the middle stack was the highest. This led them to say that drawing one of each color would be more likely than getting two reds or two blues. Aiden, for example, remarked, "You have a higher chance of getting one of each than no blue," and "Getting two reds is less likely but getting one of each color is more likely." Likewise, Tom noticed, "There's five red and five blue, that - and no blue is tough because it's not all blue - but, um, there's, like, five red and five blue, it's easy to, like, get, because there's the equal amount of red and blue." As discussion shifted to Emilia's data, difficulties reading behind the data emerged. Emilia's graph differed from the other three in that all three stacks were of equal height. Our observations did not allow us to determine if this was simply because of variation or because of an error she made in generating or recording the data. Her graph showed the same number of trials as the others. In any case, Emilia's data prompted the following exchange with Laura:

Laura:	Actually, can I change my mind from, um, well I either think each color could get the most or, um—
Teacher:	You think which one?
Laura:	I think 1 of each color could get the most or like you never know because you might—because we thought 1 of each color would get the most—
Teacher:	Yeah?
Laura:	-- but hers [Emilia's] was that each of them were equal so you never know which is—
Teacher:	We never know which one is going to come out?
Laura:	Um-hmm.

Emilia and Aiden then expressed thoughts similar to Laura's. Emilia said, "So, it could be, um, like, the same amount for all three of them, so you never really know." Similarly, Aiden remarked, "You never know if it's gonna be blue-blue or blue-red." Emilia's data temporarily complicated students' discernment of underlying probabilities in the game situation.

In response to the discussion caused by Emilia's graph, we asked students to temporarily set aside the empirical data from game play. In an attempt to draw their attention to the underlying theoretical probabilities, we posed the writing prompt shown in Figure 8. It used the familiar cube context and an empty table to encourage students to think about sample space. We asked students to individually write about why the game was fair or unfair. This prompted Laura to move away from focusing solely on Emilia's data in favor of considering probabilities of specific outcomes. Her work and reasoning are shown in Figure 9. Laura's work suggested that she had begun to think about the ideas of sample space composition and probabilities of outcomes in tandem, indicating a budding connection she did not exhibit during pre-interviews. The other three students in the group did not show evidence of making the same connection, even though they each completed the chart shown in Figure 8 correctly. Tom, Emilia, and Aiden all decided that the game was fair after completing the chart. Tom wrote that the game was fair because "There are 2 red and 2 blue and there is a 50/50 chance of a win." Emilia wrote that it was fair because "You never know what you're going to draw." We conjectured that such reasoning persisted because of the earlier discussion of Emilia's data. To help students move beyond Emilia's data and attend to underlying mathematical structure, we decided to introduce a structurally similar game during the next lesson, but to start by drawing students' attention to its theoretical rather than empirical probabilities.

You are playing a carnival game. You win a prize if you pick two blue cubes. The carnival worker tells you this is a fair game. Use the table below to see if the worker is telling the truth.

	Cube 1		
Cube 2		Red	Blue
	Red		
	Blue		

Figure 8. Empty table to prompt students to attend to sample space in the cube-drawing scenario.

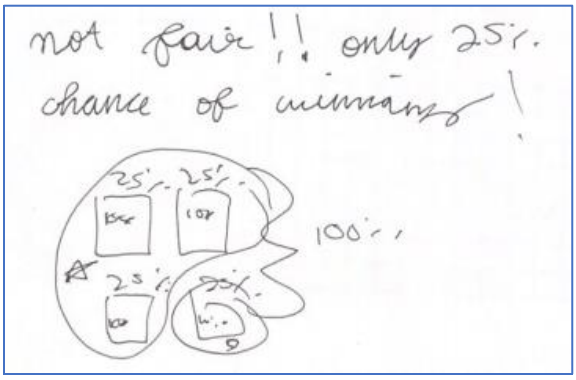


Figure 9. Laura's work on the two-cube carnival game task chart shown in Figure 8.

5.3. INSTRUCTIONAL SCENARIO 2: FLIPPING A PENNY AND QUARTER SIMULTANEOUSLY

We used the context of flipping a penny and quarter simultaneously for the second instructional scenario. We hypothesized that beginning with a game structurally similar to the one at the end of the first instructional scenario but asking students to explore its sample space before gathering data from trials, would re-focus attention on the range of possible outcomes and not just empirical results. We began with a penny and a quarter rather than two of the same type of coin to simplify initial conversations about order in determining sample spaces; such conversations can be particularly challenging to students initially learning about compound probability (Alston & Maher, 2003; English & Watson, 2016; Iversen & Nilsson, 2019).

At the outset of the scenario, students were told that a penny and quarter were flipped at the same time. Player A would receive a point if both coins landed on heads or both coins landed on tails. Player B would receive a point otherwise. Students were asked to write, individually, if they thought the game was fair and to explain their thinking. Tom wrote that it was fair because there were four coin sides and two players, so there would be two chances for each person to win. Aiden believed it to be fair because he assumed each player would get the same number of turns. Laura and Emilia both believed the game to be fair because each player would have the same chance of winning, but they did not systematically analyze the probabilities. Notably, Laura did not bring her budding connection about the relationship between sample space and probability to bear at the outset of the scenario.

In order to help students transition to systematic analyses that would help generate connections between sample space and probability, we asked them to individually draw pictures of the different ways each player could win a point in the penny-quarter game. Laura, Tom, and Emilia each drew all four possible outcomes. Aiden accounted for only two outcomes: heads on each coin and heads on the penny and tails on the quarter. When students discussed their responses to the writing prompt as a group, Tom began to consider heads on the quarter and tails on the penny to be the same outcome as tails on the quarter and heads on the penny. As shown in Figure 10, he crossed out the "tails on quarter, heads on penny" outcome he initially drew after deciding it was the same as "tails on penny, heads on quarter." Others disagreed with him, and Tom vacillated between considering the two outcomes to be the same or different. When asked to respond to Tom's opinion that the outcomes were the same, Laura said, "Technically it's not the same but if it's like the outcome of the points it's the same." Laura's remark helped emphasize the importance of specifying what one means when calling outcomes, the "same."

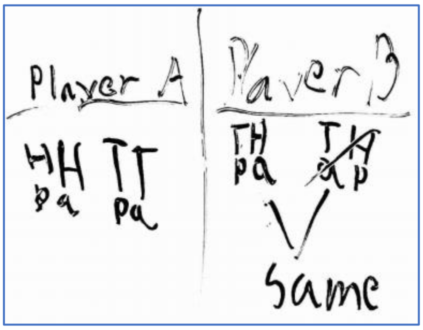


Figure 10. Tom's depiction of outcomes for the penny-quarter game at the outset of instructional Scenario 2.

To help students further discern the sample space for the penny-quarter game and ground their conversations more firmly in the game context, we had them play the game in pairs, record the results, and then once again draw pictures of the different ways each player could win a point. After playing the game, all the students individually drew all four possible outcomes, as we hoped. Unfortunately, playing the game also had some unintended effects. Tom, in particular, fixated on why he had scored only 8 points while his partner, Laura, had scored 12. He initially focused on coin weights as an explanation, just as he had focused on physical considerations related to games in the previous scenarios. He said, "The weight is the problem; this one—this one goes less, like spins less, because of the weight, and this one spins more because of the weight." Later, he returned to the idea that heads on the quarter and tails on the penny was the same outcome as tails on the quarter and heads on the penny, which led him to believe there was a problem with the rules of the game as well. Tom summarized by saying, "The problem is the weight and the rules." Laura disagreed with him, saying, "The weight didn't really matter." Tom also, at least temporarily, considered the initial positions of the coins when flipped to be important, stating,

> *If it's like this [putting quarter on his left and penny on his right] you could get a—it's an easier chance to get a tails—but, if you do it like this [putting penny on his left and quarter on his right] it will be an easier chance to get heads.*

Strength of flip was another consideration for Tom, as he remarked, "If it is (flipped) too fast, you don't know what's gonna happen … but if it's too slow, if you like do it, like really weak, it's too slow." Physical considerations again temporarily drew his attention away from considering underlying mathematical structures.

Because of the fixation on physical attributes of the two coins rather than sample space, we posed a slightly different penny-quarter task at the beginning of the next lesson. We told students to imagine a penny and quarter that were both perfectly balanced, and that the coin flipping was not "rigged" in any way (using language Tom had introduced earlier). Player A would get a point when both coins landed on tails, and Player B would get a point otherwise. Students individually wrote about whether or not the game was fair. All students but Emilia (whose work is shown in Figure 11) listed all four possible outcomes and assigned one of them to Player A and the other three to Player B. After students shared their thinking with one another, Emilia revised her response to count heads on the quarter and tails on the penny as a separate outcome from tails on the quarter and heads on the penny. All four students then assigned a one-fourth or 25% probability to each outcome when asked to individually write the chance of each one. When asked to predict how many occurrences of tails on the quarter and tails on the penny there would be in 100 trials, each student individually wrote 25. We took these events to indicate that students had begun to recognize the need to determine sample spaces and assign probabilities to outcomes to analyze games.

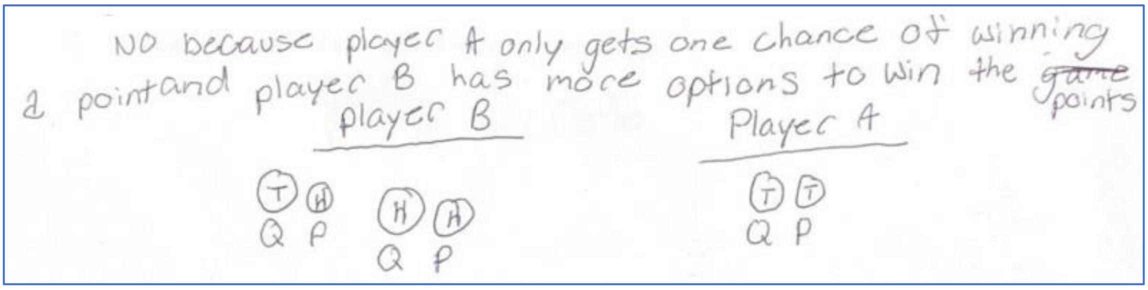


Figure 11. Emilia's initial work and reasoning on the revised penny-quarter task.

5.4. INSTRUCTIONAL SCENARIO 3: FLIPPING TWO QUARTERS SIMULTANEOUSLY

Given that students had begun to determine sample spaces and use them to assign probabilities during the penny-quarter scenario, we slightly altered the context for tasks during scenario 3 to introduce a more challenging situation. Specifically, because students had begun to attend to order with two distinct coins, we attempted to extend their reasoning by introducing two identical coins.

During scenario 3, students mainly continued to reason about sample space as they did near the end of scenario 2. To begin scenario 3, we changed just one part of the final task from scenario 2, introducing the situation of flipping two quarters simultaneously rather than a penny and a quarter. All other elements of the task remained the same, including Player A receiving a point for two tails and Player B receiving a point otherwise. When asked to write individually if they considered the game to be fair, all four students started again by listing possible outcomes. Aiden, Tom, and Laura wrote all four possible outcomes on their own at the start. Emilia wrote three of the four, not counting tails on quarter 1 and heads on quarter 2 as separate from heads on quarter 1 and tails on quarter 2. Laura briefly adopted Emilia's viewpoint as students discussed their written solutions during class before returning to her original solution later on. Ultimately, during class discussion, all students claimed that the sample space contained four outcomes.

At the conclusion of the quarter-quarter instructional scenario, we told students they had a choice of playing either the quarter-penny or quarter-quarter game at a carnival. In both cases, they would win by flipping two tails and lose otherwise. We had them respond individually, in writing, to the question of which carnival game provided a better chance of winning. All four students wrote that there was the same chance of winning each game, and they maintained the same opinions during class discussion of the written responses. Aiden and Tom explicitly quantified the chance of winning as 25%. When we asked students to write how many occurrences of tails-tails they would expect in 200 flips, all of them predicted 50. Given that the reasoning we observed closely paralleled that of the previous instructional scenario, we inferred that students were able to employ several of the connections about sample space and probability they had developed in the slightly different penny-quarter situation.

5.5. INSTRUCTIONAL SCENARIO 4: ROCK, PAPER, SCISSORS

During the final instructional scenario, we aimed to help students extend their reasoning beyond contexts involving two coins and four sample space elements. We conjectured that students were ready for new contextual and structural elements because of the connections they had generated about sample space and probabilities during the previous scenarios. Given the nature of students' work at the conclusion of the previous scenario, we also thought there may be an opportunity to introduce an additional representation conventionally used to make sense of compound probability situations. Specifically, we began the new scenario by helping students represent their thinking about previously encountered two-coin sample spaces with tree diagrams (Konold, 1996). Then, we encouraged the use of tree diagrams to map the sample space for a new context, the game of rock, paper, scissors (Nelson & Williams, 2008), which contained more sample space elements than the two-coin scenarios. Our beginning thought was that tree diagrams would be useful tools for students to use in organizing and counting outcomes as they encountered larger numbers of sample space elements in different tasks.

We began the final scenario by reminding students of the sample space for the quarter-quarter scenario and put an organized list on the board to represent the possible outcomes. Tom was able to explain what the organized list represented when asked, saying, "There's four possibilities and Player A got one possibility to win and Player B got 3 ways to win." Emilia explained that the probability of each outcome on the list was 25%, saying, "You have one option or one outcome out of four or one option out of four." Having elicited students' knowledge of the organized list for the quarter-quarter scenario, we introduced a tree diagram as another representation of the same sample space. We asked students to state the possible outcomes for the first quarter and then drew the corresponding parts of the tree diagram for them on the board. We did the same for the second quarter to complete the tree diagram. When we asked students where outcomes such as "heads on the first, tails on the second" were located in the diagram, all students but Aiden were able to circle the appropriate portions, though he was eventually able to do so with help from others.

Having introduced the tree diagram representation, we next aimed to show how tree diagrams can be used to map sample spaces for more complex situations, such as the game of rock, paper, scissors (Nelson & Williams, 2008). Students collectively created an organized list to completely map the rock, paper, scissors sample space after playing the game in pairs. Then, as with the quarter-quarter sample space, they mapped the rock, paper, scissors sample space with a tree diagram (e.g., Figure 12). When we asked individual students to explain how different outcomes were represented in the tree diagram, Aiden was again the only one of the three students to struggle. Initially, he did not see how the tree diagram represented the outcome of "rock for Player A, paper for Player B" differently from "paper for Player A, rock for Player B." Later in the lesson, however, Aiden was able to identify such outcomes in the tree diagram without assistance.

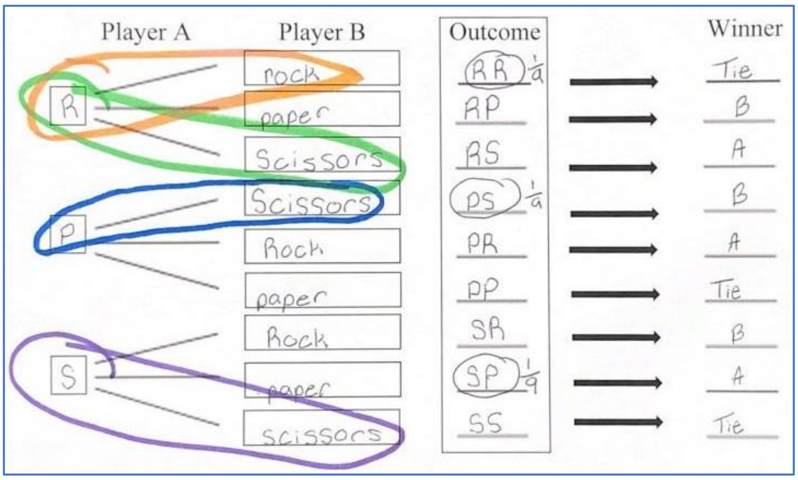


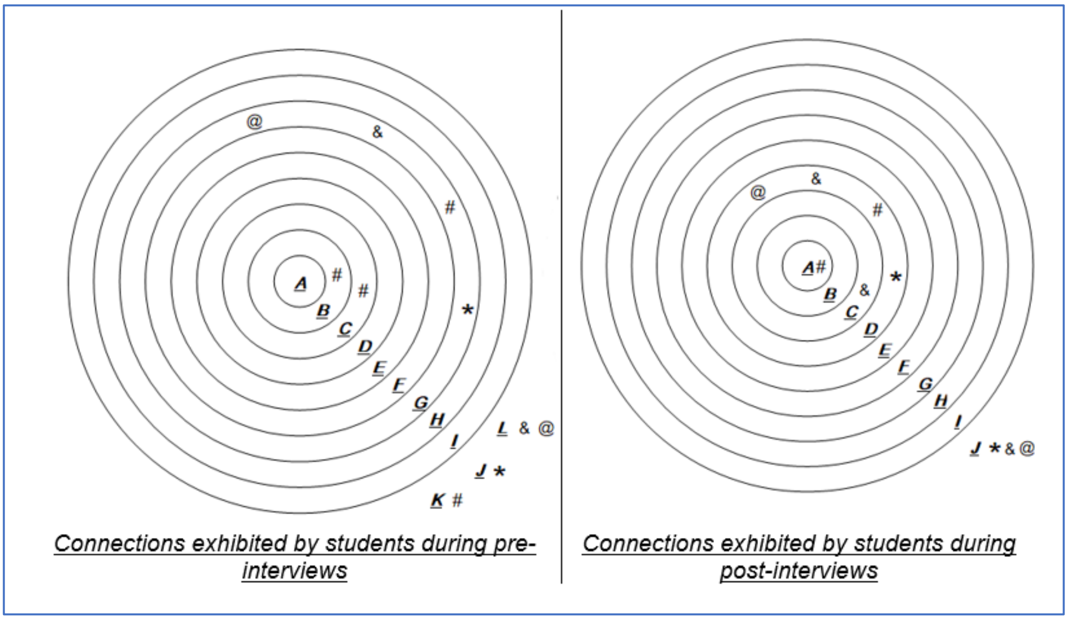
Figure 12. Emilia's tree diagram for the rock, paper, scissors scenario.

To conclude the final lesson, we asked students to use the tree diagram to calculate the probabilities of outcomes for rock, paper, scissors and determine if it was a fair game. All students calculated the probability of "paper for Player A, scissors for Player B" to be one-ninth when asked to do so individually in writing. Emilia explained that it was one-ninth "because you have one thing that we picked out of nine things" (the "nine things" being the nine outcomes listed on the tree diagram). All four students also expressed the opinion that rock, paper, scissors was a fair game. Laura explained, "We proved that it was fair because it has three ties; there's two players, a player can win three times and the second player can play three times, and there's most likely going to be three ties, so it is fair." The focus on possible outcomes and their associated probabilities differed from students' earlier conceptions of fairness that were limited strictly to accounting for the number of turns taken by each player.

5.6. POST-INTERVIEWS

Individual post-interviews were held with each student a week after the final lesson. As noted earlier, we administered the two-spinner task (Figure 2) before the two-coin task (Figure 1) to avoid implicitly suggesting that the strategy for the two-spinner task should be the same as that used for the two-coin task. We purposefully avoided spinners during instruction so we could observe the extent to which connections generated in other contexts might help them reason about the two-spinner situation.

Figure 13 contrasts the connections students exhibited during pre-interviews and post-interviews. During post-interviews, Tom and Emilia moved away from reasoning about non-mathematical features of tasks (connection L in Figure 6A). Instead, in response to the two-spinner post-interview task (Figure 2), Aiden, Tom, and Emilia all used the strategy of reasoning there was a 50-50 chance because half of the total area in the task picture was white and the other half was black. Although this was not a normative strategy, for Tom and Emilia it represented a shift toward considering a mathematically relevant aspect of the task, namely, area, rather than only physical considerations such as strength of spin. This shift in strategy occurred even though none of the instructional scenarios during the study involved spinners or area.



Note: Symbols are used to represent the four participants: Aiden (*), Laura (#), Emilia (&), and Tom (@). Underlined, italicized letters correspond to the types of connections listed in Figure 6A.

Figure 13. Comparison of connections students used during pre- and post-interviews.

During post-interviews, all four students moved beyond saying that fairness of games depended solely on the number of turns per player (connection G in Figures 6A and 13). Instead, they linked the ideas of sample space and probability (connection D in Figures 6A and 13) when given the two-coin task (Figure 1). All four began by listing the sample space for the task, as they did during instruction. Aiden, for example, explained the two-coin game was not fair by saying, "Because you only get a win if both coins are heads and heads, and there's more outcomes than just heads and heads; there's heads-tails, tails-heads or tails-tails." Emilia produced the work shown in Figure 14 and explained her diagram by saying, "Heads-tails and tails-tails, and then we said in class that tails and heads aren't the same thing so tails-heads. And then you want to have heads-heads. So, then you have ... you want this one (circles heads-heads) so that would be one-fourth which equals 25% chance of winning." When asked to explain why she considered heads-tails and tails-heads to be different outcomes, she drew upon experiences from class playing rock, paper, scissors. Emilia said that if the interviewer had scissors and she had paper, the interviewer would win; but, if Emilia had scissors and the interviewer had paper, Emilia would win. She summarized by saying, "Even though it is the same hand motions or whatever you want to call it, different people win, so it is a different outcome." Although the others did not relate the two-coin task to rock, paper, scissors, students' post-interview responses demonstrated that they were able to use strategies developed in class for the two-coin task when presented the situation again during post-interviews.

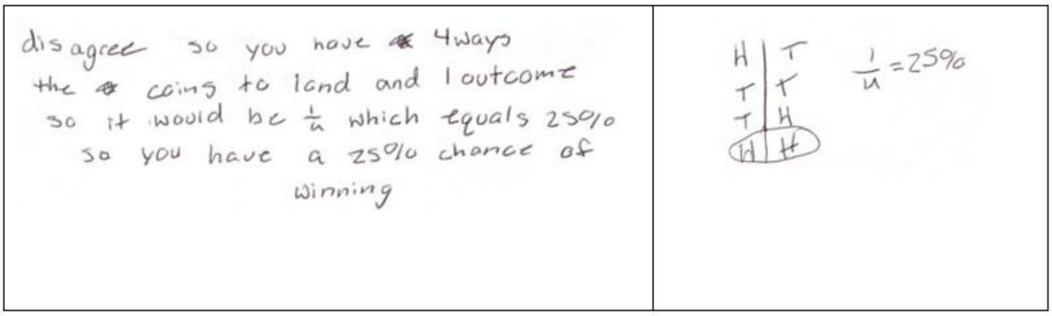


Figure 14. Emilia's work on the two-coin post-interview task.

Laura also demonstrated evidence of making additional connections between the two-coin and two-spinner tasks during post-interviews. During pre-interviews, she explicitly noted similarities between the two tasks but did not give a complete analysis of the chance of winning each game. During post-interviews, When Laura was asked if there was a 50-50 chance of winning the two-spinner game (Figure 2), the following exchange occurred:

Laura: I would say no.
Intv: Why do you say no?
Laura: Because if both of them land on black, it means that they win. And if both of them and on white, it means they lose. But if one of them lands on white and one of them lands on black, it's lose. So, it's two to one.
Intv: So how many total outcomes do you think that there are?
Laura: Four.
Intv: Four?
Laura: Mm-hmm (affirmative).
Intv: And what are those outcomes?
Laura: Black and black. White and white. Black and white. And white and black.

To summarize her strategy, Laura said, "One chance of winning and three chances of losing." Laura's strategy for the task differed from her pre-interview strategy of only partially listing the sample space and not using the sample space to judge the probability of winning.

5.7. OVERARCHING CONNECTIONS

Along with the individual student-generated connections we have reported on so far, some larger overarching connections can be discerned. Figure 15 depicts three key tasks in the study and dimensions of context, variation, mathematical structure, sample space, and theoretical probability associated with each one (D1-D5). Expertise requires the ability to attend to and coordinate all of these dimensions. The two-coin game, the two-cube game, and the two-spinner game were analogous along dimensions pertaining to mathematical structure (D3 and D4), but their context features differed (D1). Observed variation from trials (D2) at times did not match probabilistic expectations (D5). Students' varying abilities to connect and coordinate these dimensions to make sense of situations provide additional insight on the nature of their developing expertise.

	Two-coin game	**Two-cube game**	**Two-spinner game**
D1	Salient context features: weight and balancing	Salient context features: blind or non-blind draws	Salient context features: strength of spin
D2	Variation can be observed and recorded as trials are conducted		
D3	Mathematical structure used as a means to explain empirical outcomes		
D4	Lose: HT, TH, HH Win: TT	Lose: BR, RB, RR Win: BB	Lose: BW, WB, WW Win: BB
	Mathematical structures of sample spaces are determined		
D5	Probabilities associated with isomorphic sample spaces are determined		

Figure 15. Dimensions of reasoning about key tasks for the present study.

Students' responses to the three key tasks summarized in Figure 15 illustrated the influence of students' reasoning about task context in stochastic situations (Chow & Van Haneghan, 2016; Makar & Ben-Zvi, 2011). Tom's data are particularly instructive in this regard. He consistently sought to explain observed outcomes by thinking about the mechanisms used to produce them. This was true as he reasoned about all three games depicted in the columns of Figure 15. One might say he had constructed a meta-connection that irregularities in random data generation mechanisms can explain observed outcomes. This type of thinking is important, but frequently undervalued, in school curricula, which tend to include only tasks based on the assumption that such mechanisms are never flawed (Watson & Moritz, 2003). Tom's classmates were at times dismissive of his concerns along this dimension (D1 in Figure 15). Laura, for example, at one point told him that the weight of the coins being flipped did not matter. Laura's remark might actually indicate development of an "expert blind spot." Although mathematical structure and theoretical probability are important (D3, D4, and D5 in Figure 15), focusing only on those dimensions can lead to invalid conclusions when, for example, it turns out that dice or coins are not fairly weighted. In practice, most such tools used for generating data do have manufacturing flaws preventing them from being perfectly balanced. Contextual reasoning (D1 in Figure 15) has a prominent role in explaining observed phenomena in such cases (Langrall et al., 2011).

Although attention to context is important, we found that students at times needed to be prompted to focus on other dimensions as well. Tom's focus on the context dimension at times seemed over-powering, blocking him from considering the other dimensions shown in the rows of Figure 15. We used the strategy of posing idealized situations (e.g., perfectly balanced and flipped coins, simulated cube drawing with no possibility of cheating), to prompt him and his classmates to make connections among dimensions of mathematical structure (D3), sample space (D4), and probability (D5). This was not done to discourage attention to the context dimension, but to direct students' attention to other relevant aspects of situations. Similarly, we found a need to temporarily direct attention away from the dimension of variability (D2 in Figure 15) when Emilia's unusual set of results led Laura and the others to doubt their analyses of the underlying mathematical structure of the cube-drawing scenario. In that case, we shifted to a structurally similar task that was not laden with images of Emilia's unusual data set. This type of teaching move, which shifts attention to other dimensions of a situation when students become caught up in just one dimension, appears to be particularly important to support students in generating normative connections.

6. DISCUSSION AND CONCLUSION

It is important to acknowledge the core limitations of our study as we turn to a discussion of its results. As a detailed examination of a single group of students, the present study was not designed to be statistically generalizable. We cannot make claims about how prevalent the student thinking patterns we observed may be in the larger population. We also cannot draw conclusions about how effective our teaching strategies would be with different groups of students or if used by other teachers working in various instructional contexts. Nonetheless, our findings do have some thought-provoking connections to previous research and theory that are productive to consider. In closing, we reflect on: (i) a possible structure for classroom discourse among students who exhibit attention to various dimensions of a contextualized task, (ii) why it may be best to avoid analyzing student thinking through the lens of negative transfer, and (iii) the non-linear progression of learning some students have when working with two-stage compound probability problems like those used in our study.

Given the multi-dimensional nature of expertise required for reasoning about probabilistic situations, and students' sometimes uneven attention to each dimension, learning in this domain seems better characterized as visiting and re-visiting dimensions relevant to tasks rather than a simple linear progression toward deeper mathematical abstraction. Students need time and space to consider each dimension, and teachers should be alert to the need to allow students to reason along each dimension without becoming mired exclusively in it. For example, teachers using the five practices model (Stein et al., 2008) could select and sequence student strategies to include in class discussion according to the dimensions they reflect. Students like Tom can benefit from having students like Laura prompt them to consider mathematical structure, and students like Laura can benefit from having their attention drawn to contextual considerations by students like Tom. Mathematical and contextual considerations exist in tandem in robust expert analyses of situations like those summarized in Figure 15; classroom discourse should reflect this by valuing both types of contributions without letting contributions from a single dimension dominate. We acknowledge that carrying out such teaching moves is a steep challenge for classroom teachers. Complementary future studies are needed in order to provide additional insight into the types of professional development that would help teachers meet the challenge.

Our findings also suggest a need for researchers to problematize the construct of negative transfer (Chen & Daehler, 1989). Negative transfer suggests overgeneralization of previously learned ideas, but other forces might be at play. For example, during post-interviews, on the two-spinner task (Figure 2), Tom and Emilia used the commonly observed strategy of assigning a probability solely based on the total spinner area shaded (Pratt, 2000; Shaughnessy & Ciancetta, 2002). This approach has been characterized by some as the over-generalization of a strategy that works for simple probability tasks (Iversen & Nilsson, 2019). We did not, however, use spinner tasks during instruction, and Tom and Emilia did not use area when reasoning about spinners during pre-interviews, so it does not seem plausible that they generalized one-spinner strategies to the two-spinner situation. Instead, the most notable aspect of Tom and Emilia's thinking was a shift away from only physical considerations such as strength of spin during pre-interviews. Characterizing their post-interview thinking as overgeneralization or negative transfer suggests a need to extinguish a misconception; characterizing it as the beginning of reasoning about mathematical structure emphasizes that they were capable of reasoning about an abstract task dimension, area, even if their strategies needed further development. They may also have begun to see spinners as stochastic devices, as they no longer spoke of their personal actions as determining the outcomes (Pratt, 2000). This would suggest a budding connection between the notion of a random outcome and a spin rather than just misuse of previously learned strategies. Focusing on the detection of such small-scale connections using an AOT perspective holds more potential to inform instruction than simply characterizing such responses as indicative of negative transfer.

Salient findings from the present study about learning and teaching compound probability can be summarized using an extended metaphor about travel. In regard to learning, students generated connections on the journey toward expertise as they traversed within and among several dimensions (D1–D5) relevant to the content. Although all of these dimensions are important to explore (English & Watson, 2016), students' journey toward expertise was not well-characterized as a simple linear progression from dimensions related to contextual intuitions and experimental probabilities to those involving formal probabilities. Rather, the journey was most profitable when they visited and revisited dimensions of context, variation, and mathematical structure rather than lingering too long in any given one. Instructors served as tour guides familiar with the disciplinary landscape who would direct students to traverse among all relevant dimensions. At times, this required beginning an instructional scenario journey with abstract mathematical structure rather than the more concrete aspects of context and data generated from trials. Traversing to another dimension, such as mathematical structure, did not mean permanently leaving other dimensions. Rather, fluidly moving among all dimensions relevant to a problem situation was the ultimate goal. Continuously revisiting all dimensions of relevance helps students generate well-traveled connecting pathways that provide ready access to all dimensions they need to visit to understand a given compound probability problem situation.

ACKNOWLEDGEMENTS

This material is based on work supported by the National Science Foundation Grant DUE-1658968. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Alston, A., & Maher, C. A. (2003). Modeling outcomes from probability tasks: Sixth graders reasoning together. In N. A. Pateman, B. J. Dougherty, & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education, Honolulu, HI* (Vol. 2, pp. 25–32).

Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Doing qualitative research: Methodology and methods in mathematics education* (pp. 429–466). Springer.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school*: *Challenges for teaching and learning* (pp. 15–37). Springer. https://doi.org/10.1007/0-387-24530-8_2

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61–100. https://doi.org/10.3102/0091732X024001061

Chen, Z., & Daehler, M. W. (1989). Positive and negative transfer in analogical problem solving by 6-year-old children. *Cognitive Development, 4*(4), 327–344. https://doi.org/10.1016/S0885-2014(89)90031-2

Chow, A. F., & Van Haneghan, J. P. (2016). Transfer of solutions to conditional probability problems: Effects of example problem format, solution format, and problem context. *Educational Studies in Mathematics, 93*, 67–85. https://doi.org/10.1007/s10649-016-9691-x

Cobb, P., Jackson, K., & Sharpe, C. (2017). Conducting design studies to investigate and support mathematics students' and teachers' learning. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 208–233). National Council of Teachers of Mathematics.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. http://www.corestandards.org/

Corbin, J. & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). SAGE Publications.

DeCuir-Gunby, J. T., Marshall, P. L., & McColloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development project. *Field Methods*, *23*(2), 136–155. https://doi.org/10.1177/1525822X10388468

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

Fischbein, E., & Snarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105. https://doi.org/10.2307/749665

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124–158. https://doi.org/10.2307/749671

Halcyon Software. (2006). *Inspiration software* (Version 8) [Computer Software].

Iversen, K., & Nilsson, P. (2019). Lower secondary school students' reasoning about compound probability in spinner tasks. *Journal of Mathematical Behavior, 56*, Article 100723. https://doi.org/10.1016/j.jmathb.2019.100723

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory* (pp. 109-122). Oxford University Press.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education, 3*(1). https://doi.org/10.1080/10691898.1995.11910479

Konold, C. (1996). Representing probabilities with pipe diagrams. *Mathematics Teacher, 89*(5), 378–382. https://www.jstor.org/stable/27969794

Konold, C., & Miller, C. (2011). *TinkerPlots* (Version 2.3 [Computer Software]. Learn Troop.

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

Lineback, J. E. (2015). The redirection: An indicator of how teachers respond to student thinking. *Journal of the Learning Sciences, 24*(3), 419–460. https://doi.org/10.1080/10508406.2014.930707

Lobato, J. (2003). How design experiments can inform a rethinking of transfer and vice versa. *Educational Researcher*, *32*(1), 17–20. https://www.jstor.org/stable/3699930

Lobato, J. (2012). The actor-oriented transfer perspective and its contributions to educational research and practice. *Educational Psychologist*, *47*(3), 232–247. https://doi.org/10.1080/00461520.2012.693353

Lobato, J., & Siebert, D. (2002). Quantitative reasoning in a reconceived view of transfer. *Journal of Mathematical Behavior*, *21*(1), 87–116. https://doi.org/10.1016/S0732-3123(02)00105-0

Lockwood, E. (2011). Student connections among counting problems: An exploration using actor-oriented transfer. *Educational Studies in Mathematics*, *78*(3), 307–322. https://doi.org/10.1007/s10649-011-9320-7

Lysoe, K. O. (2008). Strengths and limitations of informal conceptions in introductory probability courses for future lower secondary teachers. In M. Borovcnik, D. Pratt, Y. Wu, & C. Batanero (Eds.), *Research and development in the teaching and learning of probability* (pp. 1–14). International Association for Statistical Education.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1&2), 1–4. https://doi.org/10.1080/10986065.2011.538291

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.

Moyer, P. S., & Milewicz, E. (2002). Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education, 5*, 293–315. https://doi.org/10.1023/A:1021251912775

Nelson, C., & Williams, N. (2008). A fair game? The case of rock, paper, scissors. *Mathematics Teaching in the Middle School, 14*(5), 311-314. https://www.jstor.org/stable/41183143

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*, 413–448. https://doi.org/10.3102/00346543076003413

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602–625. https://doi.org/10.2307/749889

Ron, G., Dreyfus, T., & Hershkowitz, R. (2017). Looking back to the roots of partially correct constructs: The case of the area model in probability. *Journal of Mathematical Behavior, 45*, 15–34. https://doi.org/10.1016/j.jmathb.2016.10.004

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age Publishing.

Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics, Cape Town.* https://iase-web.org/documents/papers/icots6/6a6_shau.pdf

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340. https://doi.org/10.1080/10986060802229675

Vidakovic, D., Berenson, S., & Brandsma, J. (1998). Children's intuition of probabilistic concepts emerging from fair play. In L. Pereira-Mendoza, T. Wee Kee, W.-K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics, Vol. 1, Singapore* (pp. 67–73). International Statistical Institute.

Wagner, J. F. (2010). A transfer-in-pieces consideration of the perception of structure in the transfer of learning. *The Journal of the Learning Sciences, 19*(4), 443–479. https://doi.org/10.1080/10508406.2010.505138

Watson, J. M., & Kelly, B. (2004). Expectation versus variation: Students' decision-making in a chance environment. *Canadian Journal of Science, Mathematics, and Technology Education, 4*(3), 371–396. https://doi.org/10.1080/14926150409556620

Watson, J. M., & Moritz, J. B. (2003). Fairness of dice: A longitudinal study of students' beliefs and strategies for making judgments. *Journal for Research in Mathematics Education, 34*(4), 270–304. https://doi.org/10.2307/30034785

Webb, P., Whitlow, J. W., Jr., & Venter, D. (2017). From exploratory talk to abstract reasoning: A case for far transfer? *Educational Psychology Review, 29*, 565–581. https://doi.org/10.1007/s10648-016-9369-z

Wheeldon, J., & Åhlberg, M. K. (2011). *Visualizing social science research: Maps, methods, and meaning*. SAGE Publications.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 235–268). National Council of Teachers of Mathematics.

RANDALL E. GROTH
Salisbury University
1101 Camden Ave.
Salisbury, MD 21801
USA