Toward a theoretical structure to characterize early probabilistic thinking

Randall E. Groth, Jathan W. Austin, Madeline Naumann & Megan Rickards

Mathematics Education Research Journal

ISSN 1033-2170
Volume 33
Number 2

Math Ed Res J (2021) 33:241-261
DOI 10.1007/s13394-019-00287-w

Springer

Mathematics Education Research Journal (2021) 33:241–261
https://doi.org/10.1007/s13394-019-00287-w

ORIGINAL ARTICLE

Toward a theoretical structure to characterize early probabilistic thinking

Randall E. Groth[1] · Jathan W. Austin[2] · Madeline Naumann[1] · Megan Rickards[1]

Received: 6 January 2019 / Revised: 11 August 2019 / Accepted: 16 August 2019 /
Published online: 27 August 2019

Abstract

The role of probability in curricula for children has fluctuated greatly over the past several decades. Recently, some countries have removed probability from their preschool and primary curricula, and others have retained it. One reason for such lack of agreement is that theory about early probability learning is still relatively new and under development. The purpose of this report is to sketch a tentative theoretical structure with the potential to anchor curricular decisions and inform further research on early probability learning. Toward that end, we begin by reviewing existing literature. We attend, in particular, to the probabilistic thinking tendencies exhibited by children whom researchers consider to be in the earliest stages of learning the subject. We then use these tendencies to posit two different cycles for early probabilistic thinking. One of these cycles is compatible with disciplinary norms and supports normative thinking; thinking tendencies in this normative compatible cycle include attending to the position of objects in a container, forming images of random generators, attending to the operation of random generators, and thinking about past experiences playing games of chance. Thinking tendencies in the other posited cycle lead to belief systems that conflict with normative disciplinary practice; these belief systems include elements such as myths, superstitions, animism, and determinism. We illustrate the two cycles using empirical data from design-based research. We then reflect on how the two cycles comprising our structure of early probabilistic thinking (SEPT) framework can provide a basis for further curricular and theoretical work.

Keywords Probability · SOLO taxonomy · Cognition · Beliefs

✉ Randall E. Groth
regroth@salisbury.edu

[1] Seidel School of Education, Salisbury University, 1101 Camden Ave., Salisbury, MD 21801, USA

[2] Department of Mathematics and Computer Science, Salisbury University, 1101 Camden Ave., Salisbury, MD 21810, USA

The place of probability within school curricula has been in flux over the past several decades. Piaget's observations about young children's performance on probability tasks contributed to the relegation of probability to the later grades during the mid-twentieth century (Jones and Thornton 2005). Later in the century, Fischbein's (1975) research demonstrated how primary school students could develop sound intuitions about probability, providing momentum for further research and curriculum development. This momentum continued through the conclusion of the twentieth century and the beginning of the twenty-first in the form of research programs (e.g., Jones et al. 1999a; Watson et al. 1997) and standards development efforts (e.g., Australian Education Council 1991; DfEE 1999; NCTM 1989; 2000) that established probability as a subject of study in primary school. However, in spite of the research, some countries have once again begun de-emphasizing or eliminating the study of probability in the early years of school (Langrall 2018).

The growing divergence between research and curriculum development in early probability learning has potentially negative consequences. Delaying the study of probability until later in school may allow errant early intuitions about probability to become more firmly established (Martignon 2014), making the removal of probability from the early years problematic (Greer 2014). Moreover, probabilistic reasoning is becoming more essential in society for tasks that involve dealing with uncertainty and making sense of data (Ben-Zvi 2018). Research shows that as early as the preschool level, children can begin to make sense of many aspects of probability (Nikiforidou 2018). Hence, this report begins from the premise that the current question for curriculum developers is not *if* probability should be studied by young children; rather, the question is *how* it ought to be studied. Addressing such a question requires a strong theoretical base; currently, theory concerning the earliest stages of learning probability is still in development (Leavy et al. 2018).

Purpose and overview

The purpose of this report is to sketch a tentative theoretical structure for informing curricular decisions about the appropriate nature of the earliest probability learning experiences for children. Toward that end, we begin by discussing the Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs and Collis 1991), which is the theoretical basis for the most influential existing frameworks that describe primary students' probabilistic thinking (Mooney et al. 2014). We then examine the literature on early probability learning in light of SOLO to suggest a structure of early probabilistic thinking (SEPT) framework. We illustrate the features of the SEPT framework and its potential utility by applying it to empirical data. Finally, we reflect on the implications the SEPT framework holds for curriculum development and future research.

The SOLO taxonomy and the ikonic mode

The SOLO taxonomy is a Neo-Piagetian framework postulating 5 cognitive modes: sensorimotor, ikonic, concrete-symbolic, formal, and post-formal (Biggs and Collis 1991). The five SOLO modes are believed to accumulate rather than replace one another as a child ages. Previous research on young children's thinking has found evidence of both ikonic and concrete-symbolic mode thinking in response to probability tasks (e.g., Jones et al. 1997; Tarr and Jones et al. 1997; Watson and Moritz 2000, 2003). Ikonic mode thinking emerges at approximately 18 months of age, and concrete-symbolic thinking at approximately 6 years of age (Biggs and Collis 1991). Concrete-symbolic thinking involves the use of writing, mathematical symbol systems, and other symbolic devices to make sense of the world (Biggs and Collis 1991). Ikonic mode thinking also involves making sense of the world; however, the sense-making tools differ. Oral language and storytelling are often the tools of choice because these representational systems are available to children before they have mastered concrete-symbolic devices encountered in primary school (Biggs and Collis 1991). In this report, we focus primarily on the ikonic mode because it is the first of the two that children generally acquire. At the outset, it is important to note that "replacing" ikonic mode thinking with more "advanced" thinking is not necessarily an appropriate teaching goal; ikonic thinking can complement concrete-symbolic thinking, as will be discussed.

Cycles consisting of three types of response to academic tasks occur within each SOLO mode: unistructural, multistructural, and relational (U-M-R cycles). Unistructural responses incorporate one aspect that can ultimately contribute to solving a task. Multistructural responses include multiple aspects. Relational responses include multiple aspects integrated under a unifying theme. Previous research on primary students' probabilistic thinking has mainly focused on U-M-R cycles within the concrete-symbolic mode because concrete-symbolic thinking is often required for elementary school probability tasks. However, several instances of ikonic mode thinking have also been reported among primary students (e.g., Jones et al. 1999a; Tarr and Jones et al. 1997; Watson and Moritz 2000, 2003).

In the present study, we aim to expand research on children's earliest learning of probability by characterizing U-M-R cycles within the ikonic mode. We examine children's thinking tendencies from previous research on probabilistic thinking as a starting point for this endeavor, focusing on responses from children who did not employ concrete-symbolic mode thinking in response to tasks. At least five different tendencies can be discerned among such responses: subjective thinking, focus on seemingly irrelevant task aspects, myths and imaginative stories, idiosyncratic mental imaging, and deterministic world view. Sometimes more than one of these tendencies appears within a given task response. Some manifestations of these tendencies provide a foundation for the beginnings of normative thinking, and others do not (here, "normative" is used to describe commonly accepted thinking among mathematicians; Shaughnessy 2007). Next, we discuss each tendency, illustrating each one with examples from the literature. Then, we consider how the tendencies might be used as the basis for postulating U-M-R cycles within the ikonic mode.

Subjective thinking

Children sometimes use subjective thinking when asked to determine the likelihood of an event (Mooney et al. 2014). For instance, when Jones et al. (1999a) asked why one spinner had a greater chance of landing on red than another, a child replied, "It's my favorite color" (p. 496). Similar responses incorporating favorite color have been observed across several studies (Jones and Thornton 2005; Langrall and Mooney

2005; Watson et al. 1997). When asked to deal with quantitative rather than categorical phenomena, children reasoning subjectively may respond with their favorite numbers or the ages of their siblings (Watson 2018). Memories of personal experiences are also sometimes used as basis for prediction; for instance, when asked about the probability of obtaining a 6 on a die, one student studied by Watson and Moritz (2003) replied, "6s don't come up as often as smaller numbers…especially when you want to get a 6 to start (a game)" (p. 283). Subjective responses to probability tasks, in general, are laden with egocentrism and hence tend to run counter to normative probabilistic thinking.

Focus on seemingly irrelevant task aspects

At times, children focus on aspects of a task that, from an expert's perspective, seem irrelevant to its solution. Examples include focusing on the size of the coin being flipped in an experiment (English and Watson 2016) or the shape of the random generator (Truran 1995) in a probability task (in this article, "random" is used to describe "phenomena having uncertain individual outcomes but a regular pattern of outcomes in many repetitions" (Moore 1990, p. 98); Batanero and Serrano (1999) provide a review of this conception of randomness in contrast to others). When asked to determine the typical value in a data set or its spread, students may select the highest number in the data rather than examining the aggregate (Jones et al. 2000). At times students may select numbers and combine them using an operation that does not seem to match the mathematical structure of the problem. Watson (2018) observed this phenomenon when asking a student to predict the number of red lollies that would be drawn in a sample. In response, the student added $2 + 8$; this combination of numbers had no apparent connection to a solution to the task. In some such cases, there may or may not be an underlying internally consistent thought pattern; the structures of these thought patterns may or may not be compatible with normative mathematical structures incorporating the relevant concepts.

Myths and imaginative stories

In some cases, children respond with myths and imaginative stories when given probability tasks. Superstitions and non-mathematical beliefs about luck can create a basis for imaginative narratives (Amir and Williams 1999). One common type of story incorporates the notion that a person, device, or external force has control over the outcome of a random event (Metz 1998). Children also sometimes believe that they themselves have control over the outcomes of random situations (Langrall and Mooney 2005). Perceived methods for obtaining control might include crossing one's fingers or choosing lucky numbers (Williams and Nisbet 2014). Some claim that random generators themselves want certain outcomes, reflecting animistic beliefs about devices such as dice and spinners (Truran 1995). Some children claim that some people are luckier than others (Amir and Williams 1999). Others believe that some people are especially skilled at shaking dice in a certain way to attain desired outcomes (English and Watson 2016). Although this type of thinking tends to run counter to normative conceptions, children's observations that dice or other devices can be manipulated can at times be useful; concerns about systematic, plausible manipulation can help establish the need to attend to the quality of the device and/or process used to gather data.

Another type of imaginative story involves inventing a narrative to support a predetermined answer. Watson et al. (1997) posed a task that involved predicting whose name would be drawn from a hat containing all of the names of the children in a class. One student believed that a girl would be drawn because the teacher of the class was a girl, and the student invented an imaginative narrative to support her conclusion. Jones et al. (1999b) reported similar reasoning from a child who was asked which names could be drawn from a hat containing names of students in a class. Rather than considering all possible outcomes, the student responded, "Sergio, because I think he's going to win" (p. 151). In such cases, students appear to build their response narratives upon their pre-existing beliefs about the correct answer to the task. This contrasts with the normative thinking tendency to view such situations as not depending on the personal beliefs or characteristics of the observer.

Idiosyncratic mental imaging

When asked to think about the results of drawing objects from containers, children's idiosyncratic mental images of the positioning of the objects within the container can be quite influential. When Jones et al. (1999a) asked which color bears could be drawn from a box containing various colors, one student responded "Red, it's on the top" (p. 504). The student's mental image that red was positioned near the top of the container drove her thinking about the situation. Similar types of reasoning have been observed when students are asked to predict the result of drawing lollies from a container with various colors. Students sometimes make conjectures about where the different colors of lollies are within the container or think about which section of the container they might draw from (Watson 2018). In such cases, a student's initial mental image of the container composition can become an important factor in thinking about the task. Such mental images are not necessarily counter to normative thinking. In fact, a mental image of a container that is not well mixed can help establish the need to shake it up in situations where it is appropriate and necessary to do so.

Deterministic world view

Engaging meaningfully in probability tasks requires embracing a stochastic rather than deterministic view. Doing so can be challenging. Langrall and Mooney (2005) noted that "Young children tend to view the world in a deterministic manner, often attributing causal effects to situations of chance" (p. 95). This deterministic world view sometimes causes children to expect that all events behave according to some order or purpose. One Piagetian task, for example, involves tilting a box so that marbles roll from one side to the other. Young children frequently expect predictable patterns to govern where the balls land when the box is tilted (Jones and Thornton 2005). In Piagetian spinner tasks, children often expect a great deal of regularity and order in the results (Inhelder and Piaget 1958). Such expectations of regularity may arise from deterministic explanations pervasive in formal education (Fischbein 1975). School mathematics, in particular, often involves deductive reasoning and procedures sure to produce correct answers rather than uncertainty. Developing conceptual understanding of probability requires balancing attention to the deductive and procedural aspects of the subject with the study of uncertainty.

Deterministic world views are often manifested in the language children use when asked to consider possible results of random spins or draws. Jones et al. (1999a) described a "sample space misconception" that leads students to speak as if certain outcomes are guaranteed when they actually are not. In one case, when a student was asked which colors could come up on a multi-color spinner, the student stated, "Red. It's going to win" (p. 496). No consideration was given to other possible outcomes. In the same study, when a student was asked about the outcomes that could be obtained by rolling a die, she stated, "Six [is the only chance on the die because] it will win for me" (p. 504). In each case, students gave deterministic predictions about the outcomes rather than considering all possibilities. Doing so prevented engagement with the probabilistic elements of each task. Setting aside a discourse of determinism in favor of the discourse of uncertainty that underlies normative probabilistic thinking can be challenging for adults as well (Makar and Rubin 2009).

Hypothetical ikonic mode U-M-R cycles

Task responses reflecting one or more of the five thinking tendencies described above have often been classified as *prestructural* in previous research in probability and statistics education (e.g., Jones et al. 1999a; Tarr and Jones 1997; Watson and Moritz 2003). Biggs and Collis (1991) described prestructural responses as those in which "The task is engaged, but the learner is distracted or misled by an irrelevant aspect belonging to a previous stage or mode" (p. 65). This description implies that ikonic mode thinking can impede progress toward a normative solution to a task that requires concrete-symbolic thinking. However, it is not accurate to view all instances of ikonic mode thinking as impediments. At times, ikonic mode thinking can play a complementary role in problem-solving. For example, Watson et al. (1995) found that imaginative speculation about causes of data values while examining data cards (thinking characteristic of the ikonic mode) supported the sorting of cards in a systematic fashion (thinking characteristic of the concrete-symbolic mode). Watson and Collis (1994) described how students used ikonic support to help compare the test scores for two classes; visual strategies for comparing two graphs (ikonic) were used by some students to help support strategies based on totals and averages (concrete-symbolic). In our preceding overview of the literature, we highlighted additional examples of how ikonic thinking tendencies might support normative thinking while also acknowledging that such tendencies at times run counter to normative thinking.

Given the different roles ikonic mode thinking can play in generating solutions to tasks, Collis and Romberg (1990) postulated two ikonic mode problem-solving pathways. Both pathways start with the generation of intuitions and mental images about the given task. However, pathway 1 does not lead to a normative solution because the images and intuitions are processed using criteria not related to the mathematics of the task; students may employ hunches or personal beliefs instead. Pathway 2 differs in that the ikonic images, intuitions, and narratives generated can support concrete-symbolic thinking and lead to normative solutions to tasks. For example, individuals may use their visual impressions of graphs of distributions to help select appropriate measures of center for data (Callingham 1997). This sort of ikonic mode support can also be helpful to professionals; Biggs and Collis (1991) gave an example of this in describing how a scientist's work with the structure of organic ring compounds was supported by an idiosyncratic mental image of six snakes chasing each other. Hence, pathways 1 and 2 differ from one another in regard to their potential to support normative scientific thinking.

Pathways 1 and 2 are nonetheless similar in their progression from a single idiosyncratic mental image, belief, or intuition to the generation of a coherent narrative built using these ikonic representations. Such a progression is resonant with the structure of a U-M-R cycle, which begins with a single element, collects multiple related elements, and culminates with the construction of a coherent structure to tie the elements together. Because each mode of representation may contain multiple U-M-R cycles (Pegg and Davey 1998), it is plausible that two U-M-R cycles, corresponding to pathways 1 and 2 respectively, might be identified within the ikonic mode.

Table 1 merges the hypothesis of two ikonic mode U-M-R cycles with observations about the five ikonic mode probabilistic thinking tendencies documented in previous literature, and in the process provides a working outline of the SEPT framework. Each U-M-R cycle shown in Table 1 culminates with the generation of an internally consistent narrative that uses ikonic imagery, consistent with Biggs and Collis' (1991) observation that ikonic mode thinking can largely be characterized as the construction of stories with plot elements and narratives. However, U-M-R cycle 2 has potential to support normative solutions to tasks, and U-M-R cycle 1 does not. U-M-R cycle 1 incorporates ikonic elements such as subjective preferences, myths, superstitions, and deterministic beliefs. U-M-R cycle 2 incorporates ikonic elements such as personal experiences with random number generators, images of cubes in a bag, the sizes and shapes of spinners and coins, and the manner in which different people might operate such random generators. In essence, U-M-R cycle 1 culminates with narratives that offer explanations that represent alternatives to normative disciplinary thinking in probability; U-M-R cycle 2 culminates with narratives that have the potential to complement/support normative probabilistic thinking. Subsequently, we refer to U-M-R cycle 1 as *normative incompatible* and U-M-R cycle 2 as *normative compatible*. We posit that it is valuable for teachers to encourage and elicit normative compatible ikonic mode thinking. To further illustrate and explain the theoretical structure of the two U-M-R cycles in Table 1, we next apply the cycles to empirical data.

Applying the SEPT framework to empirical data

The data we examine next came from design-based research (McClain and Cobb 2001; Smit and van Eerde 2011; Cobb et al. 2017), we conducted on teaching probability. Each child in the study had a 30-min individual problem-solving pre-interview, seven weekly 1-h group instructional sessions, and a 30-min individual problem-solving post-interview. All of these interactions were video recorded and transcribed for analysis. The study occurred as part of a larger undergraduate research project; the first two authors served as faculty mentors and the second two as instructors. Further details about the nature of the undergraduate research project and the design-based research methodology it employed have been published elsewhere (Groth 2017; Groth et al. 2016; Groth et al. 2018).

During our study, two students (pseudonyms Kate and Isaac) exhibited ikonic mode thinking in response to some tasks. Initially, we classified all of Kate and Isaac's ikonic

Table 1 Structure of early probabilistic thinking (SEPT) framework

	Normative incompatible cycle	Normative compatible cycle
Ikonic thinking tendencies	Personal preferences or characteristics, myths, superstitions, animism, deterministic beliefs	Experiences observing random generator results, image of position of marbles in a container, image of random generator, attention to operation of random generator
Unistructural response	Uses normative incompatible thinking tendencies in attending to one task aspect	Uses normative compatible thinking tendencies in attending to one task aspect
Multistructural response	Uses normative incompatible thinking tendencies in attending to multiple task aspects	Uses normative compatible thinking tendencies in attending to multiple task aspects
Relational response	Weaves an internally consistent narrative using normative incompatible ikonic thinking tendencies	Weaves an internally consistent narrative using normative compatible ikonic thinking tendencies

mode responses as prestructural, consistent with existing research on children's probabilistic thinking. Upon closer inspection, we recognized that the responses we categorized as prestructural did not all seem to hold the same potential to support normative explanations of probabilistic phenomena, suggesting that a more fine-grained analysis would be desirable. The two U-M-R cycles shown in Table 1 took shape as we sought appropriate literature-based criteria to support such an analysis. Next, we present ikonic mode data from Kate and Isaac to illustrate the structure and potential utility of the two cycles.

Kate and Isaac

Kate was a 10-year-old Hispanic female who attended a private school, and Isaac was an 11-year-old Caucasian male who attended a public school. Both of their schools had curricula aligned with the Common Core State Standards (Common Core State Standards Initiative 2010), which do not deal with probability until grade 7 (when students are approximately 13 years old). Hence, the two students were in the early stages of learning probability even though they were no longer in the earliest grade levels in school. Although each student exhibited ikonic thinking tendencies typical of younger children in the literature, we do not claim that their thinking patterns are precisely equivalent to those of students in earlier grade levels.

Kate and Isaac were involved in the study because their parents had applied for them to participate in summer mathematics instruction at the authors' university as part of the university's undergraduate research project. The application forms for summer mathematics instruction at the university asked parents to report if their children had any learning problems of which the instructors should be aware and to comment on the children's mathematical learning needs. Kate's mother reported that Kate had a reading disability and an individualized educational plan at her school for the subject. She also wrote that Kate had always struggled with mathematics and questioned whether such struggles were related to her reading disability. Isaac's parents reported that a school specialist diagnosed him to have mathematics and reading disabilities. The parents further noted that Isaac had trouble with number sense, memorizing facts, and recognizing patterns. So, each student received special attention at their respective schools for documented learning disabilities.

Next, we discuss data gathered from Kate and Isaac during video recorded individual interviews and lessons. We focus mainly on ikonic mode task responses from the two that help illustrate the two U-M-R cycles shown in Table 1. These responses were drawn from six tasks, some of which were administered during individual interviews and some during classroom instruction. Collectively, the sample responses are meant to demonstrate the value of doing a fine-grained analysis of ikonic mode responses using the two hypothesized U-M-R cycles (Table 1) rather than just categorizing all such responses as prestructural. A summary of the sample responses to be discussed is given in Table 2.

Task 1: selecting shirts from a closet

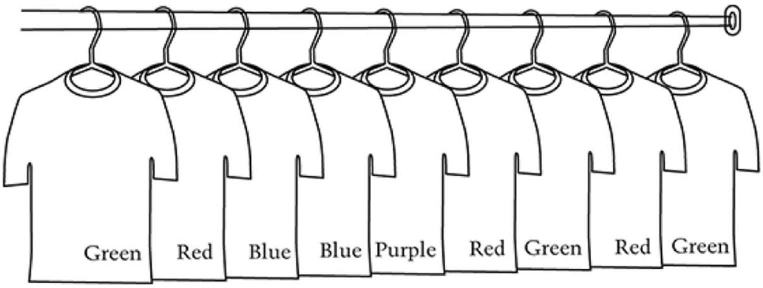
The task shown in Fig. 1 was administered during individual interviews, both before and after Kate and Isaac participated in the seven summer mathematics instruction sessions. It was drawn from the National Assessment of Educational Progress (NAEP). The task elicited ikonic mode thinking from both students during the initial interviews.

Both students believed that the person selecting shirts would reach toward the middle of the closet (shown in Fig. 1) when choosing. Using this reasoning, Isaac chose green and blue as the most likely colors to be selected. When he was asked why he chose green and blue, the following exchange took place:

> Isaac: Because green and blue are in the middle and when he reaches for the middle and it won't be the outside.
> Interviewer: So he's going to reach for shirts in the middle?
> Isaac: Uh huh

Table 2 Summary of Kate and Isaac's responses to tasks

	Kate	Isaac
Task 1: selecting shirts from a closet	Pre-interview: normative compatible ikonic (relational) Post-interview: concrete-symbolic	Pre-interview: normative compatible ikonic (relational) Post-interview: normative compatible ikonic (relational level)
Task 2: selecting marbles from a bag	Normative compatible ikonic (unistructural)	Normative incompatible ikonic (relational)
Task 3: selecting cubes from a jar	Normative compatible ikonic (relational)	Normative incompatible ikonic (relational)
Task 4: gumball machine	Normative incompatible ikonic (unistructural)	Normative incompatible ikonic (unistructural)
Task 5: predicting coin flip outcomes	Concrete-symbolic	Normative incompatible ikonic (multistructural)
Task 6: predicting dice roll outcomes	Normative incompatible ikonic (multistructural)	Normative incompatible ikonic (relational)



Mark has nine shirts in his closet as shown.

If Mark picks a shirt out of the closet without looking, which two colors have the greatest chance of being picked?

A. Blue and purple

B. Green and blue

C. Red and blue

D. Red and green

Fig. 1 NAEP question ID: 2013-4M7 #2 M170001 (U.S. Department of Education et al. 2017, n.p.). Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics Assessment

> Interviewer: OK, and why is he going to reach for the shirts in the middle?
> Isaac: Because he wouldn't walk all the way to the other side to get a shirt. He would go right from the middle.

Kate imagined a person reaching toward the middle of the closet as well, saying, "probably he would pick the middle," though she decided on blue and purple as the most likely outcomes (two colors actually closer to the middle than green and blue). Each of the responses to the item employed an imaginative story that was driven by a mental image of an individual reaching for the middle of the closet.

Kate and Isaac's pre-interview responses to the shirt task resonate with the normative compatible cycle of the SEPT framework. Their responses are relational in the sense that the idea of reaching for the middle of the closet provides a unifying thought for the narratives. Although the designers of this NAEP item intended for students to count the number of shirts of each color and make a judgment on that basis, given the task statement, it is reasonable to assume that one might use the common strategy of reaching toward the middle of the closet to select a shirt. The notion of reaching for the middle generates ikonic imagery of a sampling process. Students can understand random sampling more deeply by contrasting it with other types of sampling such as this (Garfield 2002), so ikonic mode images of various sampling processes can be valuable contributions to class discussions. Hence, there is value in eliciting this sort of thinking from students and building upon it as they develop their knowledge of the differences between various sampling procedures.

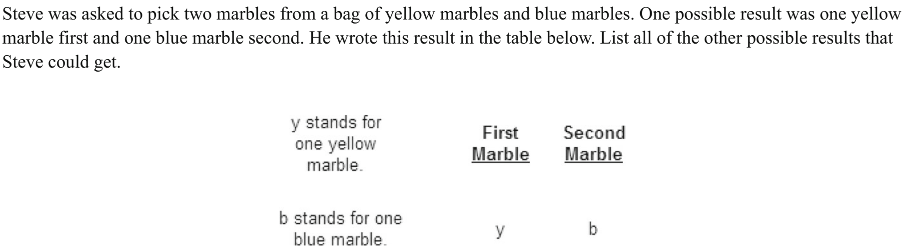
The shirt task was posed to both students again during post-interviews. Isaac once again employed ikonic imagery. He selected red and blue as the most likely colors to be drawn from the closet, explaining, "When he closes his eyes he's probably going to reach for the middle." Kate, on the other hand, looked for the most frequent colors and did not mention reaching for the middle of the closet in her explanation. Perhaps the ideal response to the task would use a combination of the post-interview thinking used by the two students. Given the absence of a clear specification of random sampling in the task, Isaac's assumption that the person would probably reach for the middle of the closet is still reasonable. However, Kate's strategy of looking for the most frequent colors is also relevant if one assumes random sampling. However, given the task statement, one could argue that the assumption of random sampling should be challenged. Isaac's ikonic mode imagery and explanation could be used to mount such a challenge. This type of challenge could be used to help students using Kate's strategy explain why they are assuming random sampling and in the process come to a deeper understanding of what random sampling entails. Ultimately, after such an exchange of ideas, students might even suggest re-wording the problem so that random sampling is a more clearly plausible assumption (e.g., changing the task so the shirts are all thrown into a hamper and shaken up before one is selected). In such a manner, ikonic mode thinking can support the concrete-symbolic thinking that leads to understanding the process of random sampling.

Task 2: selecting marbles from a bag

Another NAEP item that elicited ikonic mode thinking from both Kate and Isaac during pre-interviews is shown in Fig. 2. It was intended to assess the nature of children's beginning combinatorial reasoning.

During pre-interviews, neither student began systematically listing outcomes in order to solve the task. Kate began writing several "ys" and "bs" on her paper; when questioned about her reasoning, she explained:

> Interviewer: OK, so what are you doing, and why?
> Kate: Um, it would probably be two, three, I mean four ys and one blue.
> Interviewer: OK.
> Kate: Because there would probably be more yellows.
> Interviewer: OK, so you're saying there's more yellows than blues in the bag?
> Kate: (nods yes)
> Interviewer: OK, and why do you say that?
> Kate: Because, um, because probably like when they like mixed them together, they probably added a couple more.


Steve was asked to pick two marbles from a bag of yellow marbles and blue marbles. One possible result was one yellow marble first and one blue marble second. He wrote this result in the table below. List all of the other possible results that Steve could get.

	First Marble	Second Marble
y stands for one yellow marble.		
b stands for one blue marble.	y	b

Fig. 2 NAEP question ID: 1992-4M7 #9 M045301 (U.S. Department of Education et al. 2017, n.p.). Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992 Mathematics Assessment

In response to the same item, Isaac told an imaginative story, stating, "Because, yellow is, blue is the best color because it comes first and then that equals more, and then yellow comes last so that don't work very good." He went on to tell the story of a race between yellow and blue, with blue winning. Students offered similar types of thinking about the task during post-interviews; the lack of change in thinking likely being that we were not able to allocate time to deal with combinatorial reasoning during the summer instructional sessions.

Although Isaac and Kate both exhibited ikonic mode thinking for the marbles task, there were qualitative differences between their responses. Isaac's response started with a subjective preference for one color over the other and then culminated with a fanciful tale about why one color would win a race between the two. The response is relational in that it sketches a complete narrative, but normative incompatible because it did not contain elements that could eventually support systematic counting of outcomes. Kate's response, on the other hand, focused on how the number of blue marbles in the bag might compare to the number of yellows. Such a consideration is reasonable as one begins to form a solution to the task, though to solve it one would eventually need to realize that the precise numbers of blues and yellows need not be determined. Because her ikonic response focused on one reasonable element to consider but did not integrate it with other considerations to form a complete narrative, it could be considered unistructural in the normative compatible cycle. Kate's response could plausibly be used to start a conversation leading to a normative solution to the task, whereas Isaac's elaborate narrative may actually make it quite difficult to begin a conversation leading to a normative solution.

Task 3: selecting cubes from a jar

Another NAEP item eliciting ikonic mode thinking tendencies from both students read as follows:

> There are 6 cubes of the same size in a jar. 2 cubes are yellow. 3 cubes are red. 1 cube is blue. Chuck is going to pick one cube without looking. Which color is he most likely to pick? What is the probability of this color being picked? (U.S. Department of Education et al. 2017, n.p.).

Kate responded that yellow would be most likely, based on an idiosyncratic mental image that the yellow cubes were at the top of the jar. She said,

> If you had a bag of marbles, and you were picking out some, and someone put like yellow marbles at the top, and then put blue marbles at the bottom, and you wanted to just pick out blue, you're most likely going to get yellow.

Isaac invented a story about the individual in the problem putting the cubes on plates. When asked why he thought yellow was most likely, the following exchange took place:

> Interviewer: Okay and why do you say yellow?
> Isaac: Because if he has the yellow- if there's plates

Interviewer: Okay.
Isaac: And say that there's one plate-
Interviewer: Can you draw that for me just so, or you can write it, just so I can see what you're saying better?
Isaac: (draws a plate with three marks to represent the three different colored cubes)
Interviewer: Okay.
Isaac: So he does the yellow plate first.
Interviewer: (pointing to the student's plate) Okay, so this is a yellow plate?
Isaac: Um hum.
Interviewer: Okay.
Isaac: He's got- I don't even know what we're talking about (looks back at paper) cubes.
Interviewer: Okay.
Isaac: And then he picks up a yellow one and he looks cause it says yellow on the plate and he doesn't know it.
Interviewer: Okay. So you're just saying that he just picks it off the plate- he just picks the yellow one off the plate?
Isaac: Um hum.
Interviewer: Yeah?
Isaac: He pulls the plate to him and he picks it off and then he sees that it's yellow.

In this discussion of the task, Isaac appeared to be weaving a fanciful narrative that only tangentially connected to the task he was asked to solve.

Isaac's response to the cubes task can be characterized using the normative incompatible cycle, and Kate's with the normative compatible cycle. Isaac appeared to be delving deeply into an imaginary narrative of his own construction. Some features of the task may have served as a starting point for the narrative, but imaginary ideas about how the situation played out seemed to drive it. On the other hand, Kate's image of the composition of the cube jar was related to the task statement. She believed that the color listed first in the problem would be on the top layer, the color listed second on the next layer, and the color listed third on the bottom layer. Given these assumptions, it was reasonable for her to conclude that yellow would most likely be drawn. Nothing in the problem statement contradicted the idea that this was how the cubes might be arranged in the jar; there was no statement that the jar was shaken up so the cubes would be well-mixed. As with the shirt selection problem, Kate's ikonic mode response had potential to support normative thinking by helping contrast random sampling with other forms of sampling and by drawing attention to some hidden and perhaps unjustified assumptions in the problem statement. Both Kate and Isaac had fully developed narratives to explain their reasoning, placing their responses at the relational level of each respective cycle.

Task 4: gumball machine

Kate employed ikonic thinking once again when responding to the following NAEP pre-interview item: "In a gumball machine there are 100 red, 75 blue, 50 green, and 125 yellow gumballs. These 350 gumballs are mixed up. Sam puts money in and one gumball comes out. Which color is most likely to come out?" (U.S. Department of Edcuation 2017, n.p.). Kate chose yellow as the most likely color, explaining that it was listed at the end of the first sentence in the task. Unlike the cubes task, she did not mention a mental image of gumballs possibly being layered in a certain way to influence the outcome. Instead, she attended to a single sentence feature in the task statement that would be irrelevant from a normative perspective. Given that the response did not appear to have potential to lead to or support normative thinking and it included attention to just one task element, it can best be characterized as unistructural in the normative incompatible cycle.

Isaac's response to the gumball task appeared to be characteristic of the unistructural level of the normative incompatible cycle as well. He circled "green" when asked about the most likely color to come out. When asked to explain why, he said, "Because it has 50 in it." When asked if he could explain further, he shook his head "no" and declined to elaborate. The response was unistructural in its focus on just one specific problem element. Because he would not say why he considered 50 a relevant quantity in the problem, we assume there was again an idiosyncratic reason such as personal preference for focusing on that element of the problem; further skillful probing may or may not reveal a more nuanced reasoning pattern that would make the response normative compatible instead.

Task 5: predicting coin flip outcomes

We observed ikonic mode thinking in response to some tasks given during lessons as well as during individual interviews. In one instance during a lesson, students were given the task of predicting how many tails would be obtained when flipping a coin 20 times. Isaac predicted 5 tails, and then the following exchange took place:

> Teacher: So why um 5 times on tails? Why do you think?
> Isaac: Because…cause if you find a dime or a quarter and it's landed on tails sometimes if you see 5 coins in a row on the ground landed tails that's why.
> Teacher: So you'll never see a coin on the ground with heads on it?
> Isaac: You might.
> Teacher: You might? So do you think it's more likely to be on tails or more likely to be on heads?
> Isaac: More likely to be on tails because everybody talks about it.

Isaac's reasoning about the task appeared to be driven by an image of coins laying on the ground and a belief that tails was more likely. The belief that tails was more likely was grounded in his perceptions of what others had said about the outcomes of coin flips. His task response seems best classified as belonging to the normative incompatible cycle because using perceptions of anecdotes to make predictions differs sharply from normative practices such as conducting systematic trials. Isaac's response is perhaps best classified as multistructural within the normative incompatible cycle because it incorporates multiple elements, including an image of coins on the ground and perceptions of reports given by others, but does not build them into a narrative to as great of an extent as observed for his responses to the shirt task and the cube task. Kate's reasoning on the same task was classified as being concrete-symbolic; she reasoned there would be 10 tails because she expected equal numbers of heads and tails.

Task 6: predicting dice roll outcomes

Isaac also exhibited normative incompatible cycle thinking in response to the classroom task of predicting how many times each face on a die would be obtained if the die were rolled 60 times. He suggested that the faces of the die control the outcomes. During class discussion, he said, "You might roll a 3 a lot more than 1 and 2." When asked to explain his thinking, he replied, "Because you're not controlling the dice the dice goes everywhere and they pick their own numbers." With the suggestion that the dice "pick their own numbers," Isaac appeared to be embracing animistic beliefs about the dice, implying they can will certain outcomes. This sort of animistic belief belongs to a system of thinking separate from normative disciplinary reasoning.

Kate also expressed a belief that certain numbers were more likely to be obtained than others. She predicted that 6 would be rolled 43 times, reasoning, "the higher the number it will probably get bigger." The idea that larger numbers on a die are more likely to be obtained is contradictory to normative reasoning about the situation rather than potentially supportive of it, so the response seems best classified as normative incompatible. Unlike Isaac, Kate did not offer an explanation of why she considered the larger numbers to be more likely. Animistic reasoning may have been at the core of her reasoning, as it was with Isaac's, but it was not clear from her response. Isaac's response, which contained the unifying explanatory thread of animism, could be considered relational in the normative incompatible ikonic mode; Kate's response, which lacked such a unifying thread yet included attention to both the outcomes and their frequencies, could be considered multistructural in the normative incompatible ikonic mode.

Discussion

The SEPT framework contributes to the emerging body of research in early probabilistic thinking by drawing a distinction between normative compatible and normative incompatible ikonic mode thought and offering a hypothetical structure to characterize the two pathways. Next, we offer some final reflections on the value of encouraging normative compatible ikonic mode thinking, make conjectures about tasks to help it develop, and also make conjectures about how students might shift from normative incompatible to normative compatible ikonic mode thinking.

The value of encouraging normative compatible ikonic mode thinking

Biggs and Collis (1991) observed,

> (The) concrete-symbolic mode evolves from sensorimotor and ikonic foundations, so that any topic raised at the concrete-symbolic stage has an "ancestry" in the earlier modes. The problem with direct instruction is that it may short circuit

> this existing experiential hierarchy, substituting a network of concepts and propositions that are self-referential, and that co-exist within the concrete-symbolic mode itself (p. 69).

To illustrate this "short-circuiting" phenomenon, Biggs and Collis went on to discuss students who can give detailed explanations of photosynthesis but then are not able to explain the difference between how plants and animals process food, and other students who pass examinations in modern physics yet still harbor Aristotelean views of the subject. In such cases, students learned only at a surface level because their thinking was not grounded in earlier experiential modes. Lacking such grounding, students resort to strategies such as rote memorization to satisfy examination requirements.

Teachers can help students avoid rote learning by encouraging normative compatible ikonic thinking. Responses to the shirts task (task 1, Fig. 1) provided one example of this. Although Kate and Isaac did not initially give concrete-symbolic normative responses, the thinking they exhibited was correct. Because the task did not specify that one was reaching randomly into the closet, it was reasonable to assume that an individual might reach for the middle. One could argue that Kate and Isaac thought more deeply about the task statement than a student who just assumed random sampling was taking place even though nothing in the problem said it was. One almost needs to ignore reasonable mental imagery of the sampling process in this case to believe that random sampling is plausible. Similar remarks apply to Kate's responses to task 2 (Fig. 2, marbles) and task 3 (cubes). Injecting mental images such as these into class discussion can help all students think about different types of sampling processes and their contrast with random sampling. Because such contrasts help bring the idea of random sampling into sharper relief and also bring to light questionable hidden assumptions in task statements that may be overlooked by students reasoning on a surface level, such responses can support normative thinking rather than contradict it.

Conjectures about how to encourage normative compatible ikonic mode thinking

If one accepts the argument that encouraging normative compatible ikonic mode thinking is an important goal, it is important to develop strategies for attaining this goal. In particular, we need tasks that encourage normative compatible ikonic mode thinking and strategies for facilitating discussions about the tasks.

Although the NAEP tasks we posed elicited some normative compatible ikonic mode thinking, they may not be ideal for classroom use. Our elicitation of ikonic mode thinking with these tasks was largely accidental. Those with the initial goal of encouraging ikonic mode thinking can be more intentional in task design. Such tasks would encourage students to engage meaningfully with normative compatible ikonic mode thinking tendencies (second column of Table 1). For example, teachers might have students put together a bag of marbles with a layer of blue cubes, a layer of red cubes, and a layer of green cubes. Students could talk about the bag and perhaps draw a picture of what it looks like. Next, students could shake the bag up, talk about what changed, draw another picture, and explain why their first picture is different from the second. This sort of task, which also incorporates antecedent sensori-motor experiences, could ultimately provide ikonic support for the concrete-symbolic task of explaining how random sampling differs from other forms of sampling. Similarly, students might be encouraged to construct random generators, alter their characteristics, and discuss how they changed; this sort of attention to the characteristics of random generators is largely overlooked in school (Watson and Moritz 2003). Not providing children such experiences may lead to surface-level understandings of random processes when they encounter concrete-symbolic probability tasks in later school years. Thus, there is a pressing need to construct tasks that intentionally encourage normative compatible ikonic mode thinking and leverage sensori-motor experiences.

When concrete-symbolic probability tasks are introduced later in primary school, eliciting the thinking of students with vivid normative compatible ikonic thinking tendencies is still of value during classroom discussions. As discussed, such tendencies can support normative understanding. In selecting students to share their thinking during such class discussions, care must be taken not to overlook students who have been diagnosed with learning disabilities. In our study, Kate and Isaac, both of whom had been diagnosed with learning disabilities, at times generated normative compatible ikonic responses. The extent to which students diagnosed with learning disabilities engage in ikonic mode probabilistic thinking is an interesting research question awaiting further attention; at present, however, we can at least say that the thinking of these students should not be overlooked. When normative compatible ikonic strategies are judiciously chosen for sharing during class discussion and compared against other students' responses, they have the potential to enrich learning by introducing ikonic support for concrete-symbolic mode thinking.

Shifting from normative incompatible to normative compatible thinking

Because the goal of schooling is to introduce students to normative disciplinary reasoning, it seems safe to say that it is desirable to shift students from normative incompatible to normative compatible reasoning (though we cannot claim definitively, from our data or the existing literature, that normative concrete-symbolic always follows more directly from the normative compatible ikonic mode cycle). For instance, it seems apparent that teachers would want to help Isaac shift his normative incompatible ikonic thinking about selecting cubes from a jar toward the type of normative compatible thinking Kate displayed for the task. How this can be accomplished, however, is a question that is ripe for future research.

Confronting students who hold non-normative beliefs about probability with empirical data contradicting those beliefs is one means of prompting them toward normative compatible thinking. For example, English and Watson (2016) reported that students who initially held equiprobability biases about predicting the outcomes of tossing two coins revised their beliefs after reflecting on multiple coin flip trials. Given such research findings, it seems worthwhile to continue to employ strategies that prompt students to examine their beliefs in light of empirical data. However, a word of caution is in order. One of the most robust findings of psychological research is that when confronted with data contradictory to their belief system, humans tend to reject the data or explain the data in terms of the belief system they hold rather than re-examining and altering the belief system itself (Chinn and Brewer 1993). The more entrenched the belief system is, the less likely it is to be altered. It is not safe to assume that a few encounters with data contradictory to a normative incompatible belief structure will prompt an individual to change the structure immediately. Sustained engagement over a prolonged period of time may be necessary.

School curricula that lack attention to probability in the early grades do not purposefully allow for a prolonged period of time to challenge normative incompatible ikonic belief structures. A child with fully developed relational thinking in the normative incompatible cycle may be less likely to change beliefs about probability when those beliefs are not challenged multiple times, underscoring the questionable nature of the growing practice of removing probability from the early years of school (Greer 2014; Langrall 2018; Martignon 2014). Given the importance of ikonic support for concrete-symbolic thinking, a more productive path would be to re-cast the role of probability in early childhood and primary school rather than removing it entirely from the curriculum. The normative compatible ikonic tasks suggested earlier in this discussion section provide some starting conjectures about appropriate content for this endeavor; however, there is a great need to more fully develop standards and accompanying tasks that encourage normative compatible ikonic mode thinking in early childhood and primary school. Having students gather data from empirical trials seems essential to such standards and tasks. For example, after observing the children's thinking in response to tasks 5 and 6 in this manuscript, we had them gather and analyze data from dice rolls and coin flips (see Groth et al. (2019) for a detailed description of our approach).

Conclusion

As standards and tasks to develop normative compatible ikonic thinking are developed, theory about the structure of ikonic mode thinking in probability must also continue to develop. We view this report as an initial seed in such theory development. Theory should continue to develop in tandem with accompanying tasks and learning standards in a bootstrapping fashion. That is, as teaching materials are developed and tested, they should inform the construction of theory, and theory should inform the construction of classroom materials. Design-based research encourages this sort of cyclic, bootstrapping dynamic to produce long-term, iterative development (Cobb et al. 2017). The SEPT framework was initially generated during design-based research, and we expect that subsequent studies employing such a methodological paradigm will help refine and improve it. The SEPT framework is best viewed as a living document that provides an initial direction for further theorization and curriculum development related to ikonic mode probabilistic thinking. We encourage others to subject it to further empirical and theoretical scrutiny and suggest modifications to improve it as necessary.

As the SEPT framework continues to develop, we expect several interesting research questions to emerge as well. For example, do children who consistently exhibit normative compatible ikonic thinking more readily transition to normative concrete-symbolic reasoning? Would younger children exhibit as many different levels of response to tasks as Kate and Isaac? What other characteristics distinguish normative compatible from normative incompatible thinking? What models of classroom discourse can be used to encourage normative compatible responses? Why do children exhibit normative compatible thinking in response to some tasks and normative incompatible thinking in response to others?

As research on early probabilistic learning continues, we hope that the SEPT framework helps portray ikonic mode thinking in a new light. The SEPT framework arose because we ultimately found the label "prestructural" to be too broad to capture all of the data generated by Kate and Isaac. Re-examining their thinking using normative compatible and normative incompatible ikonic mode U-M-R cycles helped us make important theoretical and curricular distinctions. We encourage researchers investigating early probabilistic learning to look for similar distinctions in their own studies rather than considering all ikonic mode responses to be of the same value. As the structure of ikonic mode probabilistic thinking comes into sharper relief, the field will be in better position to design early childhood and primary standards and curricula that provide an experiential basis for progressively more abstract probabilistic thinking.

Funding information This material is based upon work supported by the National Science Foundation under Grant Number DRL-1356001. 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

Amir, G. S., & Williams, J. S. (1999). Cultural influences on children's probabilistic thinking. *Journal of Mathematical Behavior, 18*, 85–107. https://doi.org/10.1016/S0732-3123(99)00018-8.

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton: Curriculum Corporation.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education, 30*, 558–567.

Ben-Zvi, D. (2018). Foreword. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: supporting early statistical and probabilistic thinking* (pp. vii–viii). Singapore: Springer.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behavior. In H. A. H. Rowe (Ed.), *Intelligence: reconceptualization and measurement* (pp. 57–76). Hillsdale: Erlbaum.

Callingham, R. A. (1997). Teachers' multimodal functioning in relation to the concept of average. *Mathematics Education Research Journal, 9*, 205–224.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1–49.

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). Reston: National Council of Teachers of Mathematics.

Collis, K. F., & Romberg, T. A. (1990). "The standards": theme and assessment. In K. Milton & H. McCann (Eds.), *Mathematical turning points—strategies for the 1990s* (pp. 173–189). Hobart: Australian Association of Mathematics Teachers.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Retrieved from http://www.corestandards.org. Accessed 20 Aug 2019.

DfEE. (1999). *The national curriculum: mathematics*. London: DfEE Publication.

English, L. D., & Watson, J. M. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education, 47*, 28–62.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3), Retrieved from ww2.amstat.org/publications/jse/v10n3/garfield.html. Accessed 20 Aug 2019.

Greer, B. (2014). Commentary on perspective II: psychology. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 299–309). Dordrecht: Springer.

Groth, R. E. (2017). Developing statistical knowledge for teaching during design-based research. *Statistics Education Research Journal, 16*(2), 376–396. Retrieved from https://iase-web.org/documents/SERJ/SERJ16(2)_Groth.pdf. Accessed 20 Aug 2019.

Groth, R. E., Bergner, J. A., Burgess, C. R., Austin, J. W., & Holdai, V. (2016). Re-imagining education of mathematics teachers through undergraduate research. *Council on Undergraduate Research (CUR) Quarterly, 36*(3), 41–46.

Groth, R. E., Jones, M., & Knaub, M. (2018). A framework for characterizing students' cognitive processes related to informal best fit lines. *Mathematical Thinking and Learning, 20*(4), 251–276.

Groth, R. E., Austin, J. W., Naumann, M., Rickards, M. (2019). Probability puppets. *Teaching Statistics, 41*(2), 54–57

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. (A Parsons & S. Milgram, Trans.). New York: Basic Books.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the learning and teaching of probability. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 65–92). New York: Springer.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*, 101–125. https://doi.org/10.1023/A:1002981520728.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999a). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30*, 487–519.

Jones, G. A., Thornton, C. A., Langrall, C. W., & Tarr, J. E. (1999b). Understanding students' probabilistic reasoning. In L. V. Stiff & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K-12* (1999 yearbook) (pp. 146–155). Reston: National Council of Teachers of Mathematics.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning, 2*, 269–307. https://doi.org/10.1207/S15327833MTL0204_3.

Langrall, C. W. (2018). The status of probability in the elementary and lower secondary school mathematics curriculum: the rise and fall of probability in school mathematics in the United States. In C. Batanero & E. Chernoff (Eds.), *Teaching and learning stochastics. ICME-13 monographs* (pp. 39–50). Springer: Cham. https://doi.org/10.1007/978-3-319-72871-1_3.

Langrall, C. W., & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic thinking. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 95–120). New York: Springer.

Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (2018). Preface. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: supporting early statistical and probabilistic thinking* (pp. ix–xxii). Singapore: Springer.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105 Retrieved from https://iase-web.org/documents/SERJ/SERJ8(1)_Makar_Rubin.pdf. Accessed 20 Aug 2019.

Martignon, L. (2014). Fostering children's probabilistic reasoning and first elements of risk evaluation. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 149–160). Dordrecht: Springer.

McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics, 45*(1), 103–129. https://doi.org/10.1023/A:1013874514650.

Metz, K. E. (1998). Emergent understanding and attribution of randomness: comparative analysis of reasoning of primary grade children and undergraduates. *Cognition and Instruction, 16*, 285–365. https://doi.org/10.1207/s1532690xci1603_3.

Mooney, E. S., Langrall, C. W., & Hertel, J. T. (2014). A practitional perspective on probabilistic thinking models and frameworks. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 495–507). Dordrecht: Springer.

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

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: Author.

Nikiforidou, Z. (2018). Probabilistic thinking and young children: theory and pedagogy. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: supporting early statistical and probabilistic thinking* (pp. 21–34). Singapore: Springer.

Pegg, J., & Davey, G. (1998). Interpreting student understanding of geometry: a synthesis of two models. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 109–135). Mahwah: Lawrence Erlbaum Associates.

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

Smit, J., & van Eerde, H. A. A. (2011). A teacher's learning process in dual design research: learning to scaffold language in a multilingual mathematics classroom. *ZDM – The International Journal on Mathematics Education, 43*(6), 889–900. https://doi.org/10.1007/s11858-011-0350-5.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal, 9*(1), 39–59. https://doi.org/10.1007/BF03217301.

Truran, K. (1995). Animism: a view of probability behavior. In B. Atweh & S. Flavel (Eds.), *Proceedings of the 18th annual conference of the Mathematics Education Research Group of Australasia* (pp. 537–541). Darwin: MERGA.

U.S. Department of Education, Institute of Education Sciences, & National Center for Education Statistics. (2017). *NAEP Questions Tool*. Retrieved from http://nces.ed.gov/nationsreportcard/itmrlsx. Accessed 20 Aug 2019.

Watson, J. M. (2018). Variation and expectation for six-year-olds. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: supporting early statistical and probabilistic thinking* (pp. 55–73). Singapore: Springer.

Watson, J. M., & Collis, K. (1994). Multimodal functioning in understanding chance and data concepts. In J. P. da Ponte & J. F. Matos (Eds.), *Proceedings of the 18th Conference of the International Group for the Psychology of Mathematics Education: Volume 4* (pp. 396–376). Portugal: Lisbon.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*, 44–70.

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*, 270–304.

Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher-order thinking in statistics. *Educational Research and Evaluation, 1*, 247–275.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9*, 60–82. https://doi.org/10.1007/BF03217302.

Williams, A., & Nisbet, S. (2014). Primary school students' attitudes to and beliefs about probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 683–708). Dordrecht: Springer.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.