Chapter 9
Statistics Learning Trajectories

Pip Arnold, Jere Confrey, Ryan Seth Jones, Hollylynne S. Lee, and Maxine Pfannkuch

Abstract Statistics curricula and pedagogy are changing rapidly in response to a growing body of research findings involving students' reasoning processes, technology capability, attention to underpinning conceptual infrastructure, and new ways of statistical practice. Because many of the statistical ideas being considered are currently not in the curriculum, many researchers in statistics education have investigated students' reasoning processes through the use of learning trajectories in conjunction with design-based research methods. In this chapter, we outline the characteristics of learning trajectories and exemplify how learning trajectories have been used in three case studies in statistics education. Commonalities and differences across the learning trajectories are discussed as well as recommendations for future research.

Keywords Statistics learning trajectory characteristics • Design-based research • Statistics learning goals • Statistics classroom-based observations • Conceptual pathways in statistics • Modifying statistical learning trajectories • Problem-based research in statistics education

P. Arnold
Cognition Education, Auckland, New Zealand
e-mail: parnold@cognitioneducation.com

J. Confrey • H.S. Lee
College of Education, North Carolina State University, Raleigh, NC, USA
e-mail: jere_confrey@ncsu.edu; hollylynne@ncsu.edu

R.S. Jones
College of Education, Middle Tennessee State University, Murfreesboro, TN, USA
e-mail: Ryan.jones@mtsu.edu

M. Pfannkuch (✉)
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: m.pfannkuch@auckland.ac.nz

9.1   Introduction

Over the last 20 years, learning trajectories (LTs) have gained prominence in statistics education research. Part of the prominence may be due to a general trend toward a participationist research and design paradigm in education (Sfard, 2005). In this paradigm, there is an emphasis on understanding the teaching and learning process as it develops in actual classrooms with researchers positioning themselves as collaborating with teachers rather than studying them—"there is a remarkable blurring of the boundaries between communities of researchers and practitioners" (Sfard, 2005, p. 401). The trend in education research toward studies with ecological validity and a participationist paradigm may have set the scene for statistics education researchers to use LTs particularly as many of them were searching for new ways to approach statistical learning.

Traditionally statistics has been taught as a series of techniques to handle and display data with little regard for students' reasoning processes and the building up of conceptual infrastructure across the grade levels. With attention now focusing on students' reasoning from data and on conceptual understanding of statistics, researchers have found that the conceptual underpinnings are not only difficult to grasp but also difficult to elucidate (cf. Chap. 8). Therefore, to explicate the conceptual foundations for and across statistical topics, it has been necessary to build new LTs within and across grade levels for research and teaching purposes. Furthermore, research in statistics education is challenging traditional curricula and pedagogy with respect to the content and the lack of attention to conceptual pathways and to research findings. This challenge is coming from researchers who are concerned about problems in students' reasoning processes and the links these problems have with instructional processes. These researchers invented innovative LTs because they were attempting to scaffold new conceptual understandings in students that were not present in current curricula. They used LTs to explore and document students' thinking as they engaged with new approaches to statistics (e.g., Bakker, 2004; Makar, Bakker, & Ben-Zvi, 2011). Hence, research and curriculum development and task design and students' thinking are both strongly connected within LTs (cf. Clements & Sarama, 2004). By following the development of students' thinking as they engage in a sequence of instructional tasks, new findings and gaps in students' thinking can emerge, which can result in new research and curricular paths for learning (see Bakker & Gravemeijer, 2004).

In Sect. 9.2 we elaborate on key characteristics of LTs and then illustrate in Sect. 9.3 the use of LTs in research with three case studies. Finally, we reflect on the case studies and discuss implications and recommendations for future research.

9.2   Characterizing Learning Trajectories

In recognition that LTs were being interpreted and applied in a variety of ways within research, Clements and Sarama (2004, p. 83) stated:

> We conceptualize learning trajectories as descriptions of children's thinking and learning in a specific mathematical domain and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a development progression of levels of thinking, created with the intent of supporting children's achievement of specific goals in that domain.

A similar conceptualization of LTs is held among statistics education researchers for the statistical domains. However, to understand the characteristics underpinning LTs, we need to return to their origin.

LTs were originally conceived as hypothetical learning trajectories in the seminal work of Simon (1995) who described from a constructivist perspective how teachers could conceptualize and enact the learning process within their classrooms. He perceived the LT as hypothetical because it was based on a teacher's prediction of the learning process before it was implemented. During implementation the LT would be constantly updated in response to observations on students' interactions and reasoning processes. Because the term LT is now commonly used in the literature, we use it to describe the predicted trajectories and the updated trajectories. Other researchers (e.g., Lehrer, Kim, Ayers, & Wilson, 2014) prefer to use the term learning progressions to reflect a more open process. Although we refer to researchers using LTs, in practice teachers and researchers often collaborate on designing and studying LTs, and teachers in their own classrooms also enact the LT teaching cycle.

The LT (see Fig. 9.1) involves defining a learning goal, considering possible learning activities and the types of student thinking and understanding they might evoke, and the hypothetical learning process (Simon, 1995). To produce a LT, a learning goal is initially defined, and then a hypothesis is formed about a particular group of students' understanding within that topic domain (Fig. 9.1(1)). The hypothesis is based on information from a wide range of sources and experiences, for example, current students' experiences in a related area, the experiences of a similar group of
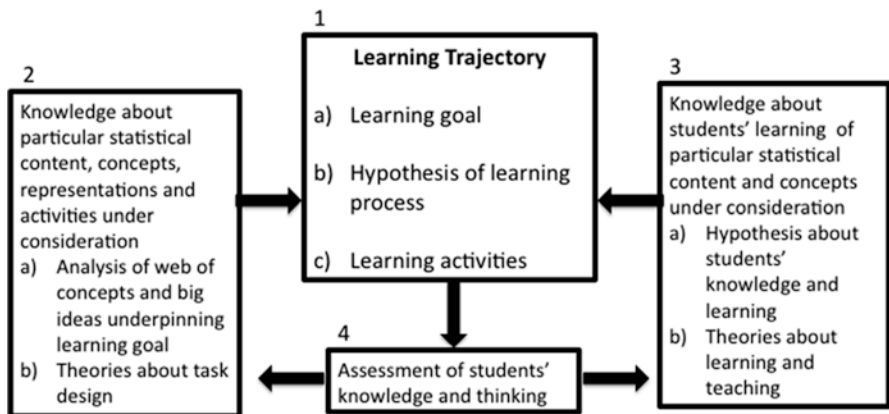

Fig. 9.1 The learning trajectory and sources drawn upon (based on Simon, 1995, p. 137)

students, information about prior knowledge that has come to light from pretesting, and data and information from the research literature (Fig. 9.1(3 and 4)). Another dimension in the creation of LTs is the undertaking of an analysis of the web of concepts including the big ideas (Ben-Zvi & Garfield, 2004) that may need to be addressed in reaching the learning goal (Fig. 9.1(2a)). For example, if the learning goal is for students to learn how to reason from distributions, then an analysis of the concepts and big ideas underpinning distributions (e.g., data, center, variability) needs to be undertaken in cognizance of future LTs that may address concepts and ideas that cannot be incorporated into the current trajectory (e.g., inference).

Based on the researchers' hypothesis of students' knowledge, skills, and possible thought processes and an analysis of the concepts and big ideas underpinning the main goal, potential learning activities and the types of thinking and learning these activities might provoke are considered. Researchers' theories about statistics teaching and learning (Fig. 9.1(3b)), their knowledge of learning in the statistics context, and their knowledge of statistics activities and representations (Fig. 9.1(2)) all intersect and come into play when considering possible learning activities (Simon, 1995). Statistical tools as mediators in the learning process need to be evaluated for inclusion in learning activities, while attention to classroom discourse and how it could be used to elicit and scaffold students' understanding is another important consideration. Other influences also impinge researchers' plans for learning activities, besides age-related development issues, such as cultural factors (Clements & Sarama, 2004), and researchers' beliefs and interests including those of the teachers that they may be collaborating with (see Chap. 10).

The learning activities can also draw on research about task design, an area of research that has only recently come to the forefront (see Watson & Ohtani, 2015). Task design is considered important because the content of the tasks affects students' learning and the nature of the learning (see Chap. 16). For research about learning, the tasks given to students have a major influence on the resultant findings about their conceptions and capabilities. Principles for designing tasks have been elucidated by Lesh and Doerr (2003) for model-eliciting activities such as personal meaningfulness to the student and the ability to generalize from the model constructed. Ainley, Pratt, and Hansen (2006) also emphasize the importance of attending to purpose and utility when designing tasks. LTs often incorporate implicit task design principles into the learning activities that are developed, suggesting more consideration is needed in this area (Fig. 9.1(2b)).

The hypothesized learning process is "a prediction of how students' thinking and understanding will evolve in the context of the learning activities" (Simon, 1995, p. 136). This is a best guess at what will happen. There is no suggestion that the instruction sequence is the only or best path for teaching and learning, only that it is *one possible route* (Clements & Sarama, 2004). A LT can also be thought of as a description of the set of intermediate behaviors (including both landmarks and obstacles) that are likely to emerge, as students progress from naïve preconceptions toward more sophisticated understandings of a target concept (Confrey, 2006). The hypothetical learning process is continually modified. This is a result of the researchers developing a broader understanding of students' conceptions in the area through

a process of reflection based on interactions with and observations of students. The researchers' thinking is modified as they make sense of what is happening in the classroom. Reflection, based on assessment of students' thinking, leads to constant adjustment and fine-tuning of the LT, the goal, the activities, and the hypothetical learning process (Simon, 1995).

The assessment of students' thinking to inform modifications to the LT (Fig. 9.1(4)) can be investigated in a variety of ways such as individual written diagnostic tests, task-based individual or group interviews requiring thematic qualitative analyses, and analyses of classroom discussion and interaction. An interesting example of addressing the problem of how to analyze classroom interaction data is found in the work of Dierdorp, Bakker, Eijkelhof, and van Maanen (2011). To determine how well conjectures about students' learning matched up with the observed learning, they used a data analysis matrix and a summary coding system for transcripts from classroom interactions in order to gain insight into how their LT supported students' inferential reasoning. More work is needed in this area to provide better evidence in research papers about how a LT supports or does not support students' learning with respect to the learning goal.

The LT systemizes and extends what good teachers do, with the difference being that within a research context, it is a deliberate act: the researchers are actively and consciously planning, reflecting, and recording actions and thoughts. As a LT is being trialed through several iterations on groups of students, the goal of the researchers is to deliberate on the observed student development together with the instructional sequence and form a localized theory of instruction (Gravemeijer, 2004). It is localized because the theory may only pertain to the group of students on which the instructional tasks were implemented, but other researchers may be able to take the theory as a framework for developing LTs for their particular group of students. Bakker and van Eerde (2015) explain that similar patterns of students' thinking can emerge across different classrooms and teaching experiments resulting in a more general theory of instruction of how a topic can be taught.

In education research, the use of LTs as a research instrument is often associated with design-based research (DBR) methodology (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Confrey & Lachance, 2000; Gravemeijer & Cobb, 2006; Prediger, Gravemeijer, & Confrey, 2015). DBR is characterized as research where students' development and progression are analyzed using deliberately designed learning activities with the aim of testing or developing theory (Bakker & van Eerde, 2015). The aims of DBR in which a new type of learning is engineered can be manifold: explanatory and advisory "to give theoretical insights into how particular ways of teaching and learning can be promoted" or predictive to state that "under condition X using educational approach Y, students are likely to learn Z" (Bakker & van Eerde, 2015, p. 431). Another characteristic of DBR is its iterative nature where cycles of preparation and design, teaching experiment, and retrospective analysis are conducted. During the teaching sequence, researchers can ascertain how the learning occurs in actual practice and through reflecting critically can then adjust or modify the plan for the next lesson. Typically these are small changes from lesson to lesson. After the teaching sequence is implemented, larger-scale modifications

can be made to the LT. DBR has recently undergone further development (see Design-Based Implementation Research, 2016). Hence, DBR methodology forms a natural partnership with LT research. Mixed methods research methodology can also be used in conjunction with LTs.

9.3 Three Case Studies of Learning Trajectories

The statistics education community has produced a number of studies that contribute to the knowledge base on LTs (Franklin et al., 2007; Lehrer et al., 2014; Rubin, Bruce, & Tenney, 1990). Research suggests that statistical concepts should be integrated into inquiry activities and that how students think about statistical concepts evolves as students grow in encountering accessible forms of variability (Garfield & Ben-Zvi, 2007; Konold & Pollatsek, 2002; Pfannkuch & Wild, 2004) that create a need for the concepts (Confrey, 1991).

In this section, with references to Fig. 9.1, we illustrate how LTs can be used in statistics education research. In the first case study, Jere Confrey and Ryan Seth Jones illustrate strategies to represent hypothesized construct maps to help teachers and students trace the growth of students' thinking about variability. Pip Arnold, in the second study, has the learning goal of making a judgment or an inference when comparing two box plots, and she exemplifies how students were scaffolded, using a hypothetical learning process, toward that goal. In the third study, Hollylynne Lee, in collaboration with Helen Doerr, designed a LT to advance teachers toward an understanding of repeated sampling for inference. All these studies used DBR. At the heart of these case studies is the big idea of variation, from the need to invent a statistic to describe the variation observed to the need to take variation into account when making an inference.

9.3.1 Case Study 1: Two Preparatory Learning Trajectories for Sixth-Grade Students toward Inventing a Statistic for Variability

9.3.1.1 Introduction

The first case study addresses students' introduction to the concept of variability, a topic studied by numerous scholars (e.g., Ben-Zvi, 2004; Garfield & Ben-Zvi, 2005; Konold & Pollatsek, 2002; Lehrer et al., 2014; Makar & Confrey, 2005; Wild & Pfannkuch, 1999). Confrey and Jones chose to approach the topic using a learning map organized around big ideas, which were broken down further into constructs with underlying LTs (Confrey, 2015). These LTs accurately characterize typical responses from students in increasing levels of sophistication. The map is used for two primary purposes: to provide professional development opportunities for teachers and to develop diagnostic assessments to gauge student progress.

9.3.1.2   The Learning Goals and the Designed Learning Process

Confrey and Jones started with the learning goals from the Common Core State Standards for Mathematics (Common Core State Standards Initiative, 2010) in the United States for sixth-grade (age 11) statistics. Through analyzing the web of concepts and big ideas underpinning the learning goals (cf. Fig. 9.1(2a)), they designed a learning map that was hierarchically organized around nine big ideas identified by Confrey. The big ideas were subdivided into one to five relational learning clusters, which were made of sets of mutually supporting constructs. Each construct was described with a corresponding learning trajectory consisting of an ordered set of indicators of increasing sophistication. These reflect the likely student behaviors and thinking that would emerge as they progressed through instruction (see Table 9.1 for the first two constructs). In statistics one big idea was "display data and use statistics to measure center and variations in distributions." This big idea was divided into three relational learning clusters: (1) displaying univariate data, (2) measuring data with statistics, and (3) displaying bivariate data. Each learning cluster was divided further into a set of connected constructs. The constructs for displaying univariate data were (1) gathering data and describing variability, (2) displaying data in novel and traditional ways, (3) comparing different displays of the same data, and (4) shape of univariate data.

The LTs were based on a synthesis of literature from statistics education research and previous iterations of the learning trajectory. For example, prior to this study, many of the behaviors and thinking about variability were articulated in the related learning cluster on modeling. However, after the foundational role of this thinking was observed in their studies for making sense of data displays and statistics, they restructured the map to include these ideas in the data display cluster. In each iteration of the LT, patterns in student thinking are reinforced, but nuanced variations or even new ways of thinking emerge and are added into the LTs.

The overarching learning goal of the trajectory for displaying univariate data was to support students to develop a conception of variability that was represented in various data shapes created by displaying data and to lay the groundwork for needing a measure of variability in later trajectories (Konold & Pollatsek, 2002; Lehrer & Kim, 2009; Petrosino, Lehrer, & Schauble, 2003). The goal was influenced by Confrey and Jones' theories about learning (cf. Fig. 9.1(3b)) the key elements of which include the role of invention and of transformation (accommodation in Piagetian terms). Another key element of their approach was to foster discourse among the students, so they could learn from each other's ideas and contributions. Teachers play a central role in bringing forth this thinking and building classroom norms valuing articulation and sharing of ideas. Their belief is that the LTs should also communicate the kinds of student statistical thinking teachers should attend to and how they fit together into trajectories of increasingly sophisticated thinking. Thinking about variability, displaying one's data, and comparing those displays prepare the ground for a discussion of data shape and statistics (Lehrer, Kim, & Jones, 2011; Petrosino et al., 2003; Schwartz & Martin, 2004). Only after students have productively struggled with these ideas are they ready to invent statistics and learn conventional definitions.

9.3.1.3   The Learning Activities and the Observed Learning Process

Confrey and Jones developed instructional materials by drawing on prior work by Lehrer (2016) and Confrey (2002), and they made use of *TinkerPlots* (Konold & Miller, 2005) and *Data Games* (Finzer, Konold, & Erickson, 2012) for data exploration and display. Hence, the learning activities were based on their knowledge of teaching strategies and resources for statistics, their knowledge of how students might learn about univariate data displays, and their understanding of the current knowledge of the students who would be in their study (cf. Fig. 9.1(2 and 3)). Diagnostic assessment to gauge student learning was also coordinated with the LTs (cf. Fig. 9.1(4)).

The following case involved 15 sixth graders (age 11) who met for 3 hours per day in a classroom on their research site for 1 week. The purpose of the study was twofold: (1) to confirm or modify the LT and (2) to collect samples of student work for professional development purposes. Thus, the research question under investigation was: What patterns of behavior, forms of representation, and ways of talking are in evidence among students when introduced to the ideas of multiple sources of variability and displaying univariate data, and how might these patterns be represented so that they are intelligible and useful to teachers?

The case study provides an image of student learning and how this learning is represented in the two constructs in Table 9.1. Throughout the description of student activity, the relevant levels are referenced within that construct. Note that in this description, Confrey and Jones are assessing students' knowledge and thinking (cf. Fig. 9.1(4)) in order to inform them whether the observed patterns of behavior are consistent with the indicators listed in Table 9.1.

Gathering Data and Describing Variability

To engage students with the problem of creating variable data (the first two levels of this construct), they asked students to consider three different questions: What is the circumference of the fountain in our courtyard?, How many M&Ms. are in one individually wrapped package?, and What is the circumference of a middle schooler's head? To highlight the challenge of variability, they left the data collection strategies open-ended and provided crude measurement tools, such as string and rulers. Under these circumstances, students produced significant measurement error.

Student measurement mistakes, though, were a resource for them to make sense of the various sources of variability in the data. To elicit a conversation about sources of variability in the data (level 4 in Construct 1, Table 9.1), the teacher posted unordered lists of the students' data and asked them in a whole class conversation "what do you notice when you look at all this variability?" They also discussed relative magnitude of the different sources they identified (level 5) by asking questions such as "what caused the variability in the different types of data?" Table 9.2 provides short examples of the kinds of student comments that are common to this discussion.

Table 9.1 LT indicators for two constructs

Constructs
Construct 1: Gathering data and describing variability
1.   Recognizes target phenomenon and asks questions about it
2.   Creates and uses data as information to answer a question
3.   Describes that some questions have uncertain answers because of variability in the data
4.   Identifies sources of variability in data
5.   Estimates magnitude of different sources of variability
6.   Categorizes sources of variability (measurement error, natural variability, production error)
7.   Anticipates variability in data across different samples
8.   Describes or predicts how a change in process affects variability
Construct 2: Displaying data in novel and traditional ways
1.   Displays data without reference to investigation
2.   Shows basic familiarity with bar graphs, pie charts, and dot plots
3.   Identifies or creates titles, labels, or keys
4.   Orders data from least to greatest without distinguishing scale from data
5.   Stacks individual values or within groups, intervals, or bins
6.   Scales using equal intervals
7.   Creates dot plots and bar graphs, knows the distinction between categorical and continuous data, and explains choices about scale, order, and grouping
8.   Creates histogram and circle graph, explains choices about intervals, and provides either count or percent within each interval
9.   Makes and justifies choices about displaying data in traditional and novel ways

Table 9.2 Key concepts and student comments about variability

Data context	Student comments
Individual measurements of the same object (in our case, the fountain in the courtyard)	"…with the fountain it's like whoa! What happened here? I see 461 and I see like 2010!" "people could have made mistakes when they measured" "people might leave a gap when they flip over the ruler"
Number of M&Ms. in different bags	"the factory probably didn't measure out the exact amount of M&Ms" "we also could have counted wrong" "It's possible to miscount 10, but it's not extremely likely"
Individual head circumferences	"…it's kind of common that everyone got different head circumferences, because not everyone's head is the same size" "the variation might be the result of different sized heads, but they also might be the result of mistakes people made"

Students observed that the variability in the fountain data was a fundamentally different kind of variability than the other two types, and they drew on their data creation experiences to generate theories about the kinds of errors that likely produced the variability. Students then shifted from describing data as measurements to calling it "opinions," indicating their feeling that there was so much variability in it that it was not "scientific" enough to be called data. This conversation ended with the teacher asking, "What would the data look like if a class of students similar to us measured the same fountain?" This question was asked to evoke early ideas about sample-to-sample variability (level 7). Students quickly responded that the data would look "similar to ours", that their data would have "the same kinds of chaos as our data," and that it would "have a similar median and mean, but the numbers will be different."

These themes ran throughout the rest of the activities. For example, they were the driving motivation for remeasuring the fountain more precisely to see how a change in process affects variability (level 8) and creating paper hats using their measurements to estimate the extent to which measurement error contributed to the variability in the head circumference data (level 5).

Displaying Data in Novel and Traditional Ways

Confrey and Jones provided opportunities for students to invent strategies for displaying their fountain data in a way that helped them think about the true length and the variability in the measurements. Students revealed a variety of strategies for displaying their data, many of which had been documented by other researchers (e.g., Lehrer & Schauble, 2002). Here two of the four displays that students invented are presented to illustrate the ways student thinking corresponds to the learning map and to show how student thinking developed as they invented and revised their displays.

Group 1 produced a dot plot without distinguishing between the data and the scale (Fig. 9.2). They explained that they wanted a display that clearly displayed every measurement observed and how often each measurement was observed. They made the decision to order the data from least to greatest (level 4), without representing gaps, but with stacking of identical values (level 5).

Group 2 created a histogram (that they referred to as a box plot) with 100 cm intervals (Fig. 9.3). Similar to group 1, they ordered the data from least to greatest (level 4), but, in contrast, they grouped and stacked all values within a 100 cm interval (level 5) and created an interval scale (level 6). These choices created a very different representation of the data, which provided a context to discuss the trade-offs between the two.

As they created these invented displays, the students sometimes showed evidence of thinking at the lower three levels of the construct as they sometimes considered decisions without referencing the question about the fountain circumference, referred to approximate notions of conventional displays, and created titles and labels. However, the most significant intellectual work for students came when they had to consider decisions about order, grouping, and scale (levels 4–6). For exam-
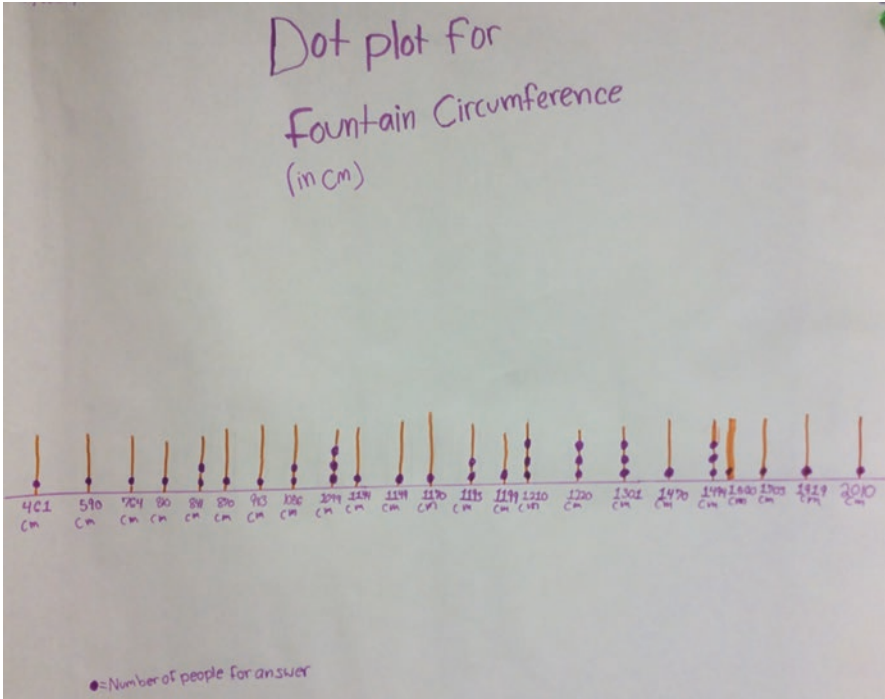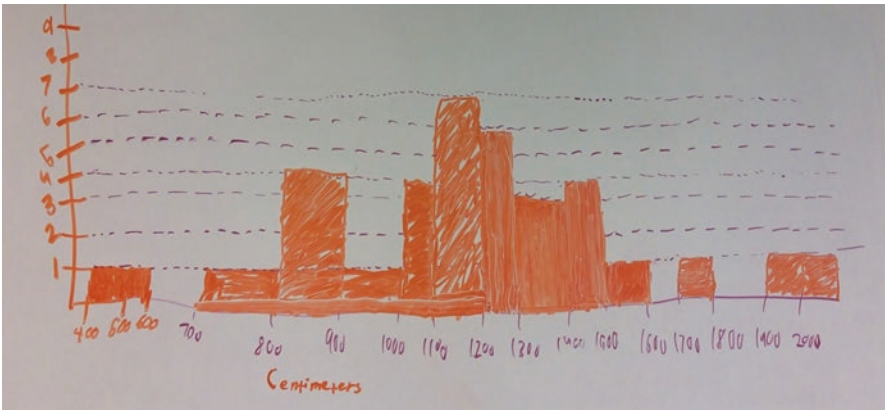
Fig. 9.2 Group 1 data display

Fig. 9.3 Group 2 data display

ple, the students had to decide if the display scale needed to include values that were not observed. Their early decisions about data displays were not driven strictly by convention, but more by their desire to make sense of variability and communicate meaning to their peers. Only after wrestling with these issues were data display conventions (levels 7 and 8) introduced, so the conventions could be rooted in stu-

dent ideas and displays. When given the opportunities to build their ideas about statistical thinking from accessible forms of variability, students often demonstrate the behaviors, strategies, and thinking described in these LTs.

9.3.1.4   Discussion and Future Recommendations

This case illustrates the potential value of designing a learning map based on an analysis of the big ideas and the web of concepts that need to be included in the LT for supporting teachers to understand student thinking. It also illustrates the need for several iterations of LTs and reflection and analysis on students' responses in the development of that map. By explicating in detail indicators of likely student thinking as they progress through the LTs, this research across multiple settings is at the stage of developing a dynamic representation of student thinking that can serve as an orienting framework for curriculum and assessment design. A product of the research for teaching is the learning map and its LTs including resource material for teaching and student work for teacher professional development which can make patterns of student thinking intelligible to teachers.

The advantage of the approach outlined in this case is that the LTs for data, variability, and statistics are related to LTs that Confrey and her team have developed and refined across all big ideas for middle grade mathematics. This provides teachers a comprehensive resource to have access to syntheses of learning trajectories. In addition, the map makes it possible to study what the effects of an overall approach informed by LTs would be as students accumulate experience with the map. Too often LT studies are difficult to continue across grades as students switch teachers and classes. In this way, research can contribute to the building up of infrastructure for supporting the long-term development of statistical concepts, a facet that is lacking in current curricula.

9.3.2   Case Study 2: Preparing Ninth-Grade Students to Make the Call—Learning How to Make a Judgment When Comparing Two Box Plots

9.3.2.1   Introduction

The second case study illustrates a LT which started with a well-defined learning goal but required thought about the underpinning concepts that students needed to experience. Because the learning goal was new to the curriculum and resources did not exist, Arnold and a research team of two statisticians and nine teachers collaboratively worked on inventing language to describe the statistical ideas and designing learning strategies and resources. The challenge in this case study was to develop a set of structured learning experiences that would enable grade 9 (age 14) students to "discover" collectively the criteria for "making the call"—making a judgment when comparing two box plots.

9.3.2.2 The Learning Goals and the Designed Learning Process

The learning goal arose from a study on the reasoning processes of students in a grade-9 class. The students were learning how to make an inference when comparing two box plots and were making the call based on a variety of criteria (Pfannkuch, 2007). From the student responses, it was clear there was no agreed understanding between the teacher and her students as to what constituted support for an inference. Furthermore, the investigative question that the students were exploring was about the populations, but the students' reasoning was based on describing the sample statistics. In New Zealand, the curriculum (Ministry of Education, 2007) and subsequent national assessment required students to make informal inferences (see Chap. 8) about populations from samples for comparative situations. This created the problematic situation. Hence, a developmental pathway was proposed for comparative situations from grade 9 to grade 12 for justifying how to make a call or make a decision about whether condition A tends to have bigger values than condition B back in the populations (Wild, Pfannkuch, Regan, & Horton, 2011). The problem for this study was how to create a LT to enable students to understand the rationale and concepts underpinning making the call using the rule as outlined in Fig. 9.4.

In cognizance of the research literature and an analysis of the web of concepts (cf. Fig. 9.1(2a)) needed for making the call, the research team determined that enabling students to make the call depended on building their understanding of a network of underlying interrelated concepts, the key concepts identified being sample, population, and sampling variability. They considered sampling variability reasoning to be at the core of statistical practice but noted it had only recently received attention in school curricula and instruction. Typically, students reach the final years of high school, where they are explicitly introduced to notions such as basic statistical inference from confidence intervals, without fundamental knowledge or experiences of sampling behavior. Despite the importance of considering variation in statistics, researchers have only in the last two decades begun to document stu-
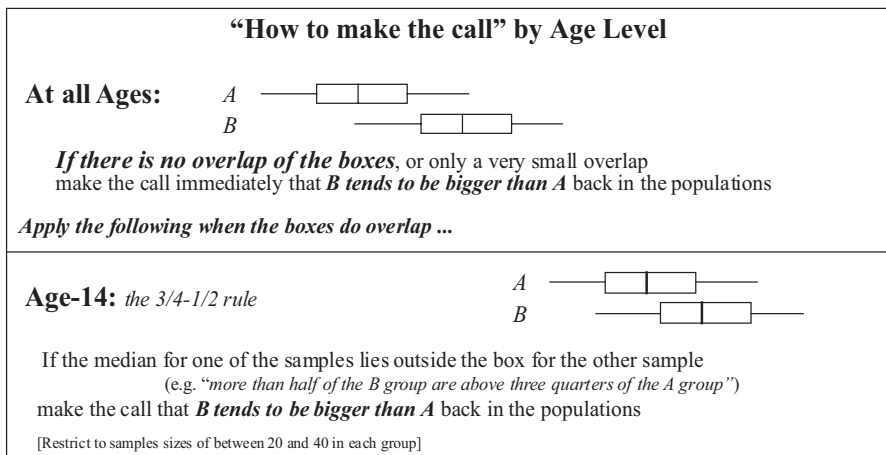

Fig. 9.4 How to make the call at ninth grade (age 14) (cf. Wild et al., 2011, p. 260)

dents' conceptions of variability. Therefore, a carefully structured set of learning experiences to support the LT was required if students were to understand and appropriate the sampling variability reasoning underpinning statistical inference. As Garfield and Ben-Zvi (2007) stated in relation to distribution, center, and variability, students "need help in developing an understanding of what these concepts actually mean and how to reason about them in an integrated way" (p. 386).

9.3.2.3 The Learning Activities and the Observed Learning Process

This case study reports on one class, although the research was undertaken with a number of classes (see Arnold, Pfannkuch, Wild, Regan, & Budgett, 2011). The planning and preparation phase involved trialing potential learning activities with the research team and making continuous changes to how the development of the three key concepts could be approached. Changes to the LT were also made when implemented in the classroom. The research question was: How can grade 9 students be facilitated to consistently and coherently make a statistical inference?

As already signaled, the three key concepts of population, sample, and sampling variability were important to support the LT for making the call. Specific learning materials and activities were created to support the development of these concepts and to support the LT (Table 9.3), which comprised 15 lessons. Some activities were deliberately planned and developed from the outset with the LT in mind, and some activities were developed as part of the ongoing reflection on the LT throughout the implementation in this class. In the description that follows are some vignettes of the learning experiences including examples of how and why the LT

Table 9.3 LT for the development of key concepts when comparing two box plots

LT for learning how to make a judgment when comparing two box plots ($n \approx 30$)
1: Sampling data from a population
• Identify population
• Pose and critique investigative questions about the population
• Recognize the need to use a sample to answer a question about the population
• Acknowledge that samples from the same population for the same variable provide similar information
• Appreciate that inferences about a population can be made from a sample
2: Sampling variability
• Recognize sampling variability in the center across multiple samples of the same size from the same population
• Appreciate sampling variability in the center and the extent of the overlap when comparing two box plots across multiple samples of the same size from two different populations
3: Developing criteria for making the call
• Recognize the salient features to focus on when comparing two groups (e.g., center, shift, overlap)
• Acknowledge that there are two situations for making the call when comparing two groups

was modified in response to the research team's observations during the preparation stage and the collaboration of Arnold and the teacher in the classroom during the implementation stage.

Population and the "Population" Bags

As the "population" of Karekare College students (a fictitious college) was going to be used extensively throughout the teaching implementation, it was important that students in the class became familiar with the data that was available. The population of Karekare College students was represented using a plastic bag filled with data cards (see Fig. 9.5). Each data card represented 1 student and contained 13 different variables relating to the student. To develop familiarity with the data, students had to work out what the different variables were on the data cards.

During subsequent lessons, whenever the teacher referred to Karekare College, she nearly always showed the population bag (see Fig. 9.5), indicating that she was referring to the whole population, not just the data cards that the students had selected. The ability to keep reminding students that they were making an inference about the population by holding up the bag was an addition to the LT by the teacher, which was regarded by her and other teachers as an important facet in aiding students' statistical reasoning processes. Giving students an image of the population was an issue that was extensively debated by the research team, because the Karekare College data had been randomly selected from a large CensusAtSchool New Zealand (2003) database and hence could be considered a sample, but then the database was also a sample itself. By considering students' understanding of these

Fig. 9.5 Karekare College data cards and the population bag

issues and the fact that they were novices (cf. Fig. 9.1(3)), the research team decided to view the Karekare College students as the population. Although the population bag provided a good visualization of the population, it was insufficient, as a posttest revealed students did not have images for or contextual knowledge about population distributions. Hence, the assessment (cf. Fig. 9.1(4)) led to the creation of an additional LT.

Developing the Idea of Using a Sample

Having established the population and the variables for which data were available, the students posed a variety of investigative questions. The teacher and Arnold together identified which of the variables would be used for the activity where the concept of sample was first addressed. From the different investigative questions that the students posed, one was selected to be explored further. The students were to answer the question: "What are typical popliteal[1] lengths of students at Karekare College?" The teacher, as part of the planned LT, asked them how they might go about answering this question, to which they ultimately replied that they would be "putting [the data] in a graph." There was then some discussion and the students, working in small groups each with their own population bag, started to graph all (616 students) of the student data, using the data cards and a pre-prepared grid. After about 10 minutes, some general discussion started about "students" not all fitting onto the grid. A student said, "I'm not going to organize the whole college into this," at which point the teacher asked, "Is there a better way than looking at the whole lot?" The ensuing discussion and action resulted in students continuing until they had filled up their group's grid or felt that the shape of the graph was not changing despite adding more data cards, i.e., they did not use the whole population, just part of it. The teacher allowed the idea of using a sample, rather than the whole population to answer the question, to come from the students—she did not say to her class at the start, "Take a sample and use this to answer the investigative question." From the observed responses of the students, it was felt that the students were developing the idea that a sample could tell them something about the population. This observation was reinforced when comparing pre- and posttest student assessment responses as in the posttest students specifically referred to the population of interest in their investigative questions and in their conclusions.

Sampling Variability

Sampling variability was explored in a number of ways. In the lesson described previously where sampling was first introduced, the students had created their graphs using the actual data cards, which provided a strong visual display. The

---

[1]The popliteal length is a measurement taken on the back of the leg from behind the knee to the floor when a student is seated.

teacher gave students time to walk around the class and see how their graph compared with other graphs in the class. The students looked at features that were similar and features that were different. All groups gave an indication of where they felt the middle of their popliteal length data was, and across the class the set of middle popliteal-lengths for the different groups lay within a 3–4 cm band. The students were able to see that the middle popliteal length was similar even though the samples were different.

Sampling variability was a focus again in a later lesson about making the call when students were looking at the patterns across different samples with respect to two variables: student heights disaggregated by gender and time taken to get to school disaggregated by mode of transportation. These two examples were deliberately chosen for the LT as they captured very clearly the two situations described in Fig. 9.4. Note that students were observing box plots, with only the box part drawn, a modification made to the LT when trialed with the research team in order to focus student attention on the salient features for making the call (see Fig. 9.6 and Arnold et al., 2011).

Making the Call

When students were looking for patterns across the sets of graphs, Arnold and the teacher realized that additional prompts were required because information about the shift and the position of medians was not forthcoming. According to Bodemer, Ploetzner, Feuerlein, and Spada (2004), leaving students to generate hypotheses about relationships on their own is very hard, and they may not pay attention to salient features. Bodemer et al. (2004) suggest that learners' interactions with learning materials should be structured so that hypotheses are formulated only on one relevant aspect of the visualization at a time, and therefore in a modification to the LT, the students were guided to first focus on the distributional shift and then on which median was bigger.

After students had sorted their samples for each question, the teacher and class reflected on the process. They described and abstracted the patterns and criteria for making a call about what was happening back in the two populations. This allowed students an opportunity to extract relevant principles (Bakker & Gravemeijer, 2004). The students noticed that in the samples for heights, the boxes were close together, whereas in the samples for time taken to get to school, the boxes were apart (Fig. 9.6). They named these two situations about the relative location of the boxes Situation 1 and Situation 2, respectively.

In the following excerpt, they explore the differences between the two situations (see Fig. 9.6):

Teacher: So in our first situation we've got the boxes. They're all overlapping; some of them are going this way and some of them are going the other way. The medians are very close together, and the medians are also within the overlap of the boxes. In the second situation, how is it different? What's different about the
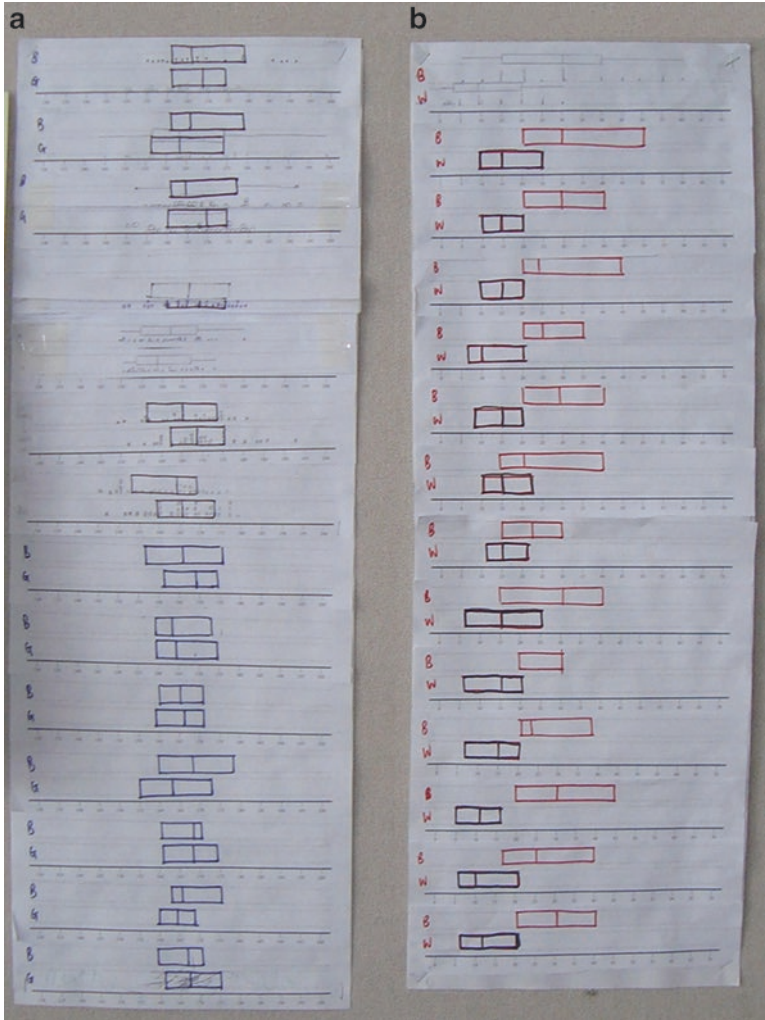
Fig. 9.6 Box plots of two situations: (**a**) samples comparing heights of girls and boys (on the *left*) and (**b**) samples comparing time taken to get to school by bus and walking (on the *right*)

overlap here? Is there no difference between the overlap on these boxes and these boxes?

Student: They're not overlapped so much.

Teacher: They're not overlapped so much. No, they're not. Okay, do they all overlap?

Student: No.

Teacher: No, so when they do have an overlap, they don't overlap much and otherwise they don't overlap at all. What can you tell us about the medians in this one?

Student: They're not overlapped.
Teacher: They're not in the overlap.

Visually and verbally, the students and teacher described differences in the two situations in terms of shift, overlap, and location of the medians. The students and teacher started to develop the criteria and language for making or not making a call. Collectively they spontaneously used hand gestures to describe the two situations, *close* (Figs. 9.6a and 9.7a) and *apart* (Figs. 9.6b and 9.7b), with vibrations to show the effect of sampling variability. Gestures according to Radford (2009) are a precursor to verbal conceptualization. The use of these gestures and the naming of the two situations, as Situations 1 and 2, by the teacher and students were built into the LT in subsequent classroom implementations.

The students also noticed that in Situation 2, there were consistent messages from the samples about the relative location of the two medians to one another back in the populations, allowing them to determine the larger of the two population medians, i.e., the median time to school by bus was always longer than the median time to school by walking. This was not the case in Situation 1. The students noted that sometimes the boys' median height was higher than the girls' median height and sometimes it was the other way around. Through recognizing and reasoning from the patterns in the two situations, they "discovered" collectively the criteria for making a call when two box plots are compared and the boxes overlap (age 14, Fig. 9.4) and do not overlap (at all ages, Fig. 9.4).

After further reinforcement of how to make the call for comparative situations, the students were given some practice material. The practice material given to the students had each student use a different sample from the same population as they worked on the same investigative question. However, this had the effect of reinforcing the idea that they could use multiple samples to make the call—an unfortunate side effect that had not been anticipated. Therefore, in a modification to the LT, all the practice material involved the same single sample for all students in the class, reflecting what happens in reality, for each investigative question. The use of multiple samples from the same population was appropriate for developing the understanding of making the call and sampling variability; however, it was not appropriate for subsequent practice as it created an unintended confusion for students.
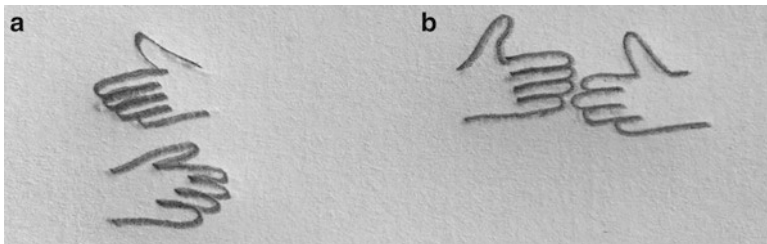

Fig. 9.7 (**a**) Hands close together mimicking two box plots overlapped (on the *left*) and (**b**) hands apart mimicking two box plots with little overlap (on the *right*)

By the end of the LT, based on an analysis of posttest data and individual student interviews, these students were beginning to understand how to make a statistical inference. They were (1) articulating the uncertainty embedded in an inference by drawing upon ideas about sampling variability, (2) making a claim about the population from the sample, and (3) explicitly providing the evidence they used from the data such as distributional shift, overlap, position of the medians, and the decision guide that enabled them to make or not make a call (cf. framework of Makar & Rubin, 2009; Chap. 8). They also seemed to understand how and why the use of the overlap and position of the medians relative to the overlap informed their use of the rule to consistently and coherently make an inference (see Arnold, 2013; Arnold et al., 2011).

9.3.2.4 Discussion and Future Recommendations

Working together to plan the LT and the carefully structured set of learning experiences to support the LT allowed the teacher, Arnold, and wider research team to get a better sense of the possible responses and outcomes for students. Modifications to the LT occurred through extensive debate within the research team, in response to students' difficulties during the lesson, from spontaneous reactions in the classroom to the issue under consideration, through reflection on the lesson or an in-depth analysis of student data after the lessons. The LT for developing the concept of making the call with grade 9 students has been the basis for teacher professional development and subsequent use in their classes.

Defining the learning goal and analyzing the web of concepts are essential ingredients for the construction of LTs. The rich interrelated conceptual repertoire underpinning statistical ideas needs further research including finding ways of developing new conceptual understandings that are not present in current curricula. As this case illustrates, LTs using DBR can assist in the development of new approaches to statistics and in understanding students' reasoning processes. Other topics in statistics need a similar focus to understand teaching and learning processes better, to generate local theories of instruction, and to explore and identify interesting phenomena.

9.3.3 Case Study 3: Preparing Teachers to Develop a Conceptualization of Repeated Sampling for Inference

9.3.3.1 Introduction

The third case presents a LT for assisting adult learners (mostly secondary and postsecondary mathematics and statistics teachers) in conceptualizing repeated sampling approaches to statistical inference, with particular attention to the role of probability models in that conceptualization. The teachers had already been exposed to formal hypothesis techniques. The intent of this case is to illustrate how and why

a team of instructors working in real graduate-level classrooms with a designed LT added further learning experiences in response to their observations on the teachers' reasoning processes.

9.3.3.2   The Learning Goals and the Designed Learning Process

The focus of the LT in this case study was to assist teachers in conceptualizing a repeated sampling approach to inference and to consider their learning with this approach. In a repeated sampling approach to inference, students and teachers should be conceiving of the observed outcome (from an observational study or an experimental design) as resulting from a process that is repeatable and that repeating the process may result in a different outcome. Thus, the question becomes: How unusual is what happened in the particular instance that we know about already? In other words, what is the likelihood of a particular outcome occurring if a process is repeated many times?

Lee and Doerr considered learners' use of probability models as essential to conceptualizing a repeated sampling approach to inference. To produce the LT, they considered the research literature and curriculum development in recent years that had focused on understanding inference and using simulation to enact resampling approaches (cf. Fig. 9.1(2 and 3)). For example, Saldanha and Thompson (2002) reported that when students can visualize a simulation process through a three-tier scheme, they develop a deeper understanding of the process and logic of inference. This scheme is centered around "the images of repeatedly sampling from a population, recording a statistic, and tracking the accumulation of statistics as they distribute themselves along a range of possibilities" (p. 261). Lane-Getaz (2006) offered the simulation process model (SPM) to describe the process of using simulation to develop the logic of inference starting with a question in mind, "what if," to investigate a problem including three tiers: population parameters, random samples, and distribution of sample statistics. In line with Lane-Getaz's suggestion, Garfield and Ben-Zvi (2008) and Garfield, delMas, and Zieffler (2012) used a generalized structure to the logic of a simulation approach to inference in their curriculum materials. Their structure includes specifying a model, using the model to generate simulated data for a single trial and then multiple trials, each time collecting a statistic of interest, and finally using the distribution of collected summary measures to compare observed data with the behavior of the model.

Saldanha and Liu (2014) described work with learners in repeated sampling tasks and made the case that students should develop a stochastic conception of an event that "entails thinking of it as an instantiation of an underlying repeatable process, whereas a non-stochastic conception entails thinking of an event as unrepeatable or never to be repeated" (p. 382). Such a stochastic conception includes seeing an event as an expression of some process that could be repeated under similar conditions that produces a collection of outcomes and "reciprocally, seeing a collection as having been generated by a stochastic process" (p. 382). All this research literature fed into the development of the LT (cf. Fig. 9.1(2)) includ-

ing the influence of the modeling perspective of Lesh and Doerr (2003) and the importance of a careful model development sequence for learners. Such a model development sequence emphasizes how learners develop their own models of a context within a LT.

Lee and Doerr's learning goal was for teachers to develop a stochastic conception of events and a generalizable model that they could use to approach inference situations using a repeated sampling approach and for them to be able to assist others in using such an approach (cf. Fig. 9.1(1a)). This model includes understanding the relationships among the problem situation, physical enactments of sampling, representations of those enactments, computer representations, and the underlying randomization (i.e., the probability models discussed above), the distribution of the statistics of interest, and how to interpret and use such a distribution (a sampling distribution) to make a decision. In order for learners to develop that model (and the entailments needed for teaching that model), they hypothesized that they should be able to make connections to and use the underlying probability model of repeatable actions with unpredictable outcomes.

The initial LT of Lee and Doerr is depicted in Fig. 9.8. This represents the key experiences they felt would lead to a generalizable model for how to use a simulation approach to inference. The key experiences in the trajectory are bolded in the center, while the statistical concepts that should be emphasized at each phase in the trajectory are noted on the right, and pedagogical considerations that could be useful in participants' own teaching practices are noted on the left. Both the statistical ideas that needed to come to the fore and pedagogical issues could help inform the development of teachers' understandings.
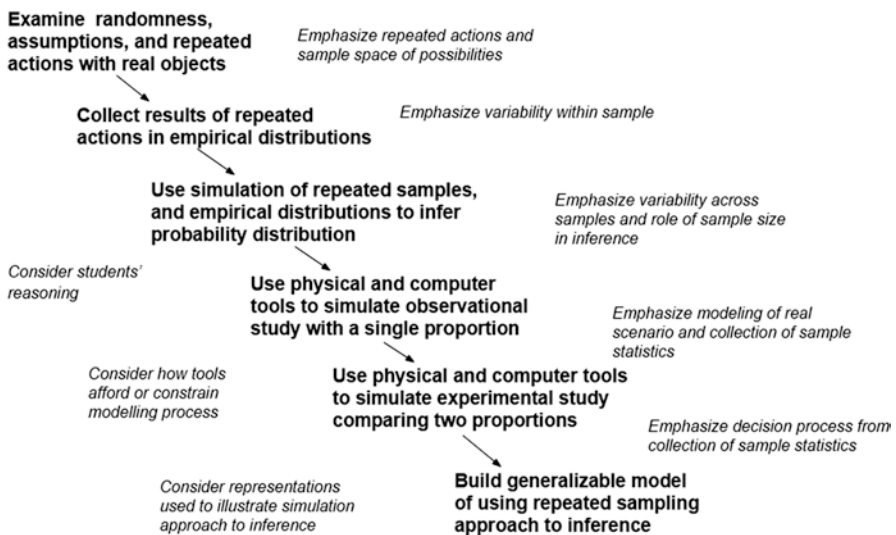

Fig. 9.8 Initial planned LT for a repeated sampling approach to inference

9.3.3.3 The Learning Activities and the Observed Learning Process

Lee and Doerr's research goals were to (1) develop and test a sequence of tasks in a LT that could achieve their learning goals for a particular group of adult learners and (2) identify key conceptualizations that seem to afford a stronger development of a generalized model of repeated sampling approach to inference. The approaches used in DBR (Bakker & van Eerde, 2015), their understanding of the literature on probability models and repeated sampling approaches to inference, and the representations and activities used by others (e.g., Lee, Angotti, & Tarr, 2010) informed their design of the LT (cf. Fig. 9.1(2)). The plan for the initial LT was designed during the 4 months before the course began and then revised during the first 7 weeks of the course as they got to know their learners. The course was taught by 4 instructors (led by Lee and Doerr) over 15 weeks in a once-a-week 3-hour meeting format to 27 teacher participants across 2 institutions.

What follows is a description of the LT at the point where teachers are comparing two proportions, the fifth task, and the consequent adjustments made to the LT based on their ongoing analysis of their learners' successes and struggles.

For the *fifth task*, they wanted teachers to apply their developing repeated sampling model for understanding the likelihood of a single proportion to the comparison of two proportions from an experimental design study (see fourth and fifth bolded goal in the initial LT in Fig. 9.8). They modified the *Dolphin Therapy* task (Catalysts for Change, 2012) to ask teachers to create a by-hand simulation using index *cards* that would answer the question: Can swimming with dolphins be therapeutic for patients suffering from depression? In the experiment, in the dolphin-swimming group (treatment), 10/15 patients improved their depression, while 3/15 improved in the control group. The question is whether that result indicates that swimming with dolphins is therapeutic for depression. The teachers were given 30 index cards marked with results from the study (13 cards marked "YES" for those benefiting with swimming with dolphins, and 17 cards marked "NO").

Lee and Doerr anticipated that how to conceive the random assignment in groups as a repeatable action would not be obvious, an important consideration when designing a LT. A variety of methods were created by teachers. After the discussion to draw out the importance of the assumptions of random assignment and that a patient's outcome does not change regardless of group assigned, the class eventually agreed to shuffle the cards representing the 30 patient outcomes and deal cards into 2 groups of 15. By repeating this action and computing the difference in proportion of YESs, they could examine a distribution of the difference in proportions on a shared class dot plot and consider how likely it is that the benefits of therapy reported in the original study happened by chance alone.

The *Dolphin Therapy* hands-on experience was followed by a *sixth task* that was another model exploration activity where the sampling distribution was explored again in *Statkey* (Lock, Lock, Morgan, Lock, & Lock, 2013) and *TinkerPlots* (Konold & Miller, 2005). Many of their teachers seemed to struggle with the multi-tiered process involved in doing a simulation through repeated sampling for this comparing proportions task. It was sometimes difficult for them to keep in mind all

the steps of the process that were happening in the computer. They also struggled with interpreting the sampling distribution in terms of how to use it to make an inference. The *seventh task* provided an opportunity for teachers to further explore the structure of their developing models by reading two articles (Lane-Getaz, 2006; Lee, Starling, & Gonzalez, 2014) in which diagrams were used to illustrate the simulation approach.

In the weekly team meetings, the four instructors (including Lee and Doerr) discussed the teachers' struggles with the repeated sampling approaches used in the two simulation tasks. They were not convinced that their learners had developed a general model for how to use a simulation approach to inference that they could apply to other situations and use for teaching students to use such an approach. Thus, they designed a new *eighth task* to allow teachers an opportunity to express their developing conceptions of the simulation process in terms of how they would help students understand the process. They considered that this task was an opportunity for teachers to explore their representations of the structure of models of repeated sampling for drawing inferences that would serve a pedagogical purpose. That is, the intended audience for this representation would be the future students of the teachers, and this representation hence served a perceived purpose of explaining the structure of models of repeated sampling to other learners. Teachers worked in small groups to do the following:

> Suppose you were going to use a repeated sampling approach with your students to help them use a simulation (with physical objects or computer models) to investigate if an observed statistic is likely or unlikely to occur. Draw a diagram you could use to help students understand the general process used for applying randomization techniques for solving these types of tasks.

Both during class and in the post-class analysis, the instructors noticed the wide variety of representations expressed in teachers' diagrams. Many teachers expressed some aspect of the modeling process from the real-world problem (though not always explicit) and that a collection of statistics is used for examining likelihood; however, their diagrams were much less explicit about the "randomize and repeat" phases in a simulation approach (e.g., see sample diagrams in Fig. 9.9).

Lee and Doerr's analysis of teachers' diagrams and the classroom conversations led to the design of an additional *ninth task* that was structurally similar to the *Dolphin Therapy* task but required an adaptation of their previous model since it involved comparing means for two unequally sized groups. In addition, they deliberately changed the form of the manipulatives (using unmarked flat wooden craft sticks rather than pre-marked index cards) to further push the learners in understanding the role of randomization in their model of repeated sampling. The teachers had varied approaches to recognizing what the repeatable action was in the scenario. Many used the craft sticks in some way, with slight variations from each other, to indicate scores and repeatedly reassigning those scores into two different unequal sized groups. Some teachers really struggled and did not create viable ways of representing the scores or reassignment to groups. Their attempts at applying their
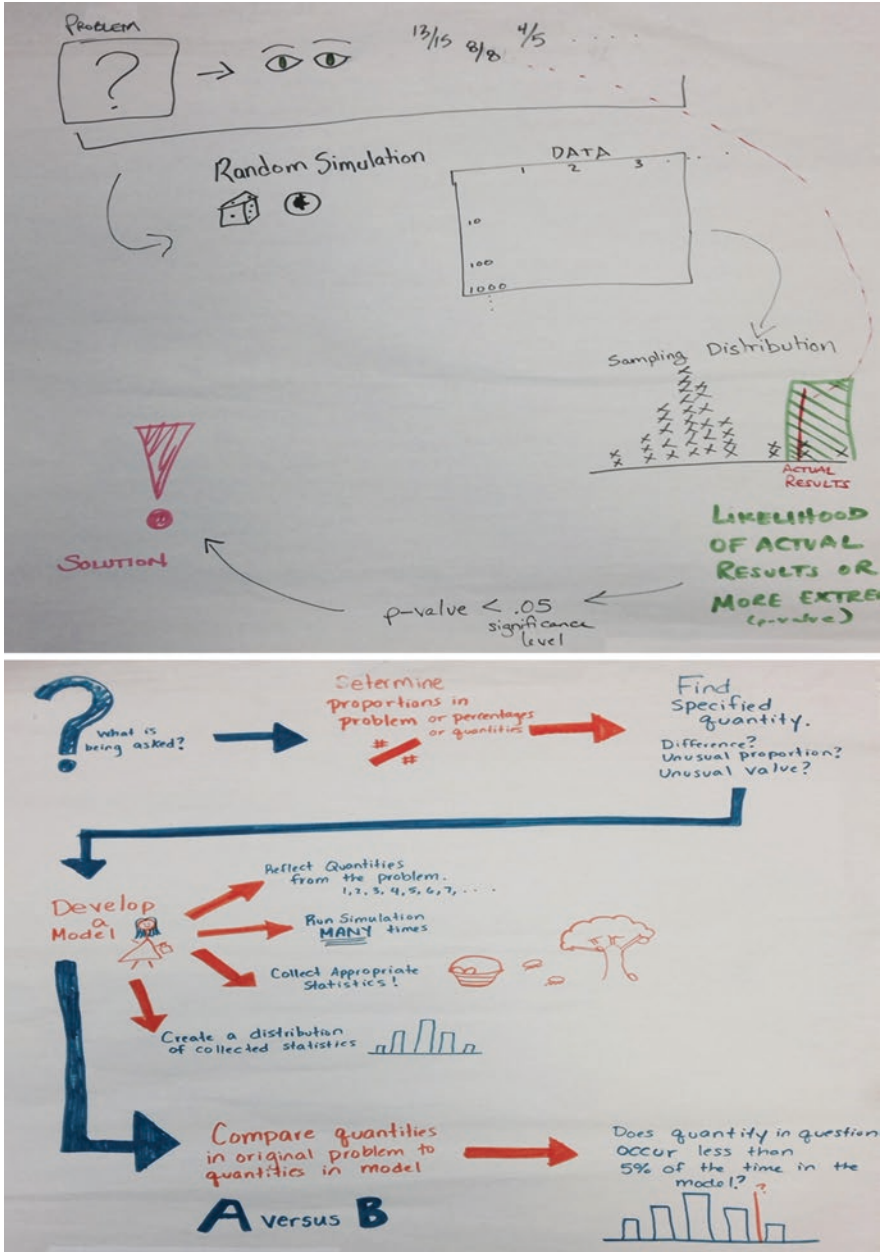
Fig. 9.9 Two samples of teachers' diagrams

model for a repeated sampling approach to inference to create this simulation in such a different context really illuminated the fragility of their models and conceptual understanding.

9.3.3.4 Discussion and Future Recommendations

This case illustrates how an ongoing analysis and instructional experiences impact the development of instructional tasks hypothesized as needed to assist learners in further developing the intended learning goals. Retrospective analysis of learners' work also can be used to modify a LT, in this case for using a simulation approach to inference. This analysis led to a realization that more attention needs to be given to the modeling process, the explicit role of probability in inference, and use of probability language. There is a two-part modeling process that should be made explicit. The first is to create a local specific *model of* the real-world context in statistical terms. The second is creating a simulation process that models the repeatable actions in the original problem and can be used to generate random samples. Most previous works have combined these two aspects into a single "model" or "population" level. There seems to also be a need to be more explicit concerning building a distribution of sample statistics, *viewing the distribution as an empirical probability distribution*, using the distribution to reason about the observed statistic, and making a claim about the chance of that observed statistic occurring. Lee, Doerr, Tran, and Lovett (2016) elaborate on these suggestions. It is important to recall that learners in this case had previous exposure and experience with learning traditional inference techniques, and some had experiences in teaching such techniques. There were only two who had previous experience in using a repeated sampling approach in their own curriculum materials with their students. Thus, the initial LT and sequence of tasks were designed with these learners in mind (cf. Fig. 9.1(3)). Researchers and teachers working with learners first engaging with inference through repeated sampling will need to adapt and adjust the LT as needed.

The LT discussed in this case study demonstrates how LTs are useful for identifying and exploring learners' reasoning processes, building new conceptual approaches for learning statistics, contributing to the research knowledge base, and directing the focus of future research.

9.4 Conclusion

LTs have been critical in the development of statistics education research and in enhancing students' learning in the classroom. LTs are not just a sequence of lessons; rather they are deliberately planned and modified based on careful analyses of the research literature, the web of concepts underpinning the learning goal, and the student responses. This chapter has focused on researchers using LTs, but we recommend that teachers, as action researchers in their own classroom, use LTs to

understand and improve their students' learning. Additionally, we recommend that teachers co-design LTs with other teachers to reflect the intentions of their curriculum and the realities of their classrooms (see Chap. 16). Co-designing LTs with researchers is also a possibility. We now reflect on what we can learn from the case studies and then propose four recommendations.

9.4.1   Reflection on the Case Studies

The LTs in the three case studies shared many commonalities. At a meta-level, all shared LTs that combined, in an interactive process, curriculum development and research and sequences of tasks and supporting students' thinking and performance. Furthermore, there was collaboration with teachers in classroom settings reflecting a participationist research paradigm (Sfard, 2005). Differences existed depending on the purpose of the research, the existing research literature, and how many cycles of teaching experiments were implemented. All the case studies, however, reflected the LT iterative process outlined in Fig. 9.1 and the components necessary to inform its design.

   All the studies started with a problem. Case Study 1 sought to model students' understanding of variability over time. Case Study 2 had a defined goal of making the call when comparing two box plots and then ascertained the myriad of concepts that underpinned making a judgment under uncertainty. Case Study 3 began with the researchers' knowledge of the literature on statistics, probability, and modeling and their belief that teachers needed to conceptualize the links among them into a general model. In line with the other two studies, Case Study 3 developed a hypothetical learning process that aimed to scaffold teachers' thinking toward a general model realization about repeated sampling for making an inference. Case Studies 1 and 3 drew on some existing learning activities for their LTs, whereas Case Study 2 invented its own. Whether inventing new tasks for LTs or not, all attended to delineating the statistical big ideas and concepts underpinning the learning goal and strived to engage students in the LT's defined abstract notions using innovative learning approaches. During teaching, as students engaged with the learning tasks, their actions, representations, and thinking were observed and analyzed. Consequently there was a feeding forward and back into the LTs, which were modified and altered from the planned LT. Case Study 3 illustrated the importance of a retrospective analysis whereby the researchers, in response to the teachers' fragile understanding of models for repeated sampling in inference, proposed some new key conceptualizations.

   Compared to research that gauges levels and types of thinking based on survey questions or explicating students' thought processes when engaging with several tasks, research that uses LTs and DBR methodology has the potential to have more impact on learning in classrooms as Case Studies 1 and 2 show (see Chap. 16 also). While acknowledging that the findings from the former type of research are vital for the designing of LTs, the latter type of research is also good at identifying gaps in

students' thinking and new avenues to explore (e.g., Case Study 3). A LT can be just one lesson or cover many lessons, but as these studies illustrated, statistical big ideas and concepts take time to experience and take root in students' cognitive infrastructure.

In a critique of LTs used in research, Baroody, Cibulskis, Lai, and Li (2004) believed some of them were overly prescriptive and detailed and consequently an inquiry-based investigative approach was lost. They conjectured that LTs "could be more comprehensible and useful to practitioners if they focused on how big ideas evolve" (p. 253). These case studies did focus on the big ideas and how these might evolve at particular levels, but there is a danger that microanalyses of students' thinking, while important to research, may lead to a plethora of types and levels of reasoning resulting in researchers and teachers using step-by-step procedures in LTs to achieve the learning goals. When designing LTs, an important criterion to consider is the degree of openness permitted in the learning process so as not to lose the investigative spirit inherent in the statistical enterprise and the process of inquiry that is central to statistical thinking and learning (cf. Chap. 10).

The statistical inquiry investigative cycle is the centerpiece of some new curricula (e.g., Ministry of Education, 2007) with students learning how to be "data detectives." As part of enculturating (Garfield & Ben-Zvi, 2008) students into statistical thinking and inquiry (see Chaps. 4 and 7), the development of concepts is essential as well as the development of coherent conceptual infrastructure across the curricula levels. These LTs illustrated how conceptual understanding might be built up in students and teachers. However, researchers may need to remind themselves not to lose sight of the big ideas and the inquiry-based investigative approach when designing LTs. That is, there is a balance between concept-focused and inquiry-based LTs.

9.4.2   Recommendations and Implications

We have four recommendations for future research regarding LTs:

1. Continue exploratory research on LTs of specific topics in statistics.
2. Scale LTs to many diverse classrooms.
3. Build coherent conceptual pathways across curricula and grade levels.
4. Attend to analysis of web of concepts, task design, and methods of data analysis.

Much of the research using LTs has been within one topic domain at one curriculum level with a few groups of students. As Case Studies 2 and 3 showed, exploratory research with one group of students that either treads into new territory or investigates a concept from a new angle can provide invaluable insights into garnering understanding about teaching and learning processes. These small-scale studies can facilitate the generation of more refined local theories about teaching and learning certain topics in statistics. Thus, our *first recommendation* is that researchers

continue using LTs in their research as they have enormous potential to explore and identify interesting phenomena and to develop theories about learning.

The *second recommendation*, which Case Study 1 attempted to address, is scalability to many classrooms. The challenge for Case Study 1 was accurately capturing typical responses, describing them in terms of increasing levels of sophistication, and communicating these ideas effectively to teachers. Another open question was how to make teachers aware of LT research results so that they could anticipate the possible student ideas and challenges, provide opportunities for ideas to emerge, and then use data on student learning to support continued progress in learning. Such challenges and questions will need to be addressed when expanding successful LTs to a broad range of classrooms. When LTs are considered to have the potential to be shared, we recommend that researchers think about collaborating in new research projects to address how to manage implementation on a larger scale. Where necessary, researchers may need to alter their LTs in response to new findings as a result of more people such as curriculum developers, professional development facilitators, and teachers being involved in the implementation (e.g., Lehrer et al., 2014).

The *third recommendation* is building curriculum coherence for teachers and students across the grade levels. What is needed is a major collaboration of researchers worldwide to work out the big ideas and web of concepts that have been researched and where more research is needed (e.g., covariation). They could then attempt to map across the curriculum the main conceptual pathways and identify the LTs that exist and may be used given the time constraints of curricula. We recommend, as a start, that researchers using LTs could devise and research a pathway for growing students' knowledge and thinking in one topic domain from grades 1 to 12 in a similar vein to Case Study 1 with its learning maps, relational learning clusters, and big ideas for grade 6. Perceiving across the curriculum, an evolving conceptual pathway together with LTs toward a big idea could be useful for curriculum developers and for the research community.

In Sect. 9.2 we identified three aspects regarding the design and use of LTs that seemed to need more attention in research. Hence, our *fourth recommendation* is that researchers conduct more in-depth analyses of the web of concepts underpinning their learning goal, carefully consider the literature on task design and the influence the task will have on students' learning, and devise more transparent ways of analyzing data gathered and providing evidence, particularly for classroom interactions. Also meta-LT research is required to study LTs as a methodological tool. Addressing these issues, which seem to be currently missing in statistics education research using LTs, would move the field forward.

In statistics education research, the use of LTs as an instrument in DBR has resulted in a fecund route for learning about students' thinking and has opened up many new challenges and avenues for future research. As technology changes approaches to learning, there is now an even greater need to focus on the big ideas and concepts that will endure despite those changes. We believe that using LTs and DBR will continue to provide a fruitful and rewarding pathway for future researchers.

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Arnold, P. (2013). Statistical investigative questions – An enquiry into posing and answering investigative questions from existing data. (Doctoral thesis). Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21305

Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education, 19*(2), 1–32.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: Freudenthal Institute.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.

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.), *Approaches to qualitative research in mathematics education* (pp. 429–466). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-017-9181-6_16

Baroody, A., Cibulskis, M., Lai, M., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning, 6*(2), 227–260. https://doi.org/10.1207/s153277833mtl0602_8

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42–63.

Ben-Zvi, D., & Garfield, J. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397–409). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*(3), 325–341. https://doi.org/10.1016/j.learninstruc.2004.06.006

Catalysts for Change. (2012). *Statistical thinking: A simulation approach to modeling uncertainty*. Minnesota, MN: Catalyst Press.

CensusAtSchool New Zealand. (2003). Retrieved from www.censusatschool.org.nz

Clements, D., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81–89.

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

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics. (College- and Career-Readiness Standards and K12 standards in English Language Arts and Math)*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org

Confrey, J. (1991). Learning to listen: A student's understanding of powers of ten. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 111–138). Dordrecht, The Netherlands: Kluwer Academic Press.

Confrey, J. (2002). Sixth grade pre-algebra curriculum at St. Francis School, Austin, TX. Unpublished document.

Confrey, J. (2006). The evolution of design studies as methodology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 135–151). Cambridge, UK: Cambridge University Press.

Confrey, J. (2015). Some possible implications of data-intensive research in education—The value of learning maps and evidence-centered design of assessment to educational data mining. In C. Dede (Ed.), *Data-intensive research in education: Current work and next steps. Report on two National Science Foundation-sponsored Computing Research Association workshops* (pp. 79–87). Washington, DC: Computing Research Association.

Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Design-Based Implementation Research. (2016). Retrieved from http://learndbir.org/

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1&2), 132–151. https://doi.org/10.1080/10986065.2011.538294

Finzer, B., Konold, C. & Erickson, T. (2012). Data games. Retrieved from http://play.ccssgames.com

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Alexandria, VA: American Statistical Association.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92–99.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education, 44*(7), 883–898.

Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning, 6*(2), 105–128.

Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. van den Akker (Ed.), *Educational design research* (pp. 17–51). London: Routledge.

Konold, C., & Miller, C. (2005). *TinkerPlots dynamic data exploration: Statistics software for middle school curricula*. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Lane-Getaz, S. J. (2006). What is statistical thinking and how is it developed? In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth NCTM yearbook* (pp. 272–289). Reston, VA: National Council of Teachers of Mathematics.

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

Lee, H. S., Doerr, H. M., Tran, D., & Lovett, J. N. (2016). The role of probability in developing learners' models of simulation approaches to inference. *Statistics Education Research Journal, 15*(2), 216–238.

Lee, H. S., Starling, T. T., & Gonzalez, M. D. (2014). Connecting research to practice: Using data to motivate empirical sampling distributions. *Mathematics Teacher, 107*(6), 465–469.

Lehrer, R. (2016). Data modeling. Retrieved from http://modelingdata.org/

Lehrer, R., & Kim, M. J. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*(2), 116–133.

Lehrer, R., Kim, M. J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In A. Maloney, J. Confrey, & K. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 32–59). Charlotte, NC: Information Age Publishers.

Lehrer, R., Kim, M. J., & Jones, R. S. (2011). Developing conceptions of statistics by designing measures of distribution. *ZDM Mathematics Education, 43*(5), 723–736.

Lehrer, R., & Schauble, L. (Eds.) (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York, NY: Teachers College Press.

Lesh, R. A., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lock, R., Lock, P., Morgan, K., Lock, E., & Lock, D. (2013). StatKey. Retrieved from http://lock5stat.com/statkey

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.

Makar, K., & Confrey, J. (2005). Variation talk: Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

Makar, K., & Rubin, A. (2009). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 8*(1), 82–105.

Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131–156.

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2*(3), 149–167.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Dordrecht, The Netherlands: Kluwer.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes – An overview on achievements and challenges. *ZDM Mathematics Education, 47*(6), 877–891.

Radford, L. (2009). Why do gestures matter? Sensuous cognition and the palpability for mathematical meanings. *Educational Studies in Mathematics, 70*(2), 111–126.

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics, Dunedin, New Zealand* (Vol. 2, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A9-4.pdf

Saldanha, L., & Liu, Y. (2014). Challenges of developing coherent probabilistic reasoning: Rethinking randomness and probability from a stochastic perspective. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 367–396). Dordrecht, The Netherlands: Springer.

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Sfard, A. (2005). What could be more practical than good research? On mutual relations between research and practice of mathematics education. *Educational Studies in Mathematics Education, 58*(3), 393–413.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.

Watson, A., & Ohtani, M. (Eds.). (2015). *Task design in mathematics education*. New York, NY: Springer.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2), 247–295.