

SAMPLING IN THE WILD: HOW ATTENTION TO VARIATION SUPPORTS MIDDLE SCHOOL STUDENTS' SAMPLING PRACTICE

MICHELLE E. FORSYTHE

*Department of Curriculum and Instruction
Texas State University
mforsythe@txstate.edu*

ABSTRACT

Sampling is a fundamental practice of many scientific disciplines. However, K–12 students are rarely asked to think critically about sampling decisions. Because of this, open questions remain about how best to support students in this practice. This study explores the emergent sampling practice of two classes of sixth-grade students as they investigate the ecology of a local creek. It draws on student interviews, pre/post-tests, student artifacts, and video recordings of classroom activity to identify and trace shifts in the ways in which students approached collecting data. The findings suggest three ways in which students' attention to variation within the context of their ecological investigations supported their development of a more sophisticated practice of sampling.

Keywords: *Statistics education research; Science education; Sampling; Ecology; Variation*

1. INTRODUCTION

In science, claims are made and evaluated in light of how data are constructed. This focus on how data are built has led many scientific disciplines, including the field of ecology, to develop sophisticated practices for collecting and analyzing data—practices of observation, measure construction, sampling, and representation (Coe, 2008; Eberhardt & Thomas, 1991; Kenkel, Juhász-Nagy, & Podani, 1990). Current frameworks for science education advocate for K–12 students to develop scientific literacy by engaging in these investigative practices in learning contexts that are personally meaningful to students (e.g., National Curriculum Board, 2009; National Research Council, 2012). Frameworks for statistics education, such as the *Guidelines for Assessment and Instruction in Statistical Education (GAISE) Report*, offer parallel recommendations that student-driven questions and data collection be used to foster statistical literacy (Franklin et al., 2007).

Although approximations of scientific practice and the foundations of statistical concepts are developmentally accessible to students (Franklin et al., 2007; National Research Council, 2012), young students are rarely invited to grapple with the complexities of practices such as sampling when conducting their own scientific investigations. Explorations of sampling within observational investigations, which characterize much of the early science curricula, have been especially overlooked. In the United States, none of the K–8 performance expectations in the *Next Generation Science Standards* (NGSS Lead States, 2013) explicitly engage students in making sampling decisions. This gap in practice raises the question, what initial aspects of sampling might emerge as meaningful to students as they conduct observational investigations? In response to this question, this paper reports on a design study which supported middle school students in conducting their own observational field investigations as they sought to understand the ecology of a local creek. In particular, it focuses on how ideas about variation, a fundamental facet of ecological research, played out in students' emergent approaches to sampling.

2. THEORETICAL FRAMEWORK

2.1. SAMPLING—A STATISTICAL CONCEPT AND A SCIENTIFIC PRACTICE

A conceptual understanding of samples and sampling is a fundamental aspect of statistics education (Franklin et al., 2007). A statistically literate student should be able to explain how the results of different types of studies and polls come from a sample and be able to critique those results based on the way in which the sample was created. In doing so, the student must recognize that a sample is not a smaller piece that mirrors the whole population, but rather a proportional piece with the potential to take on a variety of outcomes. This understanding simultaneously balances the ideas of a sample as representative of the population with the idea of a sample as variable (Rubin, Bruce, & Tenney, 1991; Saldanha & Thompson, 2002). Decisions about a sampling design, such as sample size and selection, are thus grounded in how variability and representativeness play out in the specific context of any given study.

In science education, sampling is often positioned less as a concept and more as a part of the practice of science. This construct of scientific practice has primarily been informed by studies of how scientists conduct their own work. These studies have elucidated how scientists form an epistemic community with specific means of legitimizing how knowledge is constructed (Knorr-Cetina, 2009; Latour & Woolgar, 1979; Lave & Wenger, 1991; Nersessian, 2008). The activity within this community is orchestrated by historically developed social norms and arrangements of performances that shape both the individual and the communal scientific search for a better explanation of nature (Ford, 2015). Within this community, practices such as sampling emerge as stabilized interactions of performances that have been evaluated, critiqued, and reified as appropriate means, at least for the present, to conduct this search. As knowledge and norms evolve, what the community counts as appropriate scientific practice also shifts.

For the past decade, science education has increasingly advocated for student participation in this practice of science (Ford, 2015; Mody, 2015; Stroupe, 2015). Although the practice shift in science education emphasizes engaging K–12 students in meaningful approximations of the disciplinary work of scientists (Ford & Forman, 2006; Kelly, 2011; Osborne, 2014; Richard & Bader, 2010), the goal is not for students to learn how to replicate current stabilizations of disciplinary performances. Nor is it for practices to merely serve as a means of promoting the development of core content knowledge. Rather, the objective is for students to learn how to progressively refine both their explanations of nature and the ways in which they develop those explanations. Though students' evolution of practice can be informed by the norms of how the larger scientific community has itself stabilized practice (Duschl, 2008), the goal is to create a local, meaningful need for change in a particular practice, rather than presenting disciplinary conventions a priori to students (Manz, 2012, 2014). Although every branch of science engages in similar disciplinary work, each is also characterized by subject-specific forms of practice. Because of this, science educators have to make decisions about which forms of scientific practice to emphasize in K–12 learning (Stroupe, 2015). Research studies (e.g., Berland et al., 2015; Manz, 2014; Nielsen, 2011; Sandoval & Reiser, 2004; Schwarz et al., 2009; Svoboda & Passmore, 2011) have detailed, nuanced ways to support the practices of modeling, explanation, and argumentation. However, only a few science educators (e.g., Lehrer & Schauble 2012, 2017; Metz 1999) have explicitly examined how to support the development of students' sampling practice.

2.2. THE ECOLOGICAL PRACTICE OF SAMPLING

One of the dilemmas scientists face when conducting research in field settings is that it is impossible to collect data exhaustive of the system. Scientists must therefore make decisions about where and when to measure. These methodological issues of time and space are essentially questions about the epistemic practice of sampling: where should plots be set, what size should they be, how often should they be checked, and so forth. (Coe, 2008; Eberhardt & Thomas, 1991). Although sampling is a practice inherent to many field sciences, this paper focuses on how sampling is enacted within the domain of ecology. Ecology is the study of the characteristics, abundance, and distribution of living organisms and the

relationships within and between these organisms and their environments (Korfiatis & Tunnicliffe, 2012). Like many scientists, ecologists utilize laboratory experiments and modeling environments in their studies; however, a preponderance of ecological research is conducted in field settings (Eberhardt & Thomas, 1991; Korfiatis & Tunnicliffe, 2012; Lefkaditou, Korfiatis, & Hovardas, 2014).

Mörsdorf et al. (2015) articulated the feelings of many of their ecological colleagues when they bluntly stated, “Sampling in ecology can be challenging” (p. 1). This difficulty stems in part from the inherent complexities of field settings. However, Kenkel et al. (1990) explain that this difficulty also stems from how ecological studies frequently pursue different objectives than those emphasized in classical statistical sampling theory. Some ecological investigations are concerned with estimating parameters of populations with discrete sampling units, such as the mean tree height, and these studies are similar to those that ground sampling theory in statistics. However, other ecological investigations focus on uncovering patterns of distribution or variation in continuous settings, such as clumped patterns of floral diversity. In these type of studies, ecologists often make methodological decisions that purposefully maximize variation between samples or create arbitrarily defined sampling units, issues not readily addressed by classical sampling approaches. Though ecologists have to struggle to manage and interpret variation in each type of investigation, their response to that struggle can differ from study to study. Thus, the appropriateness of any one sampling procedure is contextually dependent on the study objectives, variables measured, and characteristics of the specific ecological setting.

Ecologists’ struggle with variation has been influential in their disciplinary evolution of the practice of sampling. In the premier issue of *Ecology*, the first ecological journal in the United States, ecologists described their field settings in narrative form but either did not detail how they selected units to measure within that setting (e.g., Hofmann, 1920; Praeger, 1920; Wherry, 1920) or else used a convenience approach and selected the most readily accessible units (e.g., Douglass, 1920; Esterly, 1920). Ecologists’ attention to how they were selecting units for analysis seemed to emerge as a response to the conflicting knowledge claims that resulted from unexpected variations in data. For example, Esterly (1920) noted that some of the anomalies in his findings might have stemmed from what he had originally thought to be inconsequential differences in how he collected his samples. Over time, the need to minimize the bias in data caused by variations across known gradients gave rise to systematic forms of dividing up space and time (e.g., DeWoskin, 1980; Ewald, Hunt, & Warner, 1980; McClure, 1980; Rogers, 1980; Stephenson, 1980; Tobiessen & Werner, 1980). In addition, randomization began to take hold as a way to reduce bias from unknown gradients of variability. Though contemporary ecologists still justify aspects of sampling based on either convenience or purposive consideration of the phenomenon, most also apply some form of systematic or randomized approach to location, timing, or unit subdivision (e.g., Alberto et al., 2010; Biswas & Mallik, 2010; Bridgeland, Beier, Kolb, & Whitham, 2010; McLellan, Serrouya, Wittmer, & Boutin, 2010; Patterson, McConnell, Fedak, Bravington, & Hindell, 2010; Ravet, Brett, & Arhonditsis, 2010).

Sampling methods permeate the ecological literature, offering general procedures for sampling everything from fuel loading in forests to plant diversity (Bacaro et al., 2015; Sikkink & Keane, 2008). This literature serves as a key social resource ecologists use to construct initial sampling plans. However, when ecologists try to enact their initial sampling plans in the field, these plans become problematized by unforeseen complexities, such as spatial and temporal variation (Latour, 1999; Lorimer, 2008; Roth & Bowen, 2001). As their initially fixed protocols become more nuanced and flexible, ecologists struggle to balance the need to adapt their sampling to the local context with the need to preserve the social normality of their approach. Because of this, when ecologists want to shift the normative disciplinary approach to sampling, they often design studies that specifically argue how different sampling protocols generate different findings (e.g., Bacaro et al., 2015; Kenkel et al., 1990; Mörsdorf et al., 2015; Schweiger, Irl, Steinbauer, Dengler, & Beierkuhnlein, 2016; Sikkink & Keane, 2008). Further social distribution and discussion of these methodological studies in formal and informal settings allows for the practice of sampling to evolve in the larger disciplinary community through progressive evaluation and critique.

2.3. STUDENTS, SAMPLING, AND FIELD ECOLOGY

Sampling is a potentially powerful practice for supporting students' understanding of science because it foregrounds how a data set is constructed. It highlights how different scientific studies might produce different findings, how clear forms of communication are essential within the scientific community, and how the objectives of scientific research permeate investigative and interpretive decisions (Kenkel et al., 1990; Mörsdorf et al., 2015). As such, sampling is foundational in helping students “understand the conclusions from scientific investigations and offer an informed opinion about the legitimacy of the reported results” (Franklin et al., 2007, p. 3). Curriculum designs in science, however, rarely invite K–12 students to wrestle with how sampling methods impact the data collected and the claims drawn from that data. Rather, typical pedagogical approaches often dictate sampling procedures or have students arbitrarily select protocols (e.g., Council for Environmental Education, 2006). Many of these approaches undermine the complexity of the practice by assuming the reliability of small samples and overlooking issues of variability. Even relatively sophisticated curricula, such as Stier's (2010) explorations of sampling and bias, at times simply promote randomization rather than explore the relationship between a study's question, context, and sampling design.

The practice of sampling has similarly been overlooked in most science education research, with only a few studies beginning to tease apart how to support students' sampling practice in ecology. Lehrer and Schauble (2012) found that when middle school students engaged in ecological field investigations they most often initially focused on collecting as much of something as possible, no matter their research question. If they mimed any complex sampling practice, such as replication, it was to “double-check” their answer or ensure that they had not missed anything. When Metz (1999) explored sampling with elementary and middle school students, less than half of students used ideas such as sample size, replication, or stratification during post-interviews to critique their studies of plant growth and animal behavior. Most still insisted that they needed to test every member of a population to be confident of their findings, particularly in contexts with variability. Neither of these studies, however, specifically explored progressive shifts in students' sampling practice as they conducted their investigations. More recently, Lehrer and Schauble (2017) conducted a follow-up study with sixth-grade students who had conducted a year-long investigation of a local pond. For these students, a representative sample of the pond had to account for the various strata present (e.g., shallow and deep water). However, whereas most students recognized that observations within a given strata would vary, only a few attributed this variability to chance.

For professional ecologists, encounters with variability were crucial to the disciplinary evolution of sampling (e.g., Esterly, 1920). It may be that similar experiences could provoke the development of students' sampling practice. Variation is an inherent part of ecological fieldwork. As soon as students step into the field, they are confronted by variability—even if they only perceive it on a gross level. Wildflowers might clump in one area and grasses in another. The currents in a creek might be constantly shifting. Students can easily notice these differences. Uncovering the sources contributing to these differences, however, is more complicated. Consider a student who observes that some clumps of grass in a field are taller than others. Both random natural variability and directed causal processes (induced variability) have contributed to this variation in perceived height. Should the student choose to explore this phenomenon by measuring the height of the grass, this introduces another source of variation as natural variation, causal processes, and now measurement error would all be contributing to variability in the measured height of the grass clumps. Sorting out the source of differences, be they from measurement variability, natural variability, or induced variability motivates both ecological and statistical endeavors (Franklin et al., 2007). Studies have shown that middle school students can successfully reason about distributions of variation due to repeated measures as well as causal forms of variability (Lehrer & Kim, 2009; Lehrer & Schauble, 2004, 2017; Petrosino, Lehrer, & Schauble, 2003). These ideas about variability have the potential to support students in making sense of ecosystem processes. However, students' ideas about variability often remain disconnected from their approaches to collecting data and thus fail to be translated into action when designing scientific investigations. Because of this, many

questions remain about how to structure learning environments that might support students' sampling practice in science.

Some of the solutions science educators are seeking can be found in the inroads statistics educators have made in advancing students' understanding of sampling and variability. Statistical reasoning is grounded in an understanding of variation (Moore, 1990). Because of this, Shaughnessy and Pfannkuch (2002) have advocated that educators encourage students to look at data through a "variation lens" (p. 256). Students of all ages have an intuitive sense and expectation of variability when working with familiar contexts (Watson, 2009). This has the potential to be capitalized on in the design of learning environments. However, because much of the statistics curriculum has traditionally emphasized center over variability, students exposed to substantial instruction tend to rely on centers when predicting distribution rather than incorporating estimates of both center and variability (Noll & Shaughnessy, 2012). The tendency is particularly strong when students are making predictions from known populations.

These studies in statistics education suggest that there might be value in rooting students' initial explorations of sampling in familiar contexts that have strong patterns of variability but unknown underlying distributions. This would allow students to leverage their intuitive expectations of variability, while at the same time creating a need to look for patterns in the data. This perception of a need is fundamental in helping students shift from relying on their own personal beliefs about a phenomenon to relying on what the data say (Shaughnessy & Pfannkuch, 2002). In this, it is vital that students build substantial familiarity with the specific context in order to recruit their intuitive resources about variability. Without this, students often struggle to negotiate multiple sources of variability and relate these sources to the context (Metz, 1999; Pfannkuch, 2008; Watson & Kelly, 2002). In such instances, students often construct causal stories to explain away random variation, especially in contexts about which they have strong initial beliefs but little experience collecting data (Wroughton, McGowan, Weiss, & Cope, 2013).

However, familiarity alone is not enough. Probabilistic approaches grounded in familiar but not personally meaningful contexts, such as flipping coins or drawing candy, seem to work against students' initial resources for making sense of variation, especially when these approaches lead with measures of center. For example, Reading and Shaughnessy (2004) interviewed students about the number of red candies likely to be found in six handfuls of ten candies drawn, with replacement, from jars with different color proportions. When students simulated this experiment, they tried to explain away the variation in their results, especially if these results were disconfirming to their original predictions, by postulating causal relationships between the set of numbers generated and variables such as how well the candies were mixed. Likewise, Sharma (2003) found that when asked whether someone who tossed a coin ten times or someone who tossed a coin 50 times was more likely to get 80% or more heads, students did not attend to the relationship between sample size and variability. Rather, they used personal experience to reason causally about how one's actions could influence whether the coin came up heads or tails. In making sense of these findings, Sharma argued that when students invoke relevant background knowledge, such as familiarity with a specific context or understanding of other curricular areas such as physics, to support their reasoning about statistics, students often undermine the probabilistic basis of the very problems they are trying to solve. In a similar interview setting, Shaughnessy, Ciancetta, and Canada (2004) found that students acknowledged variation between repeated samples and could correctly identify both likely and surprising outcomes. However, the sample variability that students predicted was inappropriate given the population parameters of the task. Shaughnessy et al. suggested that students' devotion to expected outcomes and struggles with sample variability were likely amplified by recent classroom experiences focused on the probability of individual outcomes.

More promising methods for supporting students' emergent sampling are not only grounded in personally meaningful contexts, but also begin with direct experiences with distributions and variation. The results of experiments or computer-generated simulations can be used to foster productive discussions about the shape of data and sources of variability (e.g., Lehrer & Schauble, 2004; Stohl & Tarr, 2002; Torok & Watson, 2000). In such settings, students often have a deep contextual understanding of how the data were generated. This helps students to reason distributionally by providing them with a

sense for what is likely—the center—and what is possible—the variability (Pratt, Johnston-Wilder, Ainley, & Mason, 2008). As ecological field investigations offer contextually rich experiences in which students can physically experience both what is likely and what is possible when collecting data, they have the potential to support students’ emergent distributional thinking and advance students’ sampling practice.

3. METHODOLOGY

This paper reports on a case study embedded within a larger design study (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) investigating novel curricular supports for learning ecology. These supports centered on students’ emergent understanding of the ecology of a local creek. This creek was familiar to the students as it flowed through the center of their small rural town. However, none had investigated it from a scientific perspective. The overarching questions “What type of place is the creek?” and “How do different parts of the creek ecosystem interact?” guided students’ investigations.

3.1. RESEARCH QUESTIONS

The broader study from which this paper originates investigated the question: How does students’ scientific practice develop within the context of ecological fieldwork? However, this paper specifically explores the more focused research questions:

1. What initial aspects of sampling emerge as meaningful to middle school students during ecological field investigations?
2. How can attention to variation support middle school students’ development of the practice of sampling?

3.2. STUDY PARTICIPANTS

Two classes of sixth-grade students from a rural public middle school in the southern United States participated in this study. These two classes were taught by the same math/science and literacy/history teaching team. A total of 48 students (94%) consented to participate in the study, although all students joined in the learning activities. The work of two focus groups of four students each (one group from each class) was followed in more detail. The math/science teacher selected these eight students to be representative of the demographics of the two classes and span a diverse range of initial competency in science. All student names given in this article are pseudonyms.

The middle school in which this study took place served a student population that identified as 87.1% White, 9.4% Hispanic or Latinx, and 3.4% other races or ethnicities. Most of the students (61%) qualified for free or reduced student lunch, indicating low socio-economic status. Only a few students (7%) had limited English proficiency. In the year of this study, 77% of the sixth-grade students tested proficient or advanced on their state mathematics assessment. The prior year, 62% of this same population of students tested proficient or advanced on their fifth-grade state science assessment.

3.3. INSTRUCTIONAL DESIGN

This study took place over the equivalent of thirteen periods of science class. Each class session lasted 35-40 minutes for a total of approximately 8.5 hours of instruction. During the study, the students participated in three mini-cycles of investigation (Table 1) in which they formulated research questions and hypotheses, designed data collection plans, grappled with the materiality of the creek, and analyzed their findings. Each cycle incorporated opportunities for students to iteratively refine their practice based upon personal and collective experiences. Although ecological foci were built into the design, all instruction on sampling was student-driven and emerged in response to what students found salient about collecting and analyzing data. Once students had identified a need to attend to an aspect of sampling,

subsequent instruction was adapted to support their ideas. As this flexible form of instruction was novel to the math/science teacher, I served as the primary instructor during these science classes. However, the math/science teacher freely interacted with students throughout the study and often posed questions during small group work and whole class discussions.

Table 1. Instructional design for students' creek investigation

| | |
|--------------------------------|---|
| Cycle 1 December (2 hrs) | Day 1: Class Discussion – What type of place is the creek? Day 2: Small Group Planning of Data Collection Day 3: Creek Visit 1 ^a |
| Cycle 2 May (3.25 hrs) | Day 4: Class Discussion – How do different parts of the creek interact? Day 5: Small Group Planning for Revised Data Collection Day 6: Creek Visit 2 Day 7: Small Group Data Analysis Day 8: Class Research Meeting |
| Cycle 3 May (3.25 hrs) | Day 9: Small Group Planning for Revised Data Collection Day 10: Creek Visit 3 Day 11: Class Data Exploration Day 12: Small Group Data Analysis Day 13: Class Discussion – Could patterns we see be due to chance? |

^aWe had intended to include a day for data analysis and a day for a class research meeting during Cycle 1; however, a storm closed schools for an extended period of time.

Prior to the start of this study, the students had invented data displays and measures of center using data from repeated measures of an attribute as part of their math class, using activities similar to those described by Lehrer and Kim (2009) and Lehrer, Kim, and Jones (2011). In addition, when this study was conceptualized, the math/science teacher and I planned for additional statistical investigations of data modeling, similar to those described by Lehrer and Romberg (1996), throughout the spring term. However, local shifts in priorities and curricular changes forced the elimination of this element of instruction. Instead, during April and May the students received additional traditional instruction from the math/science teacher on measures of center (mean, median, mode), spread (range, inter-quartile range), and data displays (histograms, line graphs, box-plots). These lessons focused on procedural understanding and used decontextualized sets of data.

Design of Cycle 1: Days 1–3 The first cycle of investigation focused on familiarizing students with the setting of the creek and with participating in guided scientific inquiry. We began Day 1 with a discussion framed around the question “What type of a place is the creek?” and generated class lists of what we might see at the creek, what we might want to investigate, and what observations we might want to record. Students initially focused on “what” questions about the biotic life, such as “What types of fish are in the creek?,” and later broadened these to include different dimensions of the biotic life (e.g., size or number), different dimensions of the abiotic environment (e.g., water depth, water speed), and relationships between different elements (e.g., whether different organisms might be found in areas with different water depths).

On Day 2, the class broke into teams of 3–4 students and began planning their first visit to the creek. The teams were selected by the math/science teacher to maximize the diversity of ability within each team and minimize the likelihood of conflicts. Each team developed their own research question and their own plans for how they would collect and record data, including any measurement and sampling decisions. As students did not yet have a rich sense of the creek, the teams asked fairly simple questions, such as “How many fish are there?,” that did not position the variables in relation to each other or explore patterns across space.

On Day 3, the students took their first visit to the creek. Each team had access to a personalized selection of tools and equipment tailored to their data collection plans. Extra tools were on hand for

students to flexibly adjust their protocols when needed. After taking general observations to familiarize themselves with the creek, the teams worked independently to collect data in self-selected locations of the creek.

Design of Cycle 2: Days 4–8 The second cycle of investigation supported students' thinking about how different parts of the creek ecosystem interacted with each other. We began Day 4 by sharing our findings from our first creek visit. The students primarily highlighted lists and/or counts of organisms and general qualitative observations, with some teams adding in one or two abiotic measurements. During this discussion, I purposively brought into contact the observations of different teams who had investigated similar variables in different sections of the creek. When students found it difficult to picture where each team had worked, we created a map to share our data and used this to propose relationships that might exist between different elements of the creek.

On Day 5 the teams began planning their second creek investigations. Worksheet prompts scaffolded students to choose a pattern of covariation within two sections of the creek to explore. Although students were encouraged to apply ideas raised during the previous day's discussion to their plans, they were not provided with teacher-driven instructions. Rather the teams made all decisions about data collection, including sampling, based on their own ideas of how to improve their investigations. A few teams wished to investigate abiotic components, such as dissolved oxygen, for which they could not develop their own measures. I taught these teams standard protocols for measuring these components. However, the teams made all other data collection decisions.

On Day 6 we headed back to the creek. The protocol for this day mirrored our first visit. Because time was limited, we chose to scaffold students' data recording with a preformatted table that explicitly prompted students to record each of their observations. Though this inscription compelled students to attend to repeated observations, we introduced it only after students had expressed a need for this aspect of sampling.

During the last two days of this cycle, students participated in a class research meeting adapted from the format described by Lehrer, Schauble, and Lucas (2008). On Day 7, each team summarized their data and prepared what they wanted to present to the class about their research question, their data collection methods, their findings, and their difficulties. As their classmates presented their findings, the students who were not presenting filled in a "listening notes" worksheet on which they recorded either a question, something surprising, or a suggestion for improving the investigation. They then used these notes to provide feedback at the end of each team's presentation. Some students would share their surprise at what a team found, particularly if it was different from what they had noticed in their own investigations. Others would ask for clarification about how the team had taken a particular measure or about the number of observations on which a summary value was based. On Day 8 we wrapped up the remaining team presentations. Then, in the final minutes of class we looked holistically at the data summaries from each team. Students highlighted potential patterns of abundance, variation, and uniformity within the creek and hypothesized about the importance of spatial differences.

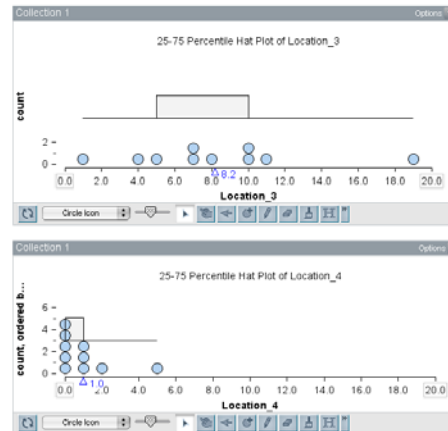
Design of Cycle 3: Days 9–13 The third cycle of investigation focused on examining whether empirical patterns of covariation uncovered likely ecological relationships. On Day 9, we used students' ideas about space and our class map to divide the creek into four sections for comparison. As before, students constructed their own data collection plans for our last creek visit. For this visit, each team was assigned a single variable (e.g., number of minnows) in which they had developed expertise and a single section of the creek in which to work. Teams would then share data in order to answer their research questions. This design allowed us to collectively measure all variables of interest to students within the time that we had in the field. It also prompted students to think about and resolve differences in how teams were collecting data on the same variable.

On Day 10, the class visited the creek for the third and final time. After each team completed their observations, they summarized their data using self-selected measures: the mean and the range. I then

copied each team's findings into a single table (Figure 1) that displayed the mean and the range for each variable in each creek section.

| | Location 1 | Location 2 | Location 3 | Location 4 |
|----------------------|---------------------------------------|-------------------------------------|--|---------------------------------------|
| 1) Water Temp | Mean = 13.5 °C Range = 0.9 | Mean = 14.0 °C Range = 2.1 | Mean = 13.5 °C Range = 0.7 | Mean = 13.8 °C Range = 0.3 |
| 2) Water Depth | Mean = 4.75 inches Range = 4 | Mean = 6.74 inches Range = 7 | Mean = 6.6 inches Range = 3 | Mean = 11.2 inches Range = 0.4 |
| 3) Water Speed | Mean = 1.64 ft/s Range = 1.48 | Mean = 0.70 ft/s Range = 1.05 | Mean = 0.59 ft/s Range = 0.58 | Mean = 1.07 ft/s Range = 4.74 |
| 4) Oxygen | Mean = 5.0 ppm Range = 4 | Mean = 5.0 ppm Range = 4 | Mean = 9.4 ppm Range = 6 | Mean = 5.2 ppm Range = 5 |
| 5) pH | Mean = 7.35 Range = 0.6 | Mean = 7.2 Range = 0.0 | Mean = 6.6 Range = 0.6 | Mean = 6.4 Range = 0.6 |
| 6) Nitrates | Mean = 10 ppm Range = 28 | Mean = 1.1 ppm Range = 1.7 | Mean = 7.5 ppm Range = 15 | Mean = 20 ppm Range = N/A |
| 7) Phosphate | Mean = 0.0 ppm Range = 4 | Mean = 0.86 ppm Range = 0 | Mean = 1.3 ppm Range = 1.5 | Mean = 2.0 ppm Range = 3 |
| 8) Hardness | Mean = 210 ppm Range = 150 | (No data) | Mean = 274 ppm Range = 305 | Mean = 163 ppm Range = 130 |
| 9) Algae/Plants | (Descriptive data) | (Descriptive data) | (A) Mean = 0.25 cup Range = 0.5 (P) Mean = 25.2 inches Range = 31 | (P) Mean = 30.0 inches Range = 30 |
| 10) Pollution | Mean = 4.7 color Range = 9 | Mean = 4.0 color Range = 5 | Mean = 4.2 color Range = 7 | Mean = 4.0 color Range = 9 |
| 11) # crayfish | Mean = 0.1 crayfish Range = 1 | Mean = 0.4 crayfish Range = 1 | Mean = 0.9 crayfish Range = 3 | Mean = 0 crayfish Range = 0 |
| 12) # minnows | Mean = 0 minnows Range = 0 | Mean = 0.03 minnows Range = 1 | Mean = 0 minnows Range = 0 | Mean = 0 minnows Range = 0 |
| 13) # water striders | Mean = 3.1 striders Range = 8 | Mean = 1.1 striders Range = 4 | Mean = 8.2 striders Range = 18 | Mean = 1.0 striders Range = 5 |
| 14) Length crayfish | Mean = 5.05 cm Range = 7 | Mean = 3.4 cm Range = 7 | Mean = 4.5 cm Range = 6 | Mean = 3.3 cm Range = 5.5 |
| 15) Invertebrates | Mean = 3.4 animals Range = 7 | Mean = 0 invertebrates Range = 0 | (Descriptive data) | Mean = 0.8 invertebrates Range = 2 |
| | Mean = 0.3 invertebrates Range = 1 | | | |

a) Data summary from third creek visit



b) Data on the number of water striders in locations 3 and 4

Figure 1. Data representations used during Day 11 and Day 12

On Day 11, we used this table to collectively review what the different teams had discovered. We searched for patterns within a single variable by looking for instances where there were similar mean values across all creek sections (e.g., pollution) and for instances where there seemed to be different mean values in different locations (e.g., water striders). The students then connected the data to their research questions. For example, we considered what the data might tell us about the best habitat for crayfish. Each team then selected their own research question to examine in more depth.

On Day 12, each team was given teacher-generated histograms and hat plots (e.g., Figure 1) of the data specific to their research question. The teams used these representations to think about whether or not there was an ecological difference in a variable across different sections of the creek. They also used a teacher-generated resampling model in the data visualization tool Tinkerplots™ (Konold & Miller, 2005) to explore the possibility that a difference of a certain magnitude could occur by chance. The detailed nuances of students' informal reasoning on this day were not a focus of analysis for this particular study.

On Day 13, students shared their ideas about their research questions, and we generated a class concept map of the relationships between different parts of the creek. Some of these relationships came directly from students' analyses (e.g., one team determined we could be fairly confident that areas with faster water speeds had fewer minnows), and some came from outside knowledge that students applied to the creek.

3.4. DATA COLLECTION AND ANALYSIS

I collected data on students' sampling practice from a variety of sources, including initial and final student interviews, pre/post-tests, student written artifacts, and video records of student activities both in class and at the creek. I began the analysis with the student interviews, as these provided the most detail about these students' sampling decisions. Using the data analysis methods described below, I uncovered three emergent aspects of students' sampling practice in these interviews:

1. Attention to having multiple observations, rather than a single point, in a sample;
2. Attention to absence as well as presence in a sample;
3. Attention to the differentiation of space and the selection of sample location.

I then analyzed the pre- and post-tests to see the extent to which these findings were evident across all students. Finally, I used students' written artifacts and the video record of students' activities to describe when and how these aspects of sampling emerged in students' practice. Details of the analysis of each data source are provided below.

Student interviews I conducted individual, 20-minute, semi-structured initial and final interviews with each of the eight students in the focus groups around a variety of hypothetical scenarios involving measurement, sampling, and variation, such as:

Benson decided to use a net to sample once in the morning and once in the afternoon. He counted four butterflies in the morning and nine in the afternoon. Given his findings, can Benson be confident that there are usually more butterflies in the park in the afternoon? Why or why not? Suppose it's true that there really are usually more butterflies in the park in the afternoon. If that's right, do you think that Benson will always catch more butterflies in the afternoon than in the morning? Why or Why not?

The same scenarios were used for both the initial and final interviews. During the final interview, I also asked students to describe various elements of their creek investigations, such as what question they explored, what they found out, and how they had made decisions about sampling. Interview findings are reported for the seven students for whom there are paired initial and final interviews. At the end of instruction, I looked holistically at each interview and used constant comparative methods to develop open codes for the various ways in which students were talking about samples and sampling, such as whether they considered sample size when critiquing a data collection plan or an analysis (Strauss & Corbin, 1990). Examples of these open codes can be found in the figures presented in the findings, under the headings initial and final perspectives. I then compared the codes across the initial and final interviews to identify axial codes, or patterns, describing how students' decisions about sampling had changed over the course of the investigation. This axial coding produced the three themes introduced above.

Student pre/post-tests Each student completed an individual pre-test at the start of cycle 2 and a post-test after the last instructional day of cycle 3. Findings are reported for the 37 students (74%) with paired pre/post-tests. The pre/post-tests asked students to design an investigation that would determine whether the number of grasshoppers in an area was related to the soil temperature. Though this scenario was designed to be similar to the types of investigations the students conducted in the creek, it was situated in an unfamiliar ecological context and focused on different variables than those with which students were familiar. I added two additional prompts to the post-test that were informed by aspects of measurement and sampling that had emerged within the study:

- “At the creek, our measurements were often different when we repeated them. Explain why our measurements of crayfish length might vary.”
- “If you are studying the invertebrates in the creek and in your first scoop you find no invertebrates, do you count that scoop and write zero on your data sheet?”

I coded the pre/post-tests for how students addressed the three themes, described earlier, that had emerged from my prior analysis of the student interviews.

Student written artifacts I collected all student worksheets from across the study and examined each for the presence/absence of the three themes about sampling. I elected to code the planning phase of each cycle separately from the observational phase at the creek so that I could document when in the investigative process students attended to sampling. For many of the teams one student served as the primary recorder, particularly when at the creek. Because of this I coded the investigation plans and data reports at the team level and looked holistically across all of the team members' written work to assign a team code. However, as students individually completed their own listening notes during the research meeting presentations, I coded these notes at the student level.

Video records of student activity I collected video records of each whole class discussion as well as the students' small group work both in the class and at the creek. As with the pre/post-test and written artifacts, I used the themes highlighted from my coding of the interviews as a lens to analyze these records. I looked across the video record for evidence of when and how these themes emerged in students' practice and used this to add depth and context to the findings from other data sources.

4. FINDINGS

This study began with the question: What initial aspects of sampling emerge as meaningful to middle school students during ecological field investigations? As highlighted above, the data analysis revealed three themes about students' emergent sampling practice: attention to repeated observations in a sample, attention to absence as well as presence, and attention to differentiated space and sample location. The findings explore each of these aspects of sampling in turn. For each aspect, I begin by explaining the nature of the change in students' sampling practice as revealed by the student interviews. I then connect this to evidence from the pre/post-tests about similar shifts across all students. Finally, I use the students' written artifacts and the video record to detail how that aspect of sampling emerged during students' ecological investigations. This story of emergence addresses the second research question: How can attention to variation support middle school students' development of the practice of sampling? To conclude the findings, I report briefly on students' views about sources of variation.

4.1. ATTENTION TO VARIATION HIGHLIGHTS PATTERNS, NOT POINTS, OF DATA AND CREATES A NEED FOR REPEATED OBSERVATIONS IN A SAMPLE

At the start of the investigation most students in the focus groups were confident that a single data point could adequately characterize a phenomenon if that data point was collected in what was perceived to be an appropriate measurement approach. For example, when asked whether she could be confident that there are usually more butterflies in the afternoon if in the morning she went out and took a single sweep of a net and caught four butterflies and then in the afternoon she did the same thing and found nine butterflies, Sharra said "Yes," so long as the sweeps were of the same size and with the right net (Figure 2). Mary was likewise confident in a single sweep. When probed whether or not it would be good to take more than one sweep each time, Mary continued, "Um, yeah, because you could always double-check yourself." This idea of double-checking that they had found the right answer and had not made a mistake during measuring was the primary reason students would initially pursue any form of repeated observation. Students thought repetition improved data collection by fixing the mistakes that had stopped them from getting the "true" measurement. Only one student, Gary, suggested during the initial interview that he would need to take multiple sweeps when collecting data because each sweep would likely have different numbers of butterflies. In his words, "Because if you're just doing one scoop, there might be a place behind you that has a whole bunch of them and you only saw that one spot that has just a few of

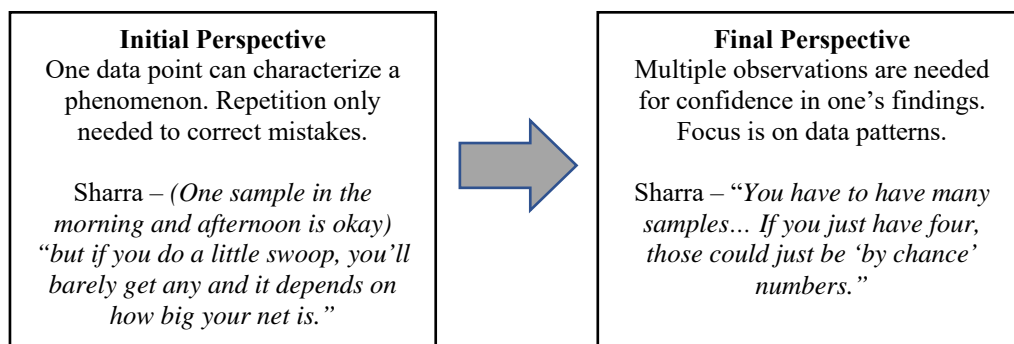


Figure 2. Shift in students' perspective of repeated observations in a sample

them.” Even here, Gary’s reasoning favored a “catching” mentality rather than a true sampling perspective of variability.

During the initial interview, students were also given displays of already collected data and asked to decide whether to continue or to stop gathering data. Here, when relying on one data point was not a choice, two students suggested that you could be more confident of your estimate of an attribute by gathering more measures, particularly if there was not yet a discernable “clump” in the data display. However, these same students viewed a single data point as satisfactory in data collection plans, indicating that this search for patterns in data displays was not connected to their plans to construct data.

However, by the end of their investigations, students’ notions about including repeated observations in a sample had undergone a dramatic shift, although they were still fairly simplistic by disciplinary standards. During the final interviews, every focus student considered a sample with a single observation to be insufficient to have confidence in one’s findings. This epistemic commitment was consistent across students’ descriptions of their creek investigations as well as their critiques of hypothesized scenarios. However, students still struggled to make sense about the extent to which they should repeat their observations. More than once was essential. But the question remained as to how much more. In trying to decide how much to sample, students would now often talk about taking observations until you are able to see the “main part” or “clump” in your data, even when developing a data collection plan. This shift from focusing on data points to focusing on data patterns seems to lie at the core of students’ perspective of including repeated observations in a sample.

Students’ ideas about chance also seemed to play out in their ideas about the need for repeated observations, even though these students did not have any formal experiences with probability and chance outside of this study. In the final interview, two students suggested that repeated sampling might help them account for random variation in their measures. For example, in describing her group’s decisions about how much to sample in the creek, one student, Sharra, explained that taking a small number like four samples would not always be enough to uncover the underlying pattern in the data because “if you just have four, those could just be ‘by chance’ numbers.” Here Sharra is referencing the idea that it is possible that the first observations you take may, just by chance, not be indicative of the pattern of data that would emerge after taking more measurements.

Pre/Post-Test findings This strong shift in attention to multiple observations in a sample was not as evident in students’ pre/post-test responses (Table 2). On their pre-tests, a plurality of students (43%) either did not describe the number of times they intended to measure or planned to measure only once in each condition. A few suggested measuring “many times” or to “repeat” their measurement process, with many suggested measuring two to six times. (Measuring two to six times was collapsed into a single category because these were the values where at least one student explicitly referenced using repetition to “double-check” that they had gotten the right answer.) Only one student planned to repeat the process at least 10 times. On their post-tests, which were completed after the third investigation, a plurality of students (38%) still did not describe the number of times they intended to measure or planned to measure only once in each condition. However, the number of students who planned to measure 10 or more times did increase to ten (27%). The maximum number of times any student suggested was 20.

Table 2. Students’ attention to repeated observations in pre/post-test data collection plans

| | Pre-Test Num students (%) | Post-Test Num students (%) |
|-----------------------------------|------------------------------|-------------------------------|
| Not described <i>or</i> only once | 16 (43%) | 14 (38%) |
| “Many times” or “repeat” | 7 (19%) | 3 (8%) |
| 2-6 times | 13 (35%) | 10 (27%) |
| 10-20 times | 1 (3%) | 10 (27%) |

The emergence of repeated observations in a sample In tracing how repeated observations emerged in students' practice, it was evident that most students initially considered a single data point to sufficiently characterize what they were studying. None of the teams included repeated observations in their data collection plans for their first visit to the creek (Day 2), and only 23% of the teams included multiple values for a variable in the data they collected (Day 3). For example, one team exploring the creek depth recorded just a single value of 12 inches. From this, one might think that the creek was a uniform entity. However, the written record tells only part of the story of students' first creek visit. The video record reveals that students were frequently surprised by the amount of variation found when trying to gather data in the creek. A student measuring depth would stand in the same spot and dip a yardstick in the creek multiple times with the water rising to a different level each dip. A team would take turns dropping a ping-pong ball in the creek to measure the water speed and the ball would take a different amount of time to float the same distance. A student would scoop one time with a net and catch four small crayfish, and the student next to them would scoop one time and catch one large crayfish. One team would find lots of minnows, and another lots of water striders. Because students were working in teams and because teams were working side-by-side, news of these differences would travel up and down the creek. Thus students, in unplanned and unstructured ways, were experiencing the variable results of repeated observations. This experience of variability helped call into question the reliability of a single measurement.

Consequently, repeated observations began to emerge as an essential aspect of sampling practice during the second investigation cycle. We began this cycle by discussing what students had seen during the first visit to the creek (Day 4). As a student or team shared what they had observed about different variables they compared their findings to what they had noticed. These comparisons highlighted the degree of variation within the creek. I also purposefully probed students about the consistency of their experiences. For example, I asked one team if the ball they had used to measure speed always floated down the creek in the same way. By this time many students were outright laughing at the suggestion that they would get the exact same value each time they took a measure in the creek, even if they stayed in the same location. I asked students how they could be confident that one part of the creek was deeper than another or had a better habitat for crayfish if measures could be so different even in the same area. Multiple students suggested we "take more than one sample" in each location. Students said we could "look for differences in the pattern" at each location or "compare the means" or other measures of center. This attention to pattern mirrored what some students had shared during the initial interviews about data displays—that you could be more confident of your estimate of an attribute by gathering more measures, particularly if there was not yet a discernable pattern in the data display. However, prior to this day's discussion, no student had invoked this reasoning when critiquing or designing data collection plans. The first visit to the creek seemed to create a shared experience around variation that supported students in bringing this reasoning to the foreground.

Thus, when preparing for the second creek investigation, 86% of the teams now included specifics about the number of observations that they intended to collect in their written data collection plans (Day 5) and reported on multiple observations in their findings (Day 6). For example, one team wrote, "Use a net or a hula hoop to check the amount of fish in the area. Do each (area) five times." This disposition towards the practice of including repeated observations in a sample was later reinforced when teams reported on their findings from the second visit during the research meeting (Days 7–8). Students often asked teams to describe how many "samples" (meaning observations) their findings were based on if they failed to share this detail in their report. When teams had not collected what others considered to be a sufficient number, they were encouraged to increase this number during the next cycle of data collection. Of the forty-four students present for the research meeting, sixteen (36%) included a question or comment about repeated observations in their listening notes. For example, one student suggested a team "take more samples (observations) to have more confidence" in their findings.

This emphasis on repeated observations carried over to the third cycle of investigation and was supported by students' inscriptional tools. Once again 86% of the teams included specifics in their data plans about the number of observations that they intended to take (Day 9), and all of the teams reported

findings from multiple observations (Day 10). For example, one team planned to “measure 10 to 20 times in the middle of location 2 and near the edge.” In their data report, this team ended up including findings from 30 observations. I asked one member of the team whether she thought this was a good number or if she would suggest a different number the next time. The student replied that the team had taken extra observations because they had extra time at the creek and that she didn’t think that they had needed all of them because they “could see the pattern of where most of the numbers would be after taking only twenty samples.” This hints that, through the emergent process of repeated observations, students were starting to build initial ideas about sample saturation as well.

4.2. ATTENTION TO VARIATION HIGHLIGHTS ABSENCE AS WELL AS PRESENCE AND CREATES A NEED TO “VARIABLEIZE” DATA

What students counted as a sample also shifted over the course of their creek investigations. During the initial interview, students in the focus groups only highlighted the material aspect of sampling, as all but one talked explicitly about a sample as a piece of something that they had cut from nature’s complexity. As Mary said, “(A sample) is a little bit...whatever we caught.” For students, a sample was the actual minnows caught in a scoop or the actual polluted cup of water pulled from the creek’s edge. However, by the end of their investigations, all of the students were also talking about a sample as the set of data they collected to help answer their questions. Thus, in the final interview the students’ conception of a sample encompassed Latour’s (1999) chain of transformation from material objects, in this case the minnows caught in a scoop, to inscriptions, in this case the numbers recorded for that scoop (Figure 3).

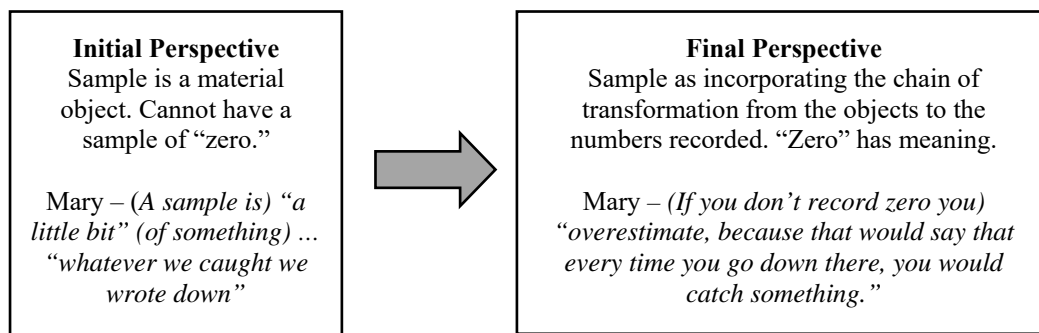


Figure 3. Shift in students’ perspective of absence and the nature of a sample

The shift in students’ view of the nature of a sample was accompanied by a parallel shift in how students attended to absence in their data. This was most evident in how students talked about constructing counts of different living organisms. Initially, students viewed an empty sweep of a net or a scoop of water with no organisms as a failure. They had not sampled the crayfish because they had not caught a crayfish. Absence was not a signal of the organism; it was a signal of incompetence. However, by the end of their investigation students had created a need for variable-like dimensionality in their measures. A “scoop of zero,” as students called it, was now meaningful. Absence as well as presence could be used to infer relationships between organisms and their environment. Mary highlighted this connection in her final interview when explaining why recording samples of zero was important in their creek investigations. She said that if you don’t record zero you “overestimate, because that would say that every time you go down there, you would catch something.” By the end of the investigation, students in both focus groups spontaneously referenced the importance of recording “zero” in either their research meeting notes, their data collection plans for the third creek visit, or their final interviews.

Pre/Post-Test findings The pre/post-test scenario failed to reveal students’ thinking about the nature of a sample or the function of absence in ecological investigations. No student explicitly addressed

absence on either the pre- or post-test. However, this is not surprising given the nature of the pre/post-test question. When writing a data collection plan, ecologists do not explicitly state that they'll be sure to record the observation if they don't happen to catch any organisms. Rather, their treatment of absence is revealed by their actions in collecting data. Similarly, students' actions in collecting data (and their critique of others' actions) likely revealed more about this perspective than their written plans.

On the post-test, students were asked an additional question: If you are studying the invertebrates in the creek and in your first scoop you find no invertebrates, do you count that scoop and write zero on your data sheet? In this case, all but one of the students (97%) indicated they would. In explaining their reasoning, students wrote that the scoop counted because it was "a part of the data," or because you had taken action to do something, or because it would "change the value of your mean" if you didn't include it.

The emergence of "scoops of zero" The shift in what counted as a sample emerged as students began to value absence—a notion highlighted though students' experiences with variation in the creek. In their initial investigations, students frequently wandered the creek in an attempt to capture some organism, such as a crayfish. In some areas they could catch a crayfish virtually every time they scooped a net into the creek. However, in other areas students would have to scoop over twenty times before they caught a single crayfish. Initially, students only recorded the organisms that they successfully caught without accounting for any scoops that came up empty. None of the teams recorded information about zero or the absence of an organism in the data from the first creek visit (Day 3). Rather, students recorded lists, total counts, or qualitative comparisons (more/less) of organisms.

However, this approach led to dissonance between how students were experiencing the creek and how they were representing it. Students would be recording similar counts for areas in which they had dramatically different material experiences. This dissonance created a dilemma for students who either noticed it on their own or who had it called to their attention. All of those previously ignored empty scoops offered students a way to resolve the problem that their sample was not representative of the underlying ecological phenomenon. During their second visit to the creek (Day 6), some students began to document every scoop they took—not just those with organisms. In their data records from this trip, for the teams that had to make a decision about absence, 55% recorded at least one "scoop of zero." However, during the following research meeting (Days 7–8) none of the students asked other teams about absence or wrote about attending to zero in their research meeting listening notes.

Because the students were already attending to absence in their actions, even if they weren't yet talking about it, I brought up the issue during our class discussion the next day (Day 9). I presented students with a scenario in which two teams had each found the same number of crayfish but one team had taken more scoops than the other, and I asked students to discuss whether the empty scoops of the second team mattered. All of the small group discussions concurred that the empty scoops mattered if you wanted to be able to make a comparison because, as one student said, "it changes what you are likely to find (in a scoop)." As these students planned the next round of data collection, the importance of "recording scoops of zero" became a reified aspect of collective sampling practice. In their final data records, all but one of the teams attended to absence as well as presence and recorded observations of zero if they were investigating organism abundance.

During these investigations, attention to absence did not emerge as readily as repeated observations in students' practice. It was never included in a team's plans for data collection, and it was not highlighted during the research meeting presentation. A variety of factors might have contributed to this pattern. Absence impacts only some of the variables students measured, such as measures of organism abundance. Many aspects of the creek, such as water speed, depth, crayfish length, and dissolved oxygen were already variabilized for students. In addition, absence is more abstract and thus potentially more difficult to talk about than repetition. Students' difficulty with absence might also reflect how they typically engage with natural spaces outside of school. Most students this age explore a creek to catch things. In such cases, absence is always a signal of failure. But in science, absence can be a signal of both success and failure. You may get a scoop of zero because that is a valid representation of the ecological

functioning of that location. Or you may get a scoop of zero because you had a momentary problem with wielding the net that you were using to take that sample. The first scoop of zero would need to be attended to. But the second scoop might be legitimately discounted as you make adjustments to your data collection technique.

4.3. ATTENTION TO VARIATION HIGHLIGHTS DIFFERENTIATED SPACE AND CREATES A NEED TO ATTEND TO SAMPLE LOCATION

The way in which students talked about space also shifted over the course of the investigation. During the initial interviews, only three students in the focus groups referenced location in any way when describing sampling or critiquing data collection plans. These students primarily focused on choosing a location that secured the most access to the phenomena of interest. For example, Mary explained that someone deciding how to sample for butterflies in a park needed to “go by the flowers” because “more of the butterflies would be by flowers.” This form of attention to space illustrates how students initially emphasized the hunt for organisms in their investigations (Figure 4). The only location to attend to when sampling was the one where you could sweep and catch the most organisms.

However, in the final interviews students talked about space in a very different way. Here, every student talked in some way about the need to differentiate space and about how sample location can impact the interpretation of one’s findings. For example, Mary described how “if the other people that were doing minnows and we wanted to compare...if we just did it in the middle and all of them just did it on the sides and the middle, they might have way different results.” Decisions about sample location were no longer only about catching an organism (although securing access to the phenomenon was still important). Decisions were about preserving the ability to make comparisons across data sets. Though students’ practice was a relatively crude approximation of the systematic approaches of professionals, students had begun to recognize that the locations in which they sampled did not just impact their access to the phenomena of interest, but the locations also impacted what they could do with their data.

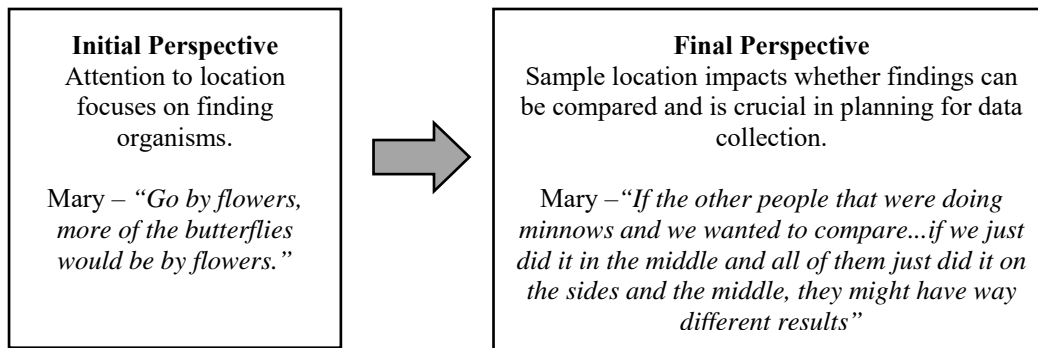


Figure 4. Shift in students’ perspective of sample location

Pre/Post-Test findings The shift in attention to sample location was also not as strongly evidenced in students’ pre/post-test responses (Table 3). On their pre-tests, the majority of students (57%) did not describe where they intended to take their measures, even though the question prompted them to do so. Those who did attend to location tended to use general proxies for the variables of interest (temperature and grasshopper abundance) to stratify space. For example, one student chose to “measure under the rock where it is shady and out in the sun.” This would likely ensure a temperature difference between locations. In addition, a few chose their sample location by stratifying the spatial structure of the field so that areas in the middle and edge were both included. On their post-tests, fewer students (38%) completely ignored sample location. However, students still did not attend to space in sophisticated ways, with the largest increase being in the number of students (16%) who suggested generally to sample

“different locations.” No student described plans to spatially locate individual samples within stratified space.

Table 3. Students’ attention to space in pre/post-test data collection plans

| | Pre-Test | Post-Test |
|--|--------------|--------------|
| | <i>n</i> (%) | <i>n</i> (%) |
| Not described | 21 (57%) | 15 (38%) |
| Sample “different locations” | 2 (5%) | 6 (16%) |
| Stratified by proxy for temperature or grasshoppers | 7 (19%) | 8 (22%) |
| Stratified by spatial structure such as middle or edge | 3 (8%) | 2 (5%) |
| Other descriptions of space | 4 (11%) | 4 (11%) |

In the interviews, the students in the focus groups had described their stratification of the creek (e.g., middle vs. sides) in relation to their specific experiences finding different organisms in those different areas. However, students did not have such experiences with the ecosystem in which the pre/post-test scenario was situated. Because of this, it may have been difficult for them to imagine which factors might influence organism abundance and would thus need to be accounted for in decisions about sample location.

The emergence of attention to space The earlier account of the emergence of repeated observations alluded to the means by which variation in a measure can create a need to see space from a new perspective. In the beginning, students viewed the creek as a singular entity. In our first discussion (Day 1) and their first data collection plans (Day 2), students asked questions such as “how deep is the creek” and “how many fish are there.” However, as students began to see differences between repeated measures of the same variable, they began to partition the creek at a gross level. For example, the area by “the car wash had fast moving water.” The area by Granny’s bridge had “lots of water striders.” These gross differences emerged in students’ observational records from the first visit to the creek (Day 3) and supported the development of our map of creek at the start of cycle 2 (Day 4).

Over time this differentiation of space supported new observations about variation, which in turn led to the development of new research questions. For example, during the second visit to the creek one focus group started exploring whether the average number of crayfish was related to water speed (Days 5–6). Because their observations of crayfish seemed to vary even at a single location with relatively slow water, they further stratified that location to test an emergent hypothesis relating crayfish and shallow habitats. In the research meeting after this second visit (Days 7–8), students began to ask the other teams about where they had sampled and how they had made these decisions. In their research meeting notes (Day 8), 41% of the students included an observation or suggestion about where teams should sample.

Although students began to differentiate spaces in the creek based on variation after their very first visit to the creek, the need to attend to sample location within data collection plans did not emerge until planning for the third visit (Day 9). Prior to this day, none of the teams’ data collection plans described where they would collect their data within the two general locations they had selected. In planning for the third visit, teams were given responsibility for collecting data about a single variable in one of four general locations in the creek. Teams would then share their findings to look for patterns of covariation in variables across locations. As methods of data collection would impact the validity of this comparison, this helped create a need to make implicit notions of sampling location explicit in students’ plans. In this, students often chose sampling locations that accounted for the breath of spatial variation and incorporated some aspect of stratification. For example, one team decided to distribute the ten samples that they intended to take so that three were equally spaced along the far bank, four were spaced through the middle of the creek, and the last three were spaced across the near bank. Though all teams talked about where they would sample when planning, only 21% of the teams detailed these decisions in their written data collection plans. These teams included hand-drawn diagrams of where they intended take each

sample. These inscriptions were not the final products of the teams' decisions, but rather emerged as students negotiated different options for data collection by means of visually representing their ideas.

4.4. STUDENTS' VIEWS ABOUT SOURCES OF VARIATION

These three findings suggest that attention to variation is important in the evolution of students' sampling practice. Given this, it is important to consider what students think contributes to variation in their data. Although sources of variation in ecological contexts is typically not explored until high school, I added an additional probe on the post-test to examine this very question. Students were asked, "At the creek, our measurements were often different when we repeated them. Explain why our measurements of crayfish length might vary." Responses were analyzed for evidence of causal (induced) variability, measurement variability, and natural variability. These categories were not mutually exclusive as students could reference multiple sources of variation in their response. This context was selected because of the ubiquity of crayfish in students' experiences at the creek and because all three forms of variability had emerged during class discussions of crayfish. Thus, although students' expressions of their ideas are reflective of this specific context, the context had potential to provoke multiple lines of reasoning about variation.

Approximately two-thirds of the students referenced some form of causal variability. Eleven of these students (30%) described a general cause, such as measuring crayfish from different locations of the creek, whereas 14 (38%) described a specific cause, such as measuring crayfish from locations with different levels of pollution. Eight students (22%) described ways differences or errors in measurement could have contributed to variation in data. For example, one student explained, "Some could fold the tail in and others don't. Or they could have gaps" (in measuring crayfish). Only three students (8%) alluded to some form of natural variation without attributing a cause. For example, one student stated, "Crayfish are all different sizes." Although some students described how both causal variation and measurement error could contribute to variation in measures, none coupled natural variation with a second source.

5. DISCUSSION

This study has detailed the ways in which ecological field investigations have the potential to foreground important aspects of sampling. Students' interest in and attention to sampling emerged from their own questions about the efficacy of their data in accounting for ecological phenomenon. Though the students' methods did not approach the complexity of practice seen in professionals, they began to attend to many of the same issues and problems that field ecologists and statisticians consider when sampling (Coe, 2008; Eberhardt & Thomas, 1991; Franklin et al., 2007). These questions about sampling emerged from moments in which students were wrestling with some form of variation. As such, variation seems important in creating a student-perceived need for more sophisticated aspects of sampling: using repeated observations in samples (sample size), sampling absence as well as presence, and attending to sampling location (Table 4). These are significant developments for middle school students. They also lay the

Table 4. Summary of students' emergent sampling practice across the creek investigations

| | |
|---------|--|
| Cycle 1 | <ul style="list-style-type: none"> No evidence of attention to repeated observations in samples, to samples of zero, or to sampling locations in data collection plans |
| Cycle 2 | <ul style="list-style-type: none"> Data collection plans include some attention to repeated observations in samples after student discussion highlighted variability in the first cycle's data Some teams begin to record samples of zero in data at the creek Students' research meeting discussion highlights sample size and samples of zero |
| Cycle 3 | <ul style="list-style-type: none"> Data collection plans include attention to sample size and, for some, sampling location Samples of zero more consistently recorded while at the creek |

foundation for students to explore more advanced aspects of sampling, such as sample saturation, sampling variability, and random assignment, in the future.

5.1. PROMOTING STUDENTS' EMERGENT SAMPLING PRACTICE

Although this study traversed scientific and statistical boundaries, it was initially grounded in a science-as-practice perspective. This perspective uses the design of the learning environment to create a local, meaningful need for students to progressively refine a particular practice, rather than presenting it as a set of a priori procedures (Manz, 2012, 2014). By applying this perspective to sampling, this study has revealed three potential features of learning environments that can be leveraged to support students' emergent sampling practice: personal encounters with variation, moments of comparison, and the problematization of practice.

Personal encounters with variation In this study students' attention was drawn first to the need for repeated observations in their samples. This early emergence was likely reflective of the degree of variation present in the ecosystem that students studied. If students were investigating a system with weaker gradients of variability, it might have proven more difficult to create a need for repeated observations.

Students' increasing attention to and recognition of variation paralleled their increasing expertise in the local ecological context. Initially, students were not expecting to need to take more than one measurement of a variable of interest and were surprised by the diversity present at the creek. The students' physical encounters in the specific ecological context challenged their initial intuitions about variability. Lehrer and Schauble (2017) described finding a similar emphasis on repeated observations or aggregate samples after middle school students had participated in year-long ecological investigations. And, in interviewing students about a variety of statistical scenarios, Watson (2009) found a comparable relationship between increased contextual knowledge and more sophisticated intuitions about variation.

Although students recognized a need to attend to and account for variability in their investigations, they did not fully understand and distinguish between the multiple sources contributing to that variability. Thus, they began to design for differences in their study before they could specifically account for the processes underlying those differences. This finding deviates from the developmental framework laid out in the GAISE report. The GAISE report recommends that students develop a conceptual understanding of some of the sources of variability (measurement, natural, induced) before they begin to design for differences (Franklin et al., 2007). This difference may reflect distinctive epistemic commitments in the disciplines of statistics and science, as science emphasizes student-driven questions and investigations even at the earliest grade levels (e.g., National Research Council, 2012). It may also simply reflect an alternative trajectory that emerged from the unique way students engaged with variation in their ecological studies.

Students' emergent attention to and valuing of repeated observations could potentially be leveraged to introduce them to more sophisticated explorations of sample size. By the end of their creek investigations, students seemed primed to consider how to use the degree of variation in a measure to make decisions about what counts as a satisfactory sample size. Students were particularly attentive to the absence or presence of clusters of values in their samples. These clusters help establish a signal in the data. In effect, the students were intuitively beginning to look for stability in the distribution as the number of observations in their sample increased (Konold & Pollatsek, 2002). Statistics educators have found that students' attention to clusters of values in repeated measurements of features such as arm span or table perimeter can be fruitful in supporting an emergent awareness of distribution (e.g., English & Watson, 2015; Lehrer & Kim, 2009; Lehrer, Kim, & Jones, 2011). Ecological fieldwork might offer a complementary context to advance a similar agenda.

Moments of comparison Throughout the students' creek investigations, moments of comparison created meaningful opportunities for students to advance their sampling practice. Students' experiences comparing different locations were formative in their attention to organism absence in their

investigations, as indicated by the emergence of “scoops of zero.” Likewise, students’ attention to the location of observations within differentiated space was tied to the instructional move to have teams collaboratively plan, rely on, and compare each other’s work.

In statistics, comparison plays a powerful role in students’ recognition of and interpretation of variability. Shaughnessy and Pfannkuch (2002) documented the value of comparison while working with students to predict the timing of geyser eruptions. “When students first look at one day’s data and then see that a classmate’s data for a day look quite different, they begin to see variability from day to day, as well as within a single day” (p. 256). Likewise, Konold and Pollatsek (2002) argued that the meaning of a signal lies in comparison, as comparison helps the signal rise from variability. In addition, Watson and Moritz (1999) found that statistical contexts involving comparison are not only more interesting to students, but that they also allow students to see the usefulness of different statistical approaches.

Field investigations, such as the one in this study, can be particularly powerful in promoting comparison because moments of comparison are embodied in students’ actions. At the creek, students physically experienced variability, in water depth or the types of organisms present, each time they took a step. Students would compare different observations within the same area, findings across different areas, how others were gathering data and using tools, and what other teams were finding nearby. These comparisons allowed students to critique their current sampling practice and elicit new insights and new approaches to sampling. In the classroom, the format of the research meeting created moments of collective comparison for students, similar to those found by Lehrer, Schauble, and Lucas (2008) during their research meetings.

The problematization of practice Attention to absence and the differentiation of space are two distinctive aspects of sampling in field ecology that do not directly carry over to other sampling contexts. The emergence of “scoops of zero” required students to be working with a feature of the ecosystem, such as organism abundance, that had a need to be variabilized. If students had solely been focusing on measures such as dissolved oxygen or crayfish length that have a built-in zero point, there would not have been the same need to attend to absence. Similarly, questions about location are vital to developing effective and efficient sampling strategies in ecology. In fact, for some ecological questions and contexts, regular grid sampling or equal-stratified sampling can produce results that more accurately model the ecological phenomena than those produced by random sampling, particularly when working with small sample sizes (Hirzel & Guisan, 2002). In contrast, methods of random sample selection are typically emphasized in statistics education for every question or context (Franklin et al., 2007).

Though these disciplinary approaches appear to potentially be in conflict, common ground can be found by positioning questions of absence and location as moments that problematize the practice of sampling. When students consider differentiating space or counting a scoop of zero that they previously ignored, they are in essence raising questions about the representativeness of the sample. Consequently, students’ attention to absence or to location could potentially be leveraged to introduce broader discussions of bias in sampling. In statistics education, bias is commonly introduced within the context of sociological surveys (e.g., Watson & Kelly, 2005). Sampling in field settings might be able to add breadth to students’ experiences with bias as students’ ecological investigations could be used to legitimize the need to attend to potential sources of bias. Students’ emergent ideas could then be investigated in more depth using traditional classroom-based explorations of models, experiments, and surveys.

5.2. IMPLICATIONS FOR TEACHING

In addition to the three themes described, this study has additional implications for the design of learning environments in science and statistics education. First, although statistics education emphasizes the importance of having students collect their own data, most student-driven studies focus on experimental settings or sociological surveys (Franklin et al., 2007; Konold & Pollatsek, 2002). Because of this, students frequently do not gain experience applying statistical reasoning to scientific disciplines, such as ecology, in which observational studies are vital to investigating questions of practical

importance. Ecology has gained increased precedence in science education as it offers a relatively accessible and compelling way for students to engage with complex systems and critical socio-scientific issues such as habitat destruction and climate change (Jordan, Singer, Vaughan, & Berkowitz, 2008; Lefkaditou, Korfiatis, & Hovardas, 2014). Integrating such contexts with statistics education could positively impact how students are able to statistically reason and make decisions about issues pertinent to their daily lives. Although the data students construct during ecological fieldwork is often messy (in addition to being muddy), observational-based ecological field studies have rich potential for initiating students' interest in questions of sampling and engaging students in what Konold and Pollatsek (2002) call "the general enterprise" of statistics: an understanding of how and why we collect and investigate data (p. 286).

Second, the creek investigations seemed to provide opportunities for students to make sense of variation due to causal processes as well as measurement error. However, students still struggled with issues of natural variation due to random processes. This struggle is not atypical for students and has been documented in other studies in which students investigated contexts that included natural variability (e.g., Metz, 1999; Torok & Watson, 2000). The overlapping influences of measurement, natural, and causal (induced) variability on a single observation can be difficult to tease out, particularly as natural variation is rarely presented from a statistical perspective in science classes. It may be that science and statistics educators need to collaboratively explore new classroom-based modeling experiences that could help tease apart these different sources of variation and ground students' perspective of the degree of variation that can be produced through random processes.

Finally, in this study the students' use of the word sample often conflicted with how the word is used in formal statistical settings. Students reflexively used sample to indicate an individual sampling unit, such as a scoop of water. When the notion of a single sample began to emerge as unsatisfactory, students would talk interchangeably about the need for repeated observations, repeated measurements, or repeated sampling. These carried the same meaning for students. In this paper, I have used the phrase "repeated observations" for clarity when writing about what students called "repeated sampling." However, I did not correct students' language during instruction as I was interested in students' emergent language use. Interestingly, the students did not have difficulty communicating the distinction between a sample and an observation with each other. Students would regularly ask and answer questions such as "How many samples (*referring to observations*) do you have in your data (*referring to sample*)?" without any miscommunication. There are two likely reasons why students' language differed from the statistical norm. First, in scientific contexts it is common to refer to remnants as samples. For example, an ecologist might talk about collecting a sample of twenty core samples from trees. Likewise, students would similarly describe their sampling plan as collecting twenty samples of water. Second, at the beginning of their investigations when students were confident in using a single observation, the sample and the sampling unit were physically equivalent. As students questioned the usefulness of that single observation, they gathered more observations by literally repeating the original sample. Thus, students' repetition of sampling created a frequency distribution, whereas in formal statistics repeated sampling creates a sampling distribution. Because of this, science and statistics educators might need to consider how to scaffold students' use of language in ways that prepare students for more sophisticated ideas, such as sampling variability, without simply authoritatively replacing students' initial language use.

5.3. IMPLICATIONS FOR RESEARCH

The findings of the study also suggest new conjectures by which future iterations of this study could build on the instructional design. First, it may be useful to streamline the planning phase of the first cycle and anchor students' initial investigations in accessible features of the ecosystem that exhibit substantial variation, specifically water depth and water speed. These are features of the abiotic environment in which students often have initial interest, strong resources for measuring, and the ability to self-monitor quantitative measures through qualitative observation. Second, as detailed above, the instructional design might benefit from new classroom-based modeling experiences that could help students tease apart

different sources of variability and ground students' perspective of the degree of difference that random processes can produce. Adaptations of Stohl and Tarr's (2002) and Lehrer's (Lehrer & Kim, 2009; Lehrer & Schauble, 2004) approaches to data visualization, simulation, and chance might be particularly fruitful in combination with students' fieldwork. Third, the revised design could also capitalize on the emergence of additional dimensions of sampling, such as sample saturation and sampling variability. Finally, as this study focused narrowly on the practice of sampling, future studies might consider how the design supports the co-development of students' knowledge and practice at a broader scale. In particular, it would be useful to better understand how students' sampling practice interacts with their performance of other scientific practices and understanding of other statistical concepts, such as informal inference.

5.4. LIMITATIONS

In streamlining this argument for how students' sampling practice developed I have had to strip away some of the nuanced complexities inherent in this work. Because of this I do not want to give the false impression that a more sophisticated sampling practice will spontaneously arise from merely engaging students in any form of ecological fieldwork. Rather, students' evolution of practice was fundamentally intertwined with the overall design of the learning environment and ecological context.

Nor do I consider attention to variation to be the sole impetus for advancing students' sampling practice. As students wrestle to develop their own measures and data collection plans, personal frustration and need can sometimes be productive stimuli for changes in practice. Likewise, it would be remiss to overlook that students' sampling practice evolved within the social context of nested, intersecting communities of learners (e.g., Lehrer et al., 2008). As has been found with professional field ecologists (Bowen & Roth, 2002, 2007; Feldman, Divoll, & Rogan-Klyve, 2009, 2013; Roth & Bowen, 2001), social interactions both within and out of the field were important for establishing and circulating knowledge within this community. An individual student's practice was refined through negotiating with their own group members, observing and jostling with other groups in the creek, reporting out to their science class, and finally sharing findings and anecdotes across the entire sixth grade.

In addition, this paper details the development of two classes of students from one rural community as they investigated one aquatic ecosystem. Though it is likely that attention to variation could support similar development in a different population of students studying a different ecosystem, it is also likely that some elements of the trajectory of development were locally contingent on the lived experiences of these students and the specifics of the ecosystem they studied. As Cobb and Moore (1997) emphasize, it is the context that provides meaning. Experience with a specific variable in a specific setting mediates the practice of even professional ecologists (Bowen & Roth, 2002, 2007; Lorimer, 2008; Roth & Bowen, 1999, 2001). Similarly, a student's personal sense of place likely influences their own sampling practice.

Finally, tracing and interpreting the sampling practice of middle school students at times proved to be a tricky endeavor. Many of the students exhibited difficulties with writing that impacted what they were able to convey in their data collection plans and on the pre/post-test. The richest signals of students' practice were found in the interviews and the video records of the creek investigations and research meetings. These moments captured students' actions while sampling and their critique of the actions of others. As was highlighted in the findings, the pre/post-test in particular failed to offer much insight into the evolution of students' sampling practice. This may have been because the post-test was given on the second-to-last day of the school year. But it may also have been because it focused on students' construction of data collection plans. Watson and Kelly (2005) have suggested that students might disproportionately struggle to create, as opposed to critique, sophisticated sampling plans in new contexts. Adapting the assessment so that it asks students to critique the sampling decisions of others might reveal more nuances in students' reasoning about sampling.

ACKNOWLEDGEMENTS

This article is based upon work supported by the Institute of Education Sciences under Grant No. R305A120217. Any opinions, findings, and conclusions or recommendations expressed are those of the author and do not necessarily reflect the views of the Institute of Education Sciences.

REFERENCES

- Alberto, F., Raimondi, P. T., Reed, D. C., Coelho, N. C., Leblois, R., Whitmer, A., & Serrão, E. A. (2010). Habitat continuity and geographic distance predict population genetic differentiation in giant kelp. *Ecology*, *91*(1), 49–56.
- Bacaro, G., Rocchini, D., Diekmann, M., Gasparini, P., Gioria, M., Maccherini, S., ... Chiarucci, A. (2015). Shape matters in sampling plant diversity: Evidence from the field. *Ecological Complexity*, *24*, 37–45.
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082–1112.
- Biswas, S. R., & Mallik, A. U. (2010). Disturbance effects on species diversity and functional diversity in riparian and upland plant communities. *Ecology*, *91*(1), 28–35.
- Bowen, G. M., & Roth, W.-M. (2002). The “socialization” and enculturation of ecologists in formal and informal settings. *Electronic Journal of Science Education*, *6*(3).
[Online: <http://ejse.southwestern.edu/article/view/7680/5447>]
- Bowen, G. M., & Roth, W.-M. (2007). The practice of field ecology: Insights for science education. *Research in Science Education*, *37*(2), 171–187.
- Bridgeland, W. T., Beier, P., Kolb, T., & Whitham, T. G. (2010). A conditional trophic cascade: Birds benefit faster growing trees with strong links between predators and plants. *Ecology*, *91*(1), 73–84.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.
- Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823.
- Coe, R. (2008). *Designing ecological and biodiversity sampling strategies*. Working Paper no. 66. Nairobi, Kenya: World Agroforestry Centre.
[Online: <http://www.worldagroforestry.org/downloads/Publications/PDFS/wp08177.pdf>]
- Council for Environmental Education. (2006). *Project WILD: K–12 curriculum & activity guide*. Houston, TX: Council for Environmental Education.
- DeWoskin, R. (1980). Heat exchange influence on foraging behavior of *Zonotrichia* flocks. *Ecology*, *61*(1), 30–36.
- Douglass, A. E. (1920). Evidence of climatic effects in the annual rings of trees. *Ecology*, *1*(1), 24–32.
- Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*(1), 268–291.
- Eberhardt, L. L., & Thomas, J. M. (1991). Designing environmental field studies. *Ecological Monographs*, *61*(1), 53–73.
- English, L. D., & Watson, J. M. (2015). Exploring variation in measurement as a foundation for statistical thinking in the elementary school, *International Journal of STEM Education*, *2*(3).
doi: 10.1186/s40594-015-0016-x
- Esterly, C. O. (1920). Possible effect of seasonal and laboratory conditions on the behavior of the copepod *Acartia tonsa*, and the bearing of this on the question of diurnal migration. *Ecology*, *1*(1), 33–40.
- Ewald, P. W., Hunt, G. L., Jr., & Warner, M. (1980). Territory size in western gulls: Importance of intrusion pressure, defense investments, and vegetation structure. *Ecology*, *61*(1), 80–87.
- Feldman, A., Divoll, K., & Rogan-Klyve, A. (2009). Research education of new scientists: Implications for science teacher education. *Journal of Research in Science Teaching*, *46*(4), 442–459.

- Feldman, A., Divoll, K. A., & Rogan-Klyve, A. (2013). Becoming researchers: The participation of undergraduate and graduate students in scientific research groups. *Science Education*, 97(2), 218–243.
- Ford, M. J. (2015). Educational implications of choosing “practice” to describe science in the Next Generation Science Standards. *Science Education*, 99(6), 1041–1048.
- Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, 30, 1–32.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., Schaeffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association.
[Online: http://www.amstat.org/asa/files/pdfs/GAISE/GAISEPreK-12_Full.pdf]
- Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157, 331–341.
- Hofmann, J. V. (1920). The establishment of a Douglas Fir forest. *Ecology*, 1(1), 49–53.
- Jordan, R., Singer, F., Vaughan, J., & Berkowitz, A. (2008). What should every citizen know about ecology? *Frontiers in Ecology and the Environment*, 7(9), 495–500.
- Kelly, G. J. (2011). Scientific literacy, discourse, and epistemic practices. In C. Linder, L. Östman, D. A. Roberts, P.-O. Wickman, G. Ericksen, & A. MacKinnon (Eds.), *Exploring the landscape of scientific literacy* (pp. 61–73). New York: Routledge.
- Kenkel, N. C., Juhász-Nagy, P., & Podani, J. (1990). On sampling procedures in population and community ecology. In G. Grabherr, L. Mucina, M. B. Dale, & C. J. F. T. Braak (Eds.), *Progress in theoretical vegetation science* (pp. 195–207). Dordrecht, The Netherlands: Springer.
- 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.
- Knorr-Cetina, K. (2009). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Konold, C., & Miller, C.D. (2005). *TinkerPlots: Dynamic data exploration*. [Computer software]. Emeryville, CA: Key Curriculum Press.
- Korfiatis, K. J., & Tunnicliffe, S. D. (2012). The living world in the curriculum: Ecology, an essential part of biology learning. *Journal of Biological Education*, 46(3), 125–127.
- Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lefkaditou, A., Korfiatis, K., & Hovardas, T. (2014). Contextualising the teaching and learning of ecology: Historical and philosophical considerations. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 523–550). Dordrecht, The Netherlands: Springer.
- 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., & Jones, R. S. (2011). Developing conceptions of statistics by designing measures of distribution. *ZDM*, 43(5), 723–736.
[Online: www.researchgate.net/publication/225632082_Developing_conceptions_of_statistics_by_designing_measures_of_distribution]
- Lehrer, R., & Romberg, T. A. (1996). Exploring children's data modeling. *Cognition and Instruction*, 14(1), 69–108.
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 635–679.
- Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, 96(4), 701–724.

- Lehrer, R., & Schauble, L. (2017). Children's conceptions of sampling in local ecosystems. *Science Education*, *101*, 968-984.
- Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development*, *23*(4), 512-529.
- Lorimer, J. (2008). Counting corncrakes: The affective science of the UK corncrake census. *Social Studies of Science*, *38*(3), 377-405.
- Manz, E. (2012). Understanding the codevelopment of modeling practice and ecological knowledge. *Science Education*, *96*(6), 1071-1105.
- Manz, E. (2014). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research*, *85*(4), 553-590.
- McClure, M. S. (1980). Foliar nitrogen: A basis for host suitability for elongate hemlock scale, *Fiorinia externa*. *Ecology*, *61*(1), 72-79.
- McLellan, B. N., Serrouya, R., Wittmer, H. U., & Boutin, S. (2010). Predator-mediated Allee effects in multi-prey systems. *Ecology*, *91*(1), 286-292.
- Metz, K. E. (1999). Why sampling works or why it can't: Ideas of young children engaged in research of their own design. In F. Hitt & M. Santos (Eds.), *Proceedings of the 21st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 492-498). Cuernavaca, Mexico: PME.
[Online: <http://www.matedu.cinvestav.mx/publicaciones/e-librosydoc/pme-procee.pdf#page=492>]
- Mody, C. C. M. (2015). Scientific practice and science education. *Science Education*, *99*(6), 1026-1032.
- Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy*. (pp. 95-137). Washington, DC: National Academy Press.
- Mörsdorf, M. A., Ravolainen, V. T., Støvern, L. E., Yoccoz, N. G., Jónsdóttir, I. S., & Bråthen, K. A. (2015). Definition of sampling units begets conclusions in ecology: The case of habitats for plant communities. *PeerJ*, *3*(e815). doi: 10.7717/peerj.815
- National Curriculum Board. (2009). *The shape of the Australian curriculum*. [Online: acaraweb.blob.core.windows.net/resources/The_Shape_of_the_Australian_Curriculum_May_2009_file.pdf]
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Nersessian, N. (2008). Model-based reasoning in scientific practice. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching scientific inquiry: Recommendations for research and implementation* (pp. 57-79). Rotterdam: Sense Publishers.
- NGSS Lead States. (2013). *Next generation science standards: For states by states*. Washington, DC: The National Academies Press.
[Online: www.nap.edu/catalog/18290/next-generation-science-standards-for-states-by-states]
- Nielsen, J. A. (2011). Dialectical features of students' argumentation: A critical review of argumentation studies in science education. *Research in Science Education*, *43*(1), 371-393.
- Noll, J., & Shaughnessy, J. M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education*, *43*(5), 509-556.
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177-196.
- Patterson, T. A., McConnell, B. J., Fedak, M. A., Bravington, M. V., & Hindell, M. A. (2010). Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. *Ecology*, *91*(1), 273-285.
- 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. (2008). Building sampling concepts for statistical inference: A case study. *International Congress on Mathematical Education* (pp. 6-13). Monterrey, Mexico.
[Online: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.3601&rep=rep1&type=pdf]
- Praeger, W. E. (1920). A note on the ecology of herons. *Ecology*, *1*(1), 41.

- Pratt, D., Johnston-Wilder, P., Ainley, J. & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, 7(2), 107–129.
[Online: [https://iase-web.org/documents/SERJ/SERJ7\(2\)_Pratt.pdf](https://iase-web.org/documents/SERJ/SERJ7(2)_Pratt.pdf)]
- Ravet, J. L., Brett, M. T., & Arhonditsis, G. B. (2010). The effects of seston lipids on zooplankton fatty acid composition in Lake Washington, Washington, USA. *Ecology*, 91(1), 180–190.
- Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Richard, V., & Bader, B. (2010). Re-presenting the social construction of science in light of the propositions of Bruno Latour: For a renewal of the school conception of science in secondary schools. *Science Education*, 94(4), 743–759.
- Rogers, R. S. (1980). Hemlock stands from Wisconsin to Nova Scotia: Transitions in understory composition along a floristic gradient. *Ecology*, 61(1), 178–193.
- Roth, W.-M., & Bowen, G. M. (1999). Digitizing lizards: The topology of 'vision' in ecological fieldwork. *Social Studies of Science*, 29(5), 719–764.
- Roth, W.-M., & Bowen, G. M. (2001). Of disciplined minds and disciplined bodies: On becoming an ecologist. *Qualitative Sociology*, 24(4), 459–481.
- Rubin, A., Bruce, B., and Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314–319). Dunedin, New Zealand. Voorburg, The Netherlands: International Statistical Institute.
[Online: <https://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A9-4.pdf>]
- Saldanha, L. & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257–270.
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345–372.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Schweiger, A. H., Irl, S. D. H., Steinbauer, M. J., Dengler, J., & Beierkuhnlein, C. (2016). Optimizing sampling approaches along ecological gradients. *Methods in Ecology and Evolution*, 7(4), 463–471.
- Sharma, S. (2003). An exploration of high school students' understanding of sample size and sampling variability: Implications for research. *Journal of Educational Studies*, 25, 68–83.
[Online: <http://www.directions.usp.ac.fj/collect/direct/index/assoc/D1175030.dir/doc.pdf>]
- Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Johnsen-Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th annual conference of the International Group for the Psychology of Mathematics Education*, (Vol. 4, pp. 177–184). Bergen, Norway: Bergen University College Press.
- Shaughnessy, J. M., & Pfannkuch, M. (2002). How faithful is Old Faithful? Statistical thinking: A story of variation and prediction. *The Mathematics Teacher*, 95(4), 252–259.
- Sikkink, P. G., & Keane, R. E. (2008). A comparison of five sampling techniques to estimate surface fuel loading in montane forests. *International Journal of Wildland Fire*, 17(3), 363–379.
- Stephenson, A. G. (1980). Fruit set, herbivory, fruit reduction, and the fruiting strategy of *Catalpa speciosa*. *Ecology*, 61(1), 57–64.
- Stier, S. (2010). Is knowledge random? Introducing sampling and bias through outdoor inquiry. *Science Scope*, 33(5), 45–49.
- Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *The Journal of Mathematical Behavior*, 21(3), 319–337.
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage Publications, Inc.

- Stroupe, D. (2015). Describing “science practice” in learning settings. *Science Education*, 99(6), 1033–1040.
- Svoboda, J., & Passmore, C. (2011). The strategies of modeling in biology education. *Science & Education*, 22(1), 119–142.
- Tobiessen, P., & Werner, M. B. (1980). Hardwood seedling survival under plantations of scotch pine and red pine in Central New York. *Ecology*, 61(1), 25–29.
- Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal*, 12(2), 147–169.
- Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal*, 8(1), 32–61.
[Online: [https://iase-web.org/documents/SERJ/SERJ8\(1\)_Watson.pdf](https://iase-web.org/documents/SERJ/SERJ8(1)_Watson.pdf)]
- Watson, J. M., & Kelly, B. A. (2002). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa (pp. 7-12). Voorburg, The Netherlands: International Statistical Institute.
[Online: https://www.stat.auckland.ac.nz/~iase/publications/1/2a1_wats.pdf]
- Watson, J., & Kelly, B. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal*, 17(1), 24–57.
- Watson, J. M., & Moritz, J. B. (1998). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145-168.
- Wherry, E. T. (1920). Plant distribution around salt marshes in relation to soil acidity. *Ecology*, 1(1), 42–48.
- Wroughton, J. R., McGowan, H. M., Weiss, L. V., & Cope, T. M. (2013). Exploring the role of context in students’ understanding of sampling. *Statistics Education Research Journal*, 12(2), 32–58.
[Online: [https://iase-web.org/documents/SERJ/SERJ12\(2\)_Wroughton.pdf](https://iase-web.org/documents/SERJ/SERJ12(2)_Wroughton.pdf)]

MICHELLE E. FORSYTHE
Education 3045
601 University Dr.
San Marcos, TX 78666