



Dot plots and hat plots: supporting young students emerging understandings of distribution, center and variability through modeling

Jill Fielding-Wells¹

Accepted: 15 June 2018 / Published online: 22 June 2018
© FIZ Karlsruhe 2018

Abstract

An important use of statistical models and modeling in education stems from the potential to involve students more deeply with conceptions of distribution, variation and center. As models are key to statistical thinking, introducing students to modeling early in their schooling will likely support the statistical thinking that underpins later, more advanced work with increasingly sophisticated statistical models. In this case study, a class of 10–11 year-old students are engaged in an authentic task designed to elicit modeling. Multiple data sources were used to develop insights into student learning: lesson videotape, work samples and field notes. Through the use of dot plots and hat plots as data models, students made comparisons of the data sets, articulated the sources of variability in the data, sought to minimize the variability, and then used their models to both address the initial problem and to justify the effectiveness of their attempts to reduce induced variation. This research has implications for statistics curriculum in the early formal years of schooling.

Keywords Statistical model · Statistical modeling · Statistical inquiry · Distribution · Variability · Model eliciting activities

1 Introduction

Statistics at the primary/elementary school level largely consists of constructing and reading variations of column and bar graphs developed from data that is either provided or collected from simple pre-populated options. The majority of student time is spent in constructing and coloring a graph type before *reading* the graph to provide descriptive responses to simple questions. After a time, students move on to learn calculations such as mean, median and mode through tasks that are again procedural, resulting in students having little opportunity to develop understanding of these measures beyond that of the algorithm (Watson and Moritz 2000). A distinction between population and sample is not raised, and thus no consideration is given to the difference between descriptive and inferential statistics.

This approach may harm students' developing statistical understanding as it facilitates a view of graphs as

illustrations, rather than as reasoning tools, and of all data as being able to be addressed using procedural, descriptive techniques, thereby masking the existence and uncertainty of sample data. Although a procedural, population-based approach is easier to teach, it neglects conceptual understandings critical for inference—distribution, center and variability.

Statistics education research highlights two current approaches to developing inferential reasoning in students: exploratory data modeling approaches (EDA) and probability modeling approaches. EDA was developed to shift data analysis from procedural based approaches towards investigative approaches that rely on visualization without assigning probabilities to findings (Makar and Rubin 2018). With EDA, students are introduced to making inferences through experiential activities in which they can see and experience variability and uncertainty; for example, growing samples or comparison of samples and groups (e.g., Ben-Zvi et al. 2012). The EDA approach somewhat weakens the relationship between probability and data analysis (e.g., Pratt 2011). This has given rise to probability modeling approaches that incorporate a stronger probabilistic focus through activities that simulate events and make more apparent the role of

✉ Jill Fielding-Wells
jill.wells@acu.edu.au

¹ Institute for Learning Sciences and Teacher Education,
Australian Catholic University, Brisbane, Australia

chance variation, such as building statistical models (see Konold and Kazak 2008).

Informal statistical inference (ISI) sits somewhere between the two approaches: an EDA foundation with an element of probability incorporated (Makar and Rubin 2018). Makar and Rubin (2018) have identified five principal components of ISI:

1. making claims that extend beyond the data at hand;
2. acknowledging the uncertainty inherent in the claim;
3. using data as evidence for the claim;
4. considering the aggregate (focusing on signal and noise); and,
5. consideration of the context.

ISI has been conceptualized as a means for affording novice students opportunities to be engaged in the more authentic practices of statistics and to engage with the key statistical ideas (Watson 2006) as well as enhancing inferential reasoning and personal meaning making (Doerr et al. 2017).

This study focuses on introducing the underpinnings of an ISI approach to a class of Year 5 students accustomed to a procedural approach. The study described here takes an inquiry approach to focus on key ISI components as identified by Makar and Rubin (2018). The topic of the inquiry was to ascertain the best paper size for building a catapult plane. The paper focuses particularly on students initial modeling of the context and interpretation of the results before moving to their use of the model to make decisions, begin to infer possible outcomes and predict extended outcomes.

This paper commences with an overview of key understandings of distribution, center and variation before discussing relevant existing literature on models and modeling in mathematics and more specifically statistics. After providing the methodology and research context, the results are provided in a chronological format to facilitate an appreciation of the development of the lesson sequence in conjunction with pivotal moments in student learning. Implications for curriculum and statistical learning are addressed along with limitations of the study.

2 Literature

2.1 Key statistical concepts in early statistical reasoning

A fundamental concept for students is the realization that data are needed, can help in making decisions and solving problems, can be collected or generated, and that there are ways of maximizing the quality of data collected. There are also several key statistical ideas that all students need to

understand at a deep conceptual level: distribution, center, and variation (Garfield and Ben-Zvi 2008; Watson 2006).

2.1.1 Distribution

Conceptualization of distribution is quite difficult (Garfield and Ben-Zvi 2007) as reasoning about distributions involves interpreting a complex structure that not only includes reasoning about features such as center, spread, density, skewness and outliers but also involves appreciation of related concepts of sampling, population, causality and chance (Pfannkuch and Reading 2006, p. 4).

Specific problems that students encounter in conceptualizing distribution include tendencies to focus on individual data scores rather than viewing data as aggregate (Wild 2006) and viewing graphs as illustrations rather than as reasoning tools (Wild and Pfannkuch 1999). Ben-Zvi and Arcavi (2001) make the distinction in EDA between students having a *local understanding* of data and data representations as distinct from a *global understanding*. In the former, students focus on an individual or small number of values within a data set or representation. In the latter, they recognize an aggregate view of data and are able to search for, describe, and explain more general patterns by either visual observation of the distribution or through statistical techniques.

Novice students often approach interpreting data distributions *locally*, by locating their contributed data point, identifying high/low values, and/or looking for frequently occurring outcomes (Konold et al. 2015). It is important that students learn to see a data distribution as an aggregate, observing the pattern and shape, rather than considering individual cases (Rubin et al. 2006).

2.1.2 Center

Introducing the notion of ‘center’ is often done through single procedural scores, such as mean, without the development of the conceptual understanding of ‘center’ or spread. Students thus learn to calculate measures of center without knowing what they tell us about the data (Garfield and Ben-Zvi 2007, 2008). Mokros and Russell argue that

until a data set can be thought of as a unit, not simply as a series of values, it cannot be described and summarized as something that is more than the sum of its parts. An average is a measure of the center of the data, a value that represents aspects of the data set as a whole. An average makes no sense until data sets make sense as real entities (1995, p. 35).

One way to enhance visualization of center is through the metaphor of ‘signal among the noise’, with argument that signal and noise, or center and variation, be taught together

to enhance student understanding (Konold and Pollatsek 2002). Another method is to maintain a constant focus on distributions and their shape which, along with use of informal language such as clumps and hills, can also be useful in drawing young students' attention to the key features of 'center' and in enabling them to visualize data sets. Later, this visualization assists students to appreciate possible difficulties in using the mean for non-normal distributions.

2.1.3 Variation

The key understanding that students need to develop is that variation is an inherent part of a data set and is able to be categorized as 'real' (that which is characteristic of the phenomenon) or 'induced' (that which comes about through collection methods) (Wild 2006). Often, statisticians seek to minimize the latter to more accurately identify the former. Distribution serves as the 'lens' through which variation can be viewed (Wild 2006).

Garfield and Ben-Zvi (2007) argue variation is key to statistical reasoning and a necessary focus from the earliest grades. As students develop their appreciation of variation, their early knowledge should be enhanced through opportunities for students to "explore ways of presenting data, especially visually, and to begin making decisions about reasonable and unreasonable variation, and about large and small differences within and between data sets" (Watson 2006, p. 218).

2.2 Models and Modeling

Mathematical models are tools used to make sense of a phenomenon we want to understand (Larson et al. 2013). A model is a construct that represents, for example, structural characteristics or general patterns with a degree of abstractness (Lesh and Harel 2003). Statistical models share commonalities with mathematical models, but with a specifically stochastic focus and function. A statistical model can be considered that which enables the location, explanation or extraction of underlying patterns in data (Graham 2006) and which adopts statistical ways of representing and thinking about the real-world (Wild and Pfannkuch 1999). Through opportunities to generate models and to use them to identify and explain variability, young students can be introduced to, and encouraged to explore, initial conceptions of distribution, center and variation. Having students develop their own models from a problem question with identifiable outcomes may assist in guiding students to use the process of modeling to focus on the signal whereas conducting simulation and trialing may serve to focus the students towards noise (Konold and Pollatsek 2002).

Modeling can be considered the entire process of using such models. This might include: identifying and

deconstructing the problem, designing an approach to gathering information about the problem, mathematizing the problem, building a model, refining a model, using the model to address the problem, predicting from the model, and so forth (Lehrer and Schauble 2010).

The complexities of 'real' statistics and statistical modeling are not realistically within the grasp of young children, and this may explain the avoidance of significant statistical practice in many school curricula (e.g., Australian Curriculum, Assessment and Reporting Authority 2017; Common Core Standards 2010). To facilitate students' access to modeling frameworks that imitate the complexity of statistical practice, researchers are making advances in developing like frameworks that are developmentally appropriate. One such example is that of Lehrer and English (2018) as reproduced in Fig. 1. This framework draws on two essential components of data modeling: the statistical investigation process and the modeling of the variability inherent in the sample generated by that process (Lehrer and English 2018).

The statistical investigation process was described by Wild and Pfannkuch (1999) in their widely adopted PPDAC cycle. The cycle is named for the stages of the statistical investigation:

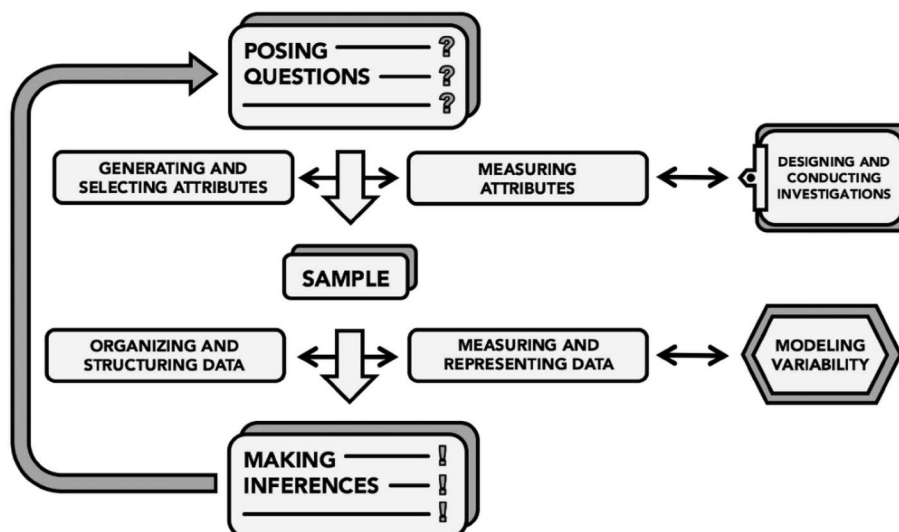
1. Problem defining;
2. Planning the statistical investigation;
3. Data collection and management;
4. Analysis of data; and,
5. Conclusions.

In working through the PPDAC cycle, students can be engaged with in the authentic statistical investigations rather than what Shaughnessy (2007) describes as impoverished views of statistics that emanate from dealing with provided data sets from which the complexity has been removed. In addition to the investigative approach, Lehrer and English (2018) incorporate a specific focus on statistical modeling, using the data generated from the investigation or inquiry, and the making of inferences to address the original problem.

English (2010, 2012, 2013) has demonstrated that young children have the facility to work with mathematical and statistical models at the earliest ages of formal schooling, including developing and applying skills and understandings required to rank and aggregate data, calculate and rank means, and create and work with weighted scores. McPhee and Makar (2014) have shown that key understandings or the "big ideas" of statistics are within the grasp of even very young students.

One aspect of modeling with which students do experience difficulty, is that of making the transition from real-world to statistical model (Crouch and Haines 2004), prompting Noll and Kirin (2017) to express a need for further research that explores students transitioning between

Fig. 1 Data modeling framework (Lehrer and English 2018)



real world and model with both novice and experienced learners.

2.3 Research question

This difficulty in transitioning between ‘worlds’ has led in part to the research question being addressed in this study: How can 10–11 year-old students reason about distribution, center and variation when first introduced to dot plots and hat plots to model an authentic problem?

3 Method

The purpose of this case study was to investigate how young children reason about distribution, center and variation through a modeling context. The study does not aim to assess the instructional design but rather study children’s emerging understandings in their usual classroom setting. This research was undertaken in collaboration with the classroom teacher to implement a statistical inquiry with a single class of children. Case study serves empirical investigation into contemporary phenomena in real-life contexts through the use of multiple sources of evidence (Stake 2006; Yin 2014) as detailed in the following sections.

3.1 Participants

The class involved in the study comprised 26, 10–11 year-old students from a suburban, government school in Australia. The teacher, Ms Thompson, is experienced both as a classroom teacher and in inquiry-based learning in mathematics. Ms Thompson had expressed keen interest in exploring alternatives for teaching statistics and this made her an

ideal choice for this study. She had no preparation for teaching statistics beyond the traditional procedural focus incorporated in her pre-service teaching qualification. Throughout these lessons, the teacher is learning and exploring with the students, developing her own statistical understandings alongside the students.

3.2 Context

This was the third inquiry problem the students had addressed during the year and so the class had developed a culture of inquiry; that is, they were accustomed to dealing with ill-structured problems, providing evidence to support conclusions, and working with uncertainty for extended periods (see Makar et al. 2015).

To provide an authentic context, an arrangement was made with an Air Museum to sell low cost catapult plane kits (a catapult plane is a paper dart plane that is launched using a rubber band) to museum visitors. To achieve this, the students needed to design and package the required materials and calculate costings. In the section of the unit addressed in this paper, students were determining the best-sized piece of paper for making the plane.

In preparation for teaching, an initial lesson was planned to introduce the task. At the end of each lesson, a focus for subsequent lessons was identified. The nature of inquiry as taught by this teacher was to follow on naturally from the students’ lines of thinking, so these foci were broad and responsive. Virtually all during-lesson decisions were made autonomously by the teacher.

As an example, an agreed focus for one lesson was to have students consider ways they could reduce variation that resulted from testing. The teacher addressed that focus by choosing to draw the students’ attention to the data displays

and asking the students why not all the planes had flown the same exact distance. This she did to establish discussion on variation and link variation back to the context.

The class teacher was unfamiliar with statistical modeling and found the Lehrer and English framework (2018) useful to consider key experiences with which the students needed to engage. The students were not introduced to the framework, but the processes described by Lehrer and English progressed naturally. As such, the framework has been used for reporting the progress of the lessons in the Sect. 4.

3.3 Data collection and analysis

Multiple sources of data were collected and analyzed to provide rich detail of the student activities and outcomes, including a classroom observation/field research log, video of the entire sequence, and collection of student artifacts.

The researcher maintained a field research log throughout, incorporating classroom observations. The research log included: general observations; identification of related student and teacher artifacts; and ideas for later discussion with the teacher. On occasion, the log was supplemented retrospectively as the researcher was periodically called upon by the teacher to engage in the lesson, in particular in the use of TinkerPlots™ software (Konold and Miller 2005). Finally, notes of informal pre- and post-lesson discussions with the teacher were recorded as soon as practical after the event to capture concerns, ideas and ‘excitements’ of the teacher.

Each lesson was videotaped in its entirety. During group work, video was obtained of a single group pragmatically selected for the purpose. This was done by eliminating all groups containing a student member from whom video permission was not obtained and then, from the remaining groups, selecting a group on the periphery of the class in order to minimize group noise interference: a constraint of working in an authentic class context. Video analysis was carried out following a process adapted from Powell et al. (2003). The videos were viewed and logged initially by a research assistant, and then reviewed in full by the researcher to develop an overall record of the unit and its progression. Video sections in which the students were engaged with the concepts of center, variation, or distribution were identified and later transcribed. These included both teacher presentation of the identified concepts and student and/or teacher engagement in discussing or representing these concepts. These sections were linked to relevant representations/models in use.

All artifacts associated with the lesson sequence were collected, providing consent to do so had been given by both parent and student. This resulted in a total of 17 sets of student work that were able to be linked to student discussions for cross referencing student understandings.

A typical instruction pattern for this classroom incorporated a problem being posed to the children (for example, *Why didn't the planes all fly the same distance?*), followed by small group and then whole class discussion, after which students recorded their responses independently. An exception to this was related to two children in the class with very low literacy levels. These students worked with a partner to record a joint response for the pair. In each instance, students were asked to provide a claim along with statistical evidence and reasoning for the claim. Student work was analyzed by recording all responses for each activity and developing common clusters. These clusters were categorized as follows: (1) no or incomprehensible response, (2) limited claim that draws on tangential or non-desired evidence, (3) claim only, (4) claim with partial supporting statistical evidence, (5) claim with sufficient supporting statistical evidence, (6) claim with sufficient supporting statistical evidence and reasoning. More specific details are included with the relevant results.

4 Results

For clarity, the unit is reported according to the sequencing of the Lehrer and English (2018) framework in Fig. 1, with a privileging of those activities that illustrate students' engagement with distribution, variation and center.

4.1 Posing questions

Ms Thompson introduced the students to the inquiry by showing them a catapult plane and demonstrating how they fly before explaining the arrangement with the Air Museum. The children's enthusiasm was clear—many already had experience in flying paper planes and began spontaneously contributing ideas. The teacher asked the students to brainstorm what would be required to achieve the task. The class decided they would need to make the kits and complete a budget to price the kits. The kits would need paper, rubber bands and a set of instructions for making the plane. Discussion of the kit contents provided the teacher with the opportunity to ask what size pieces of paper should be included in the kits and this led to the inquiry question being posed: *What is the best-sized piece of paper to make a catapult plane from?*

4.2 Designing and conducting investigations

The students were tasked to make a plan to determine the best-sized piece of paper for making the planes. On being queried, the teacher told the students the internet instructions she had found used a 10 cm × 15 cm rectangle. Working with the teacher to scale the sizes, the students listed

possible planes in the same ratio of 2:3. Their list incorporated whole number increments from 2 cm × 3 cm, to 30 cm × 45 cm. A discussion ensued with students debating how many planes should be tested, with many feeling it should be all 15 possibilities. Two students proposed trying the 10 cm × 15 cm as it had been the recommended size and 20 cm × 30 cm as it was ‘double the size’ (sic) arguing the answer could guide future testing sizes. The class felt this was a practical idea. The class decided that each student would make one of the two sizes (a total of 13 of each) and that five throws of each plane would enable measurement of flight distance before planes became damaged. The students commenced with building and testing their planes and recording their throws on a data sheet provided.

4.3 Generating and selecting attributes and measuring attributes

The students had decided that the attribute of interest was the distance the plane flew, to be measured from the throwing point to ‘the plane’. They commenced data collection and quickly began realized there were a few aspects they had not considered, such as what the throwing point actually was (the student’s foot, the line they were standing behind or the approximate release point of the plane) or what they meant by ‘the plane’: where the plane landed initially or where it finished after sliding? Should they measure to nose or tail? Students were asked to note these for later discussion.

4.4 Sample

Once back in the classroom, the teacher asked the students to identify the data they had collected: two sets of data, five iterations of each of 13 planes of the same size (10 × 15 cm or 20 × 30 cm).

4.5 Organizing and structuring data and measuring and representing the data

The teacher elected to use dot plots to record the data. A dot plot is a simple histogram used for representing small data sets with values that fall into discrete categories (bins). The plot illustrates the distribution of numerical variables with each dot representing a value. Repeat values are stacked above each other so that the height of the column of dots indicates the frequency for that bin (see Figs. 2, 3, 4). The teacher drew the basis of a dot plot on the whiteboard and explained the necessity to bin data across values, setting the bin width at 1 m. The students copied the skeleton dot plot into their grid books (once for each size plane) and then recorded the class data as all students called their five data points (Figs. 2, 3). No student had difficulty beyond accidental recording in the wrong bin.

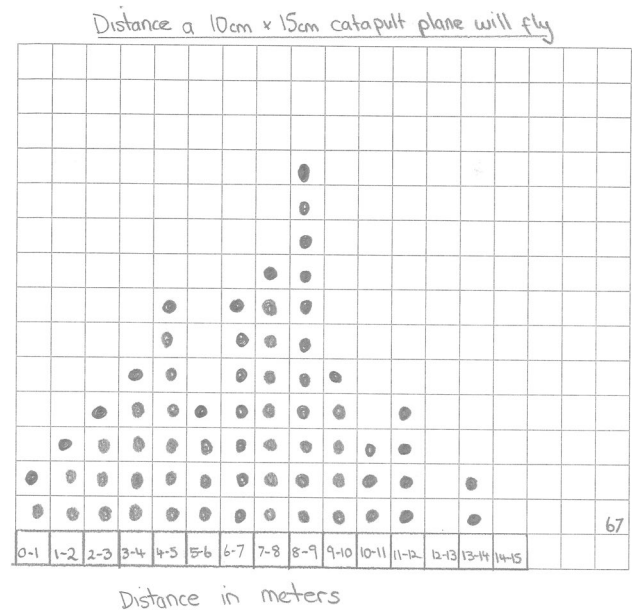


Fig. 2 Data from first round of flight testing of 10 × 15 cm paper sizes

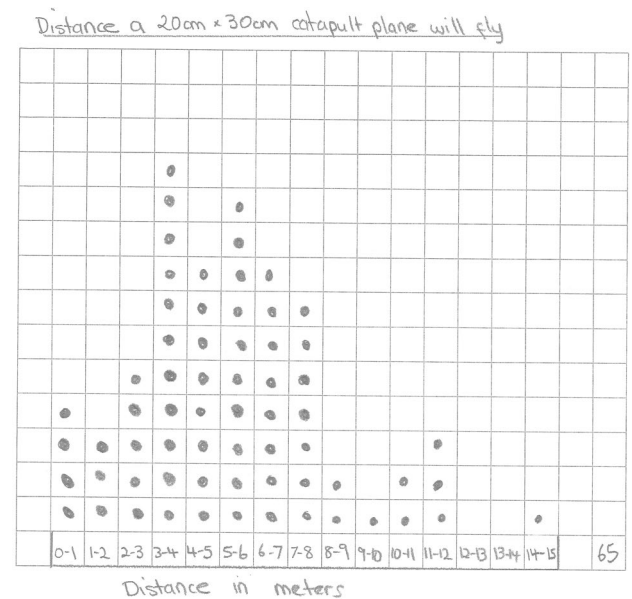


Fig. 3 Data from first round of flight testing of 20 × 30 cm paper sizes

4.6 Modeling variability

Drawing students’ attention to the dot plots, the teacher asked the students what they noted. Student responses largely focused on single data points, such as the greatest distance flown by each plane. The teacher then led a discussion that served to focus the students on the overall distribution of the data before consideration of variation and center. In each instance, the teacher also discussed the

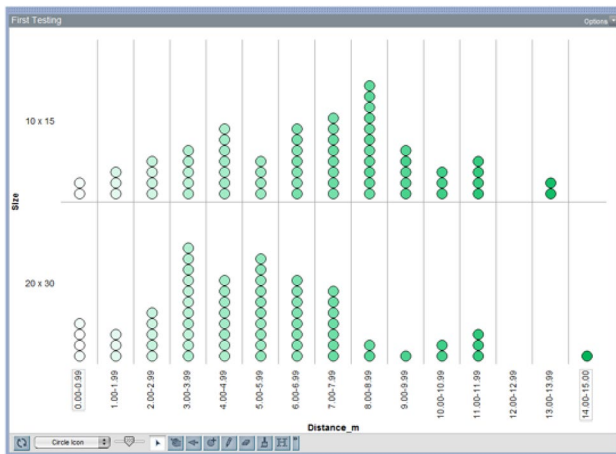


Fig. 4 TinkerPlots™ comparison of 10×15 cm and 20×30 cm flight test data (binned)

representations in terms of the context to ensure students were making the link from the interpretation of the graph to the context of the plane flights. These activities are detailed below.

4.6.1 Exploring distribution

The teacher had the students examine their constructed dot plots while identical plots constructed by the teacher were projected onto a whiteboard for display. The teacher asked the students to look for patterns in the plots, ‘Each graph has a curve or an overall pattern. I want you to look at the overall pattern’. Although unfamiliar with the notion of ‘pattern’ as it related to data, the students quickly observed and identified central ‘clumps’ to each graph. The students noted that there were points some distance from the clumps and a lengthy conversation ensued about the outliers (the teacher introduced the term):

- Patrice ...you know how we were saying [pause] we were talking about [pause] can there be a clump of outliers? There actually could. I think I’m correct um because see the 13 to 14. ... See the 13 to 14 meters on the 10 by 15 graph, that is an outlier but it’s a clump of outliers [the two dots on the far right of Fig. 2].
- Ms T Is it far enough away from the general clump to be an outlier though?
- Patrice Yeah, probably not.
- Ms T What does outlier mean? Why is it an outlier?
- Kala: Because it’s like out of the clump so it’s like far away and it’s like separated from the big clump of all the stuff. ...
- Ms T What does an outlier give you the impression of? So, if you’re looking at data and you see an outlier

what do you immediately think when you see an outlier? Something that is miles away from everything. What do you think?

- Isaac It’s either a really good thing or a really bad thing.
- Ms T What do you think?
- Cindy I think someone’s done something wrong.
- Ms T You think it’s just so far away from everything it’s wrong? It’s a mistake? What do you think Grant?
- Grant It’s unlikely, forget it.
- Ms T What do you think Callum when you see an outlier?
- Callum I think that it’s really lonely.

The conversation above highlights the students’ use of informal language as they begin to express their emerging ideas of distribution, and of any data lying away from the center ‘clump’. Students’ emerging appreciation of outliers is also made apparent through their use of informal language as their own descriptions make their understandings clear. Various student views include outliers as: entities away from the main clump (Kala); extreme values that may be accurate (Isaac); error (Cindy, and Grant); or, simply an isolated score (Callum). Although these views are varied, each of them is a valid possibility and the students were beginning to see that outliers might carry information of interest.

4.6.2 Exploring variation

The teacher introduced the students to a TinkerPlots™ version of the same data (Fig. 4) created by the researcher at the request of the teacher (who knew of TinkerPlots™ but was unfamiliar with the program).

After giving students time to ascertain the generated plot was identical to their hand drawn ones, Ms Thompson drew students’ attention to the data range and attempted to draw out reasoning for the spread of data. She questioned why there was not just the same data for every throw of the one plane type. In this way, she begins to explore the students’ notions of variation.

- Ms T What I want to know is why have you got this great big range? ... why do I have all this [indicating the spread], why don’t I have [just] this range in the middle? [indication drawing all data to a central point]. Why don’t I have one line [data in a single bin] so why, if I say OK it’s going to be like 8 meters, why don’t I just have one line because it flies 8 m? What do we call these things that give us that great big range?
- Janice Mistakes?
- Ms T They could be mistakes. ...
- Isaac Um is it called like human errors?

- Ms T It could be human error. ... Like how they make the plane or how they throw the plane. Yeah Paul?
- Paul You know like on my one I put a hole punch in it to create a drag along it so [voice drops off] ... And some of them put it [the hole] in the sticky tape as well and that would make a little bit of a difference.
- Ms T OK. What do we call all of these things? And we talk about it in science as well. All these things that change all the time. Or that could have created change. Starts with 'v'.
- St: um, variants?
- Ms T Oh, very close.
- Callum Vegetables?
- Ms T These are variables. These are things that happened that might answer the question of why do we have these [indicates outliers] down here?

After this discussion, the student groups brainstormed and then shared back as a class a list of variables that could have influenced flight testing. These included construction of the plane (measurement of the rectangle, folding accuracy, amount of tape used in construction), testing of plane (throwing method, angle of release), and measurement of flight (whether the tape measure was to be shifted to the diagonal, the part of the plane to measure to). One of the students summed this up by stating that *'the first time [they tested the planes] we did cut down some variables to make it a fair test but now we realize it wasn't really the best fair test.'*

4.6.3 Exploring centre

In the following lesson, the teacher moved towards discussions of center. She provided an unbinned distribution (Fig. 5) and asked which was the best-sized piece of paper from which to make a catapult plane. Janice argued that the smaller plane (top plot in Fig. 5) must be the better because there is a clump around 7–9 m whereas the larger plane was

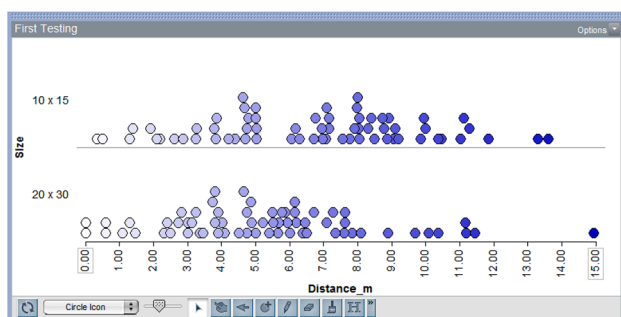


Fig. 5 TinkerPlots™ comparison of 10×15 cm and 20×30 cm flight test data (unbinned)

more clumped around 4–6 m. A few students disagreed, arguing that the spread of the larger plane (bottom plot in Fig. 5) was such that there was not really a clump. Other than Janice, the students could not see a clear difference between the plane data and the teacher became unsure how to proceed and ended the lesson.

In the teacher-researcher discussion following the lesson the teacher stated that, like the students, she could not see a clear difference either. The teacher was aware that another teacher in the school, had used TinkerPlots™ to develop hat plots, and asked if this might help the students. The researcher provided the hat plot data in TinkerPlots™ and the teacher elected to use the idea of 'middle 50%' to move forward.

The following lesson commenced with the teacher's suggestion that finding the 'middle' of the data might assist students. A student familiar with mean, in the context of a cricket batting average, made the suggestion of a mean but she redirected the student by suggesting a middle *range* of the data might be more useful. Ms Thompson challenged the students to work in groups to see if they could identify the middle 50% of the data, handing them printed versions of Fig. 5 with which to work. The students found this an extremely challenging task, due to largely unformed understandings of percentages. Most students commenced by trying to identify halfway to split the data into two lots of 50%, linking their knowledge of 50% with 'half'. Janice was the first in her group to work this out (Fig. 6) and then she verbally summarized her conclusion:

- Janice So, I personally think 10 by 15 is better because it's as you can see, the average is closer to the higher numbers, so where the average for 20 by 15 [cm] [she means 20×30 cm as this is the data she indicates] is a lot lower than the 10 by 15 [cm]. Because it is a whole 2 meters further.

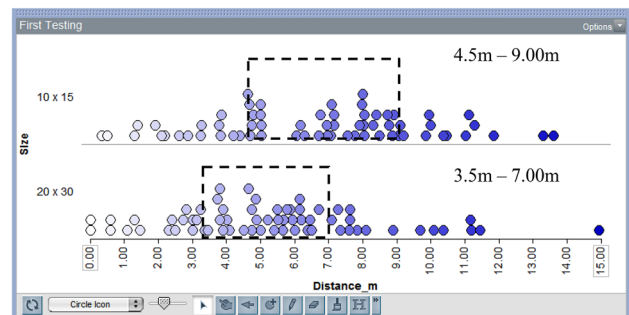


Fig. 6 Student identified center. (Original box and range were hand drawn by the students onto an existing dot plot print out—this has been replicated electronically due to the poor scan quality of the original.)

Most students were now articulating a need to look at the center region rather than individual data points. However, back in this group, Callum was still focused on a single data point. The conversation of the students working with him, as they attempted to explain the need for a more central average, provided some insight into their appreciation of what the data points were representing.

- Callum Look at this. Aha! [pointing to the extreme right data point on the lower diagram in Fig. 6].
- Janice That's not an average that's just how he [Irwin] threw it.
- Isaac That's an average [pointing to the marked 50%].
- Janice That's an average [also pointing to the marked 50%]. One person in the whole class did that [pointing to the extreme right data point]. That's not an average.
- Isaac That's the average of Irwin. [laughs]
- Callum Doesn't matter. It was really good.
- Isaac That's the average of Irwin's second last throw. [still laughing]
- Janice Yeah, this [pointing to the center 50% on the top diagram] is still further though. *It's more typical* to throw further than the 20 by 30 [emphasis added].

This conversation illustrates the difficulty Callum experienced in shifting to a global or aggregate perspective of the data, even as Janice and Isaac try to focus his attention on the center of the data set. Isaac indicates a solid contextual appreciation of the data as he laughingly indicates that a single point is only an 'average' of itself.

The students were shown with TinkerPlots™ generated hats and compared these with the location of their hand drawn hat. Again, the teacher's ordering of the activities was purposeful in ensuring the students knew the origin of the hat rather than simply 'reading' the hat procedurally. The hat plot derived from TinkerPlots™ can be seen in Fig. 7. Not all students had matching plots to those generated by TinkerPlots™; although the drawings were virtually identical, a number of groups demonstrated inaccurate reading. These inaccuracies resulted from either 'guesstimating' the values at the end points of the crown of the hat, or through using a ruler at a skewed angle and thus failing to read the x-axis at the perpendicular. Some of these were out by as much as a meter on the scale.

4.7 Making inferences

To assess the students' understanding of their hats, and to lead the students to begin inferring from data, the teacher asked the students to write a conjecture in their books, based on the data. They were asked which of two planes, one of

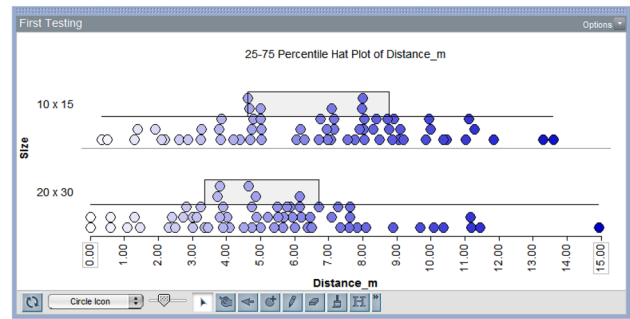


Fig. 7 Comparing TinkerPlots™ hats to students' hat plots

each size, would likely fly the furthest and to provide evidence to explain their response. Of the 17 sets of data, the following responses were noted: three provided no response; one provided an undesired response (referring to the single flight distance of one plane to make a decision); one provided insufficient evidence by identifying the 10×15 cm plane as the better plane but only providing the data for that size and not giving other data for comparison; and, 12 students provided a statement which identified the better size, citing the middle 50% of the flight test results and drawing on these results as evidence. Five of these latter 12 students further referenced the fact that the middle 50% was some form of 'average' through the use of language such as 'typical range', 'middle range' or 'middle clump'; for example,

The 10×15 catapult plane has a typical range of 4.5 meters to 9 meters and the 20×30 catapult plane has a typical range of 3 meters to 7 meters. Therefore the 10×15 catapult plane will probably go further than the 20×30. [Grant]

Although the students had not had any teaching or discussion of inference, in this response we see the student drawing on data, making a generalization about the two planes and the probabilistic nature of that conclusion. Students were not always this clear but typically there was some form of qualifier used to reduce the certainty of their response; for example, 'I think', or 'most of the time'. However, it is unclear whether the students were deliberately qualifying their statements or whether they were merely figures of speech, particularly with 'I think'.

4.8 Posing questions

The students agreed that the smaller plane (10×15 cm) was the better plane. The purpose of this initial testing had been to narrow the choices for a second round of testing and so the students had now to address the question as to what size the best plane was based on a new selection of sizes, thus recommencing the statistical PPDAC inquiry cycle by drawing on their informal inferences.

4.9 Designing and conducting investigations

The students elected to follow the same plan as in their previous experiment, with a focus on reducing variation due to error. They decided to address the variables they had previously identified and see if they could reduce the potential error through more careful design and testing. Each group selected variable/s and carefully discussed each, suggesting means of reducing variation. For example, it was decided that, for each launch, a student would use a ruler to ensure that the rubber band was pulled back precisely 20 cm before launch.

After significant debate, the students decided that the 10×15 cm plane should be retested to make a fairer comparison given the intent to reduce variation. They also established that a 6×9 cm plane was the smallest that could be physically constructed and so decided that 6×9 cm, 8×12 cm, 10×15 cm and 12×18 cm would be the tested

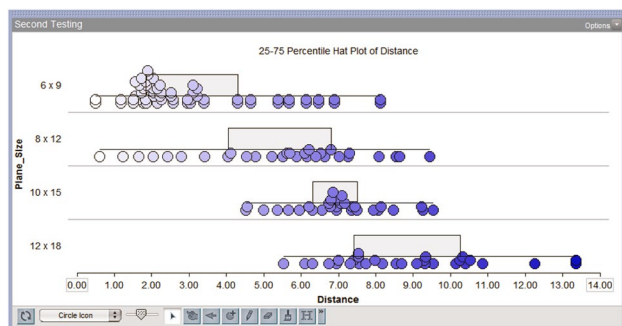


Fig. 8 Data from second round of testing comparing 6×9 cm, 8×12 cm, 10×15 cm and 12×16 cm paper sizes with hat plots

sizes. The measure of success would again be the flight distance, now carefully defined as the distance from the lead toe of the thrower to the closest part of the plane from its final resting place.

4.10 Sample and data organization and representation

The results of the second investigation were recorded and entered into TinkerPlots™ and the students provided with the display in Fig. 8.

The data that resulted from the second round of the investigation were almost disappointing in the extent to which the students were easily able to make a clear-cut response and identify the plane with the overall farthest average flight distance. To ensure that the students were individually interpreting the dot/hat plots, the teacher posed the following two questions to the students:

1. Which was the best [farthest flying] plane? Why?
2. Which was the most consistent plane? Why?

Student responses (Table 1) suggested that many of the students could interpret the hat plots both in terms of identifying the highest average (middle 50%) and the least variation and do so relevant to the context. Few students, however, could explain their reasoning with clarity.

4.11 Making inferences

Two final tasks were given the students to explore further their understandings of the data and context. For the first task, the students were provided with a copy of the hat plot

Table 1 Student responses upon being asked to interpret the TinkerPlots™ generated Hat Plots from Round 2 testing to identify the farthest flying and most consistent planes

Category of response	Farthest flying plane? No. of responses	Most consistent plane? No. of responses	Example
No response	2	3	Not applicable
Incorrect response	1	1	10:15 and 8:12 [are the most consistent] because they are only 3 and 2 cm off each other. [Duane]
Answer only	4	3	12×18 is best. [Irwin]
Partial support	5	3	The 12×18 catapult plane was the best because the middle 50% has 7 to 10 m. [Grant]
Sufficient support	2	2	I think the 12×18 cm paper plane is the best because there is a bigger [higher] range <i>from the others</i> with a range from 5.5 m to 13.5 m. [Darren]
Sufficient support and reasoning	1	2	10×15 is the most consistent plane because all of the data is clustered in the same spot from 4.5 to 9.5 m. There are no outliers. 10×15 is the only ratio with one single big clump with scores over and over again in the same area. None of the data is distant from each other like the other ratios. [Janice]

representation of the Round 2 results (Fig. 8) and asked to predict where the hat might be if the data collection were to be replicated. The students marked this on the existing hat plot sheet. Of the 13 students for whom responses were able to be analyzed, only one student suggested no change would occur. This response suggested that the student was not considering either the real or induced variation inherent in sampling. Of the remaining students, approximately equal numbers suggested: more than $\frac{3}{4}$ overlap of the plots; more than $\frac{1}{2}$ overlap of the plots; and, more than $\frac{3}{4}$ overlap of the plots with clearly less variation on the 10×15 plane (the one exhibiting less variation in Round 2). Noteworthy is that only two students included the full data range in their predicted representations by including the ‘brim’ of the hat in their response. Both students reduced the length of the tails, arguing that variation had been reduced from the first to the second trial and thus they anticipated further reduction as more care was taken to control variation.

The second task was that the students were to decide what the best-sized piece of paper was to make a catapult plane and write a justification for their decision. Most students merely replicated, or elaborated slightly, their response reported in Table 1. Grant felt his previous answer was not convincing enough and he took it upon himself to write the following (unaided). Any writing errors were retained, and the rounded brackets are the work of the student. This has not been included to demonstrate a typical or expected response, but to illustrate the possibilities for statistical understanding and informal inferencing at a young age.

The best sized piece of paper for a catapult plane would be the 12×18 plane because it has a middle 50% of 7.50 m to approximately 10 m. All the other planes had a smaller (lower) middle 50% because their ranges of distance are between 2 and 4.5 m (6×9 plane), 4–7 m (8×12 plane), and 6–7.50 m. We use the middle 50% because it is kind of like an average where half of the planes landed in a cluster and that is easy to read and easy to figure out. ... [provision of outliers for each plane]. We tried to cut down the variables but sometime there was a outlier (measurement of planes dispersed from the cluster). Also if the range was smaller there would be less variation like 10×15 which only has a range of 5 meters. This shows the 10×15 plane was more reliable. The 12×18 would be one of the most unreliable. Overall, with all the things to think about, like variables, outliers and the middle 50%. This has affected my decision, but I still choose the 12×18 paper plane because even though the 10×15 was more consistent, the 12×18 middle 50% (which gives you the

range where the most planes landed) was the largest [highest] range and that gave me a good idea of what is usually thrown.

Again, we see the nature of the informal inference with a generalization made (which is the better plane) the drawing upon of data, and the probabilistic nature of inferences made with the qualifier ‘usually’.

4.12 A final note

Approximately three months after the completion of this unit, the researcher went back to the classroom and asked the students to recall the plane unit they had been working on. She provided the teacher with the TinkerPlots™ diagrams of the first and second trials of the 10×15 cm plane (Fig. 9), which students had not previously seen side-by-side and asked the students if they had any comments to make.

Patrice It looks like on the second one we have definitely cut down a lot more variation. Because it looks like there is quite a big clump in the middle, but the range is not really wide, not like the first one.

Patrice’s comment suggests that, even 3 months later, at least some of the students could interpret the dot plots/hat plots within a familiar context.

5 Discussion

The purpose of this paper was to illustrate possible affordances of statistical modeling for developing young students’ statistical reasoning about distribution, center and variation in the context of an authentic problem.

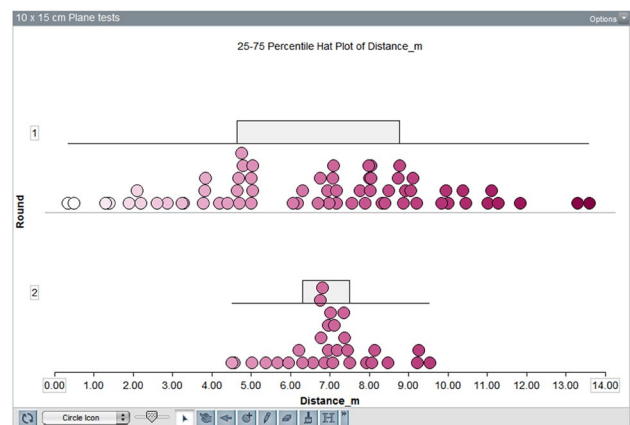


Fig. 9 A comparison between the first and second trials of the 10×15 cm planes

Distribution is difficult for children to conceptualize given the complexity of the structure (Garfield and Ben-Zvi 2007; Pfannkuch and Reading 2006). Although ideally understandings of distribution might take a broader focus than those described here, this paper serves to address *emergent* concepts of distribution with young learners. Key aspects of distribution that were attended to by students included outliers and ‘clumps’ and this was accomplished with little prompting beyond the teacher’s request to find a pattern. Of note was the constant use of informal language throughout the unit, which not only enabled students to explain their emerging understandings but also served to provide greater teacher and researcher insight into students’ developing concepts, as was seen in students’ discussion regarding outliers. An anticipated issue, a tendency for students to focus on individual data scores rather than viewing data as aggregate (Wild 2006), was noted. However, the unwillingness of Calum to relinquish his focus on a single upper score, despite the explanation and persuasion of his classmates, served to highlight the tenaciousness of these views.

To counter procedural approaches to center, which do not provide students with a conceptual understanding (Garfield and Ben-Zvi 2008), the teacher focused the students on informal observation before introducing the ‘middle 50%’. One consideration with using real data, as distinct from neatly planned data, is that it is unlikely the teacher can anticipate the data. In the first collection round, the data scores advantaged the teaching focus, with most of the class having difficulty seeing ‘signals’ in the comparison dot plots. This created a need for further analysis that led to introduction of ‘middle 50%’. Although the students had initial difficulty, it was largely due to unfamiliarity with fractional and percentage concepts. Once understood, the ‘middle 50%’ gave the students a means of identifying the signal in the data (Konold and Pollatsek 2002), resulting in students’ interpretation of the plots. The drawback is that collected data may sometimes be much more clustered which reduces the need for further exploration. Perhaps this is one argument for designing ‘messy’ data sets for students to work with.

As students developed an appreciation of distribution, they began to make conjectures about causes of induced variation (Wild 2006). By attempting to reduce this variation for the re-test, students were able to see the impact of their endeavors on the distribution. This manipulation of the data results, seen through the process of modeling, served to enhance both the students’ appreciation of variation induced through data collection, as well as the notion of data representations as tools that could provide information (*model of* a phenomenon) as well as serve a given purpose (*model for* observing variation) thus countering data representations as illustrations (Wild and Pfannkuch 1999). As an aside, it would be wise to ensure that students engage with

multiple contexts so as not be left thinking that all variation is problematic. In this instance, when the students wanted to identify, for comparison, the center in the four samples, reduction of variation was a legitimate aim as, by doing so, the center of the data became more apparent. Such an approach served to highlight the center of a distribution as being the signal in the noise of variation (Konold and Pollatsek 2002), while acknowledging that variation is an inherent part of a data set.

A key component of these lessons was the students’ engagement with various models; including, multiple dot plot representations of the data, both binned and unbinned, with and without hats and with different comparisons. These models were brought about through the affordances of TinkerPlots™, as the students were able to see graphs manipulated quickly to shift between binned and unbinned data, see hats added and removed, they were able to see how the graphs could be manipulated to give various perspectives of the data. These models thus challenged the ‘graph as illustration’ perceptions (Wild and Pfannkuch 1999) and offered a ‘graphs as tools’ alternative. Previous research has highlighted the potential affordances of technology for offering opportunities to emphasize key concepts of inference (Cobb 2007) and there are many studies that more specifically focus on the affordances of TinkerPlots™ than this study (see Ainley and Pratt 2017; Ben-Zvi and Amir 2005; Doerr et al. 2017; Noll and Kirin 2017).

6 Conclusion

When working with young children, an informal yet conceptual approach would appear to offer far greater potential to develop statistical reasoning than teaching more procedurally. The students engaged in this modeling activity were able to explore and develop some emergent understandings of key aspects of statistical literacy—a conceptual understanding of distribution, variation and center. This supports the contention that young children can address variation and center through modeling at early ages (Garfield and Ben-Zvi 2008) and demonstrates that the capacity to do so is well within their reach.

Issues of replicability and magnitude of this study should be noted as limitations to the research described. Replicability of this study would be problematic as classroom dynamic is individual to each class and teacher. If this study were to be replicated, even with this teacher but a different class, the responses from students may be quite different. This is especially notable in a class that has a strong inquiry culture in which students formulate understandings and ideas collaboratively, and the teacher is responsive to these understandings. However, the goal was not replicability or generalizability but rather to develop an appreciation for how students

might reason about key statistical ideas through modeling an authentic problem. To this end, the results suggest that modeling through statistical inquiry offers potential that procedural views have not similarly demonstrated.

A further limitation to this study is its size. The study provided insight into the possible development of informal and emerging statistical understanding with a single class of Year 5 students. However, there was nothing to suggest development of these understandings could not be commenced earlier, with prior research indicating much younger children can develop early informal statistical understandings (McPhee and Makar 2014). More work with a broader range of ages, using the affordances of statistical modeling, would provide greater insight.

In addition, focused longitudinal studies that result in the development of a continuum of learning for ISI are needed in order to inform curriculum development. This study brings into further question the value of continuing with existing curricula that limit learning to a descriptive and procedural approach until secondary or tertiary levels because young students are not deemed developmentally ready for 'real' statistics.

Acknowledgements THIS work was supported by funding from the Australian Research Council under DP170101993. The author wishes to gratefully acknowledge the contributions of the teacher and students engaged in this research.

References

- Ainley, J., & Pratt, D. (2017). Computational modelling and children's expressions of signal and noise. *Statistics Education Research Journal*, 16(2), 15–37.
- Australian Curriculum, Assessment and Reporting Authority (2017). *Australian Curriculum: mathematics v8.3*. <https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics>. Accessed 23 Nov 2017.
- Ben-Zvi, D., & Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distribution: a collection of current research studies*. Proceedings of the fourth international research forum on statistical reasoning, thinking, and literacy (SRTL-4), University of Auckland, New Zealand, 2–7 July. Brisbane: University of Queensland.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35–65.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM*, 44(7), 913–925.
- Cobb, G. W. (2007). The introductory statistics course: a Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1), 1–15.
- Common Core State Standards Initiative (2010). Common Core State Standards for mathematics. <http://www.corestandards.org/Math>. Accessed 30 Nov 2017.
- Crouch, R. M., & Haines, C. R. (2004). Mathematical modeling: transitions between the real world and the mathematical model. *International Journal of Mathematics Education in Science and Technology*, 35(2), 197–206.
- Doerr, H. M., DelMas, B., & Makar, K. (2017). A modeling approach to the development of students' informal inferential reasoning. *Statistics Education Research Journal*, 16(2), 86–115.
- English, L. D. (2010). Young children's early modeling with data. *Mathematics Education Research Journal*, 22(2), 24–47.
- English, L. D. (2012). Data modeling with first-grade students. *Educational Studies in Mathematics*, 81(1), 15–30.
- English, L. D. (2013). Modeling with complex data in the primary school. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies: ICTMA 13* (pp. 287–299). Dordrecht: Springer.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. Dordrecht: Springer.
- Graham, A. (2006). *Developing thinking in statistics*. London: Paul Chapman.
- Konold, C., Higgins, T., Russell, J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, 88(3), 305–325.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1). <http://escholarship.org/uc/item/38p7c94v>.
- Konold, C., & Miller, C. D. (2005). *TinkerPlots: dynamic data exploration*. Emeryville: Key Curriculum Press.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259–289.
- Larson, C., Harel, G., Oehrtman, M., Zandieh, M., Rasmussen, C., Speiser, R., & Walter, C. (2013). Modeling perspectives in math education research. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies: ICTMA 13* (pp. 61–71). Dordrecht: Springer.
- Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *The international handbook of research in statistics education* (pp. 229–260). Switzerland: Springer International.
- Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In M. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 9–22). Boston: Springer.
- Lesh, R., & Harel, G. (2003). Problem solving, modeling, and local conceptual development. *Mathematical Thinking and Learning*, 5(2), 157–189.
- Makar, K., Bakker, A., & Ben-Zvi, D. (2015). Scaffolding norms of argumentation-based inquiry in a primary mathematics classroom. *ZDM*, 47(7), 1107–1120.
- Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Switzerland: Springer International.
- McPhee, D., & Makar, K. (2014). Exposing young children to activities that develop emergent inferential practices in statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *International Conference on Teaching Statistics (ICOTS9)*, Flagstaff, Arizona, USA. Voorburg: International Statistical Institute.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- Noll, J., & Kirin, D. (2017). TinkerPlots model construction approaches for comparing two groups: student perspectives. *Statistics Education Research Journal*, 16(2), 213–243.

- Pfannkuch, M., & Reading, C. (2006). Reasoning about distribution: a complex process. *Statistics Education Research Journal*, 5(2), 4–9.
- Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *Journal of Mathematical Behavior*, 22(4), 405–435.
- Pratt, D. (2011). Re-connecting probability and reasoning about data in secondary school teaching. In *Proceedings of the 58th World Statistics Conference*, Dublin. <http://2011.isiproceedings.org/papers/450478.pdf>. Accessed on 21 June 2018.
- Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1010). Charlotte: Information Age.
- Stake, R. E. (2006). *Multiple case study analysis*. New York: Guilford.
- Watson, J., & Moritz, J. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, 2(1–2), 11–50.
- Watson, J. M. (2006). *Statistical literacy at school: growth and goals*. New Jersey: Lawrence Erlbaum Associates.
- Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.
- Yin, R. (2014). *Case study research: design and methods* (5th edn.). Beverly Hills: Sage.