



# Characterizing the Ways in Which Young Students Recognise, Describe, Explain, and Employ Variation When Analysing Data in a STEM Context

Jane Watson<sup>1</sup> · Noleine Fitzallen<sup>1</sup> · Suzie Wright<sup>1</sup> · Ben Kelly<sup>1</sup>

Received: 6 August 2021 / Accepted: 12 April 2022 / Published online: 14 May 2022  
© The Author(s) 2022

## Abstract

This report focuses on the ways in which 36 Grade 4 students recognized, explained, described, and employed variation during interviews conducted 1 month after participating in STEM-based activities in which they tested, adjusted, and re-tested catapults. An inductive thematic methodology was used for analysis of the interview transcripts to capture the ways in which students discussed their analyses and justified their conclusions from the activity. The results were based on 1080 instances of variation in student responses to the interview questions, which evidenced three ways students characterized variation: contextual variation, specific variation, and general variation. Findings point to the essential nature of context in building statistical understanding in relation to both specific and general aspects of variation as well as decision-making in that context.

**Keywords** Statistics education research · Variation · Students' experiences · Practice of Statistics · STEM education

## Background

### Variation

Variation in statistical terms is ill-defined. Dictionary definitions (e.g., Kirkpatrick, 1983, pp. 1437–1438; <https://www.merriam-webster.com/dictionary/varia>

---

✉ Jane Watson  
jane.watson@utas.edu.au

Noleine Fitzallen  
noleine.fitzallen@utas.edu.au

Ben Kelly  
ben.kelly@utas.edu.au

<sup>1</sup> School of Education, College of Arts, Law and Education, University of Tasmania, Private Bag 66, Hobart, TAS 7001, Australia

tion) are associated with “variable,” “vary,” and “a range in which a thing varies.” It is problematic that words associated with the etymology of variation are used to define the word. Those words themselves display a range of possible interpretations. Typical descriptions include the following: a change, a difference, a modification, a deviation, and a range. James and James (1959) in their *Mathematics Dictionary* equate variation with dispersion, as “scatteration of the data” (p. 125). Shaughnessy (2007) goes further to describe variation as the measure of variability and elaborates on variability as follows: variability in particular values, including extremes and outliers; variability as change over time; variability as whole range—the spread of all possible values; variability as the likely range of a sample; variability as distance or difference from some fixed point; variability as the sum of residuals; variability as covariation or association; and variability as distribution. Research into young children’s interpretations, experiences, and understanding of variation is limited.

In the history of statistics education, variation was not the earliest concept explicitly considered. In one of the first suggestions of a framework for teaching statistics for students aged 11–16 in England and Wales, Holmes (1980) suggested five stages of a statistical investigation: data collection, data tabulation and representation, data reduction, assigning probabilities, and interpretation and inference. The word “variation” was not included in the expanded description of these five stages. Towards the end of the twentieth century, there was growing interest in teaching statistics at the primary school following the National Council of Teachers of Mathematics (NCTM) publication of its *Curriculum and Evaluation Standards for School Mathematics* (1989), which also had no specific mention of variation for the early grades. In 1997, Friel and Joyner produced a teachers’ manual for the US Grades K–6 including an extensive concept map of the process of a statistical investigation, which briefly mentioned variation. Based on a representation of four corners of a square—pose the question, collect the data, analyse the data, and interpret the results—there were 22 linked boxes indicating the ingredients and processes involved in the process of a statistical investigation. Only one of the 22 boxes included the word “variation”: under analyse the data, descriptive statistics “may include measures of variation.”

About that time, Wild and Pfannkuch (1999) were analysing the work of their applied statistician colleagues. The PPDAC model they developed—Problem, Plan, Data, Analysis, Conclusion—brought planning more to the forefront of an investigation cycle. Although variation was not explicitly mentioned in the PPDAC description, Wild and Pfannkuch included variation in the second dimension of their larger model as follows: Types of Thinking. Here they included “Consideration of variation” as one type of thinking fundamental to statistical thinking. As part of perhaps the most extensive description in the literature, it was characterized as requiring the following: noticing and acknowledging; measuring and modelling for the purpose of prediction, explanation, or control; explaining and dealing with; and investigative strategies (p. 226). Earlier, Moore (1990)

went so far as to focus on aspects of variation in four of his five core elements of statistical thinking as follows: (1) the omnipresence of variation in processes; (2) the need for data about processes; (3) the design of data production with variation in mind; (4) the quantification of variation; and (5) the explanation of variation (p. 135).

The need for specific research on students' understanding of variation was one of three major issues raised by Shaughnessy (1997) in an early keynote related to statistics education. Since that time, there has been an enormous growth in research on the topic, acknowledging both its foundational necessity and its application at all stages of statistical investigations. Recognizing the omnipresence of variation across all aspects of a statistical investigation, Watson (2009a) suggested the representation of a statistical investigation at the school level as seen in Fig. 1 and later related it to the Practice of Statistics at school (Watson et al., 2018).

Research into student understanding of variation has ranged from general large-scale surveys related to the implementation of curriculum innovations (e.g., Shaughnessy et al., 1999; Watson & Callingham, 2003; Watson et al., 2003; Zawojewski & Shaughnessy, 2000), to reports of classroom interventions (e.g., Bakker, 2004; Petrosino et al., 2003; Reading, 2004), to detailed focus on one or two students grappling with the concept during an activity (e.g., Ben-Zvi, 2004), to comparison studies of two classroom interventions (e.g., Noll & Shaughnessy, 2012), to long-term learning outcomes from classroom interventions (Watson & Kelly, 2002a, 2002b, 2004), to interviews with teachers (e.g., Peters, 2011) or pre-service teachers (e.g., Makar & Confrey, 2005), to focused individual student interviews while completing an activity (e.g., Lehrer & Schauble, 2002; Reading & Shaughnessy, 2004; Torok & Watson, 2000; Watson, 2009b; Watson & Kelly, 2005). As well, when interviewing pre-service teachers, Makar and Confrey focussed on the vocabulary used to articulate notions of variation used as the teachers compared two distributions presented in graphical form. Much of this research about understanding of variation has focused on utilising interviews and surveys as research methods. It has demonstrated that

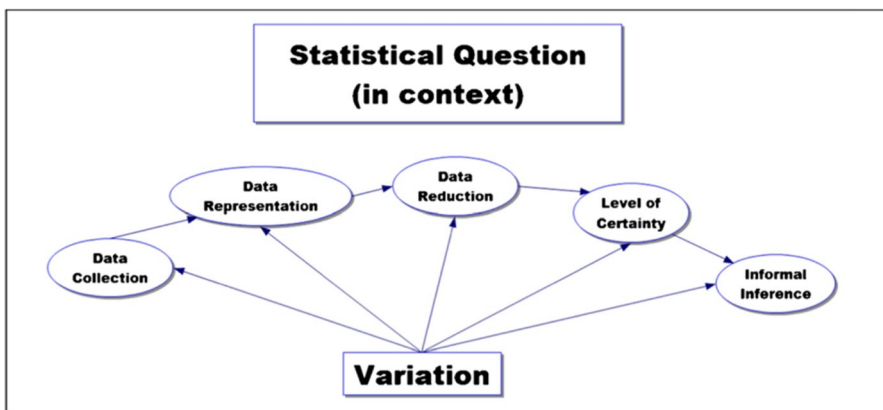


Fig. 1 The Practice of Statistics at school (Watson et al., 2018, p. 108)

young students can express ideas about variation and some studies have reported the development of those ideas within hierarchical frameworks (Watson & Callingham, 2003; Watson et al., 2003, 2007). Little research, however, has taken place explicitly linking variation and STEM topics or been carried out with students reflecting on extended structured learning experiences, such as a STEM inquiry. For the most part, studies have described what students are capable of but have not gone as far as conceptualising the post hoc characteristics of student reflections about variation in complex contexts, nor determined the ways in which students reason about variation when making decisions from the data and the language they use to do so.

### Context of the Study: STEM

Recognizing that there can be no variation (or statistics) without context (e.g., Cobb & Moore, 1997), the range of topics that have the potential to be used as the basis for research studies is enormous. The context chosen for the study reported in this paper is based on a STEM (Science, Technology, Engineering, Mathematics) context. As STEM has grown in importance to national economies (e.g., Office of the Chief Scientist, 2013), and has been acknowledged by school curricula (e.g., Australian Curriculum, Assessment and Reporting Authority [ACARA], 2016), the opportunity to find meaningful topics across the school curriculum has expanded greatly.

In recent years, the governmental focus on STEM fields has created great interest across organizations associated with education at the school level, from early childhood (e.g., Early Childhood STEM Working Group, 2017), across the years (e.g., Australian Council for Educational Research [ACER], 2016; Education Council, 2015) with specific focus on investigations (Australian Academy of Technology & Engineering, 2016; Harland, 2011; Rosicka, 2016). The focus of the project reported in this article was to build upon the *Australian Curriculum* (ACARA, 2019) subjects of Science, Digital Technology, Design Technology, and Mathematics with activities over 4 years, illustrating the power of linking the subjects through using statistics to solve interesting problems. A summary of the activities over the 4 years is presented in Fitzallen and Watson (2020) and Watson et al. (2020a). Within all contexts chosen, variation was a key concept receiving attention.

The activity related to catapults, which is the context for this report, was chosen in relation to the topic of Force in the Science curriculum (ACARA, 2019), in particular, to help students appreciate the different aspects of variation that are involved in statistical investigations. It took place in Grade 4 following a Grade 3 activity intended to lay the foundation for the concept itself. In Grade 3, students took part in a hands-on activity making “licorice sticks” from PlayDoh® by two different methods (Watson et al., 2020b). The sticks were to be 1 cm in diameter and 8 cm long, made either by hand or with a PlayDoh® “factory.” Each student made three sticks by each method and weighed them, putting the mass (in grams) on a sticky label. This happened over 2 days with the results of each day displayed on a large stacked dot plot on the wall. Having completed the hand-made sticks first, the teachers were careful to use the same scale endpoints on the plot for the machine-made masses. Although the centers of the plots were nearly the same, the large difference in the

spread of the two plots was the starting context for the discussion of variation with the class. The “Big V” word (variation) was introduced in Grade 3 and discussed through every activity for the remainder of the 4 years.

The catapult activity (Watson et al., 2022) was introduced in Grade 4 to expand students’ appreciation of variation from a context where the expectation of two different conditions (the mass) is the same, to a context where the expectation of two conditions is different. In the case of catapults, the purpose was to test and then retest “modified” catapults to determine if they would launch ping pong balls further. Often when students are introduced to such activities, they focus on the “before” and “after” data values without considering the variation created within the data for each condition. The aim for this activity was to create for children an appreciation of both “middles” and “spread” visually from graphical representations of the data, many years before they were introduced to means and standard deviations via statistical calculations. This was considered a very straightforward way of creating an awareness of potential difference in variation in two samples, a topic that has proved difficult for both students and teachers (e.g., Cooper, 2018; Vermette, 2016).

## Research Approach and Questions

This report concerns an activity carried out in the second year of a 4-year longitudinal project based in STEM contexts for Grade 3 to 6 (8–12 years old). Although the research approach for the entire project was in some aspects design-based (Cobb et al., 2003), using the outcomes from earlier activities to inform the planning and execution of later activities, each individual activity followed a pragmatic paradigm (Mackenzie & Knipe, 2006) in being problem-centered and oriented to real-world practice. Specifically, as noted above, an extension of the Grade 3 activity, where the expectation (mass) of a product created by two methods was the same, introduced a different expectation (distance travelled) associated with two methods of launching ping pong balls from catapults. As well, the data analysis software *TinkerPlots* (Konold & Miller, 2015) was introduced, extending students’ repertoire for graphing data. For this activity, in keeping with Wiliam (2011), the formative thinking of the students is considered in Watson et al. (2022), where the details of the classroom implementation of the activity were presented with analysis of responses to workbook questions based on the application of neo-Piagetian cognitive learning theory (Biggs, 1992; Inhelder & Piaget, 1958) as developed in the Structure of Observed Learning Outcomes (SOLO) model by Biggs and Collis (1982, 1989). In particular, the characterization of students’ initial learning was presented in terms of two adaptations of the SOLO model. Groth et al. (2021) focused on the Ikonc (IK) mode of development (from approximately 18 months to 6 years) and the Concrete Symbolic (CS) mode (from around 6 years), extending the Unistructural (U), Multistructural (M), Relational (R) analysis of responses from the CS mode to the IK mode. This was useful with the hands-on nature of the experience with catapults. As well, multimodal functioning (Biggs & Collis, 1991) occurred with IK support for many CS

responses. Given the diversity of the expectations of the workbook questions, IK responses ranged from 4 to 100%, CS responses, from 2 to 41%, and multimodal responses, where relevant, from 8 to 59.2% (Watson et al., 2022, p. 17).

This report considers the summative learning (Wiliam, 2011) a month after the completion of the activity. The research was underpinned by the version of constructivism and situated cognition espoused by Cobb and Bowers (1999). This theoretical perspective acknowledges learning is situated, in this case, in the mathematics classroom and learning is both individually and socially constructed. As Cobb and Bowers assert,

A situated perspective on the mathematics classroom sees individual students as participating in and contributing to the development of the mathematical practices established by the classroom community (cf. Cobb & Yackel, 1996). From this point of view, participation in these communal practices constitutes the immediate social context of the students' mathematical development. (p. 5)

They also go on to say that analyses within these contexts “attend to qualitative differences in individual students' reasoning” (p. 5) and the choice of unit of analysis is a pragmatic one that should be based on the purpose of the research. Therefore, in attempting to reflect on students' individual experiences and thinking expressed after the implementation of a classroom learning experience, a qualitative interpretative approach was employed (Creswell, 2013) that followed an inductive thematic methodology (Gioia et al., 2012).

In this study, the optimal outcome was to investigate the students' reporting of their decision making in relation to the activity, which included the following: comparing groups and drawing informal inferences (Makar & Rubin, 2009), as well as appreciating the vocabulary used (Makar & Confrey, 2005). That is, the overall aim was to determine the way in which the students used their understanding of variation to explain the results of their catapult trials 1 month after the activity was implemented. Hence, the following three research questions were posed:

- (1) What language do young students use, derived from the word “variation,” when reflecting on a STEM activity?
- (2) In what ways do young students *recognise*, *describe*, and *explain* variation when reflecting on the data collection and analysis from a STEM activity?
- (3) In what ways do these students *employ* variation in making comparisons and inferences when reflecting on a STEM activity?

## Research Method

### The Student Activity—Catapults

The catapult activity involved launching ping pong balls from pre-fabricated catapults (Fig. 2), making adjustments to the catapults, and launching ping pong balls again to determine if the adjustments improved the performance of the catapults.

The context of improving the performance of the catapults was designed to be a motivating context within which to explore variation and challenge the students' understanding of force. In particular for this activity, the importance of variation was reinforced in a context where expectation in the two parts of the activity was different. It was considered important to use the variation displayed in distributions for each way the launches were carried out, to assist in making a decision about the "improved" performance (expectation) of the catapults.

## Participants

Two classes of Grade 4 students at an urban independent Catholic school participated in the catapult activities with 50 students having parental and student permission for their data to be collected. Of these students, 36 were interviewed 1 month after the completion of the final catapult activity. The 36 students, consisting of 18 girls and 18 boys, with an average age of 10 years 4 months, were those present who had participated in all parts of the catapult activities and were able to be interviewed during the time available when in the school. The project had ethics approval from the Tasmanian Social Sciences Human Research Ethics Committee (H0015039). To maintain anonymity and confidentiality for reporting purposes, students were assigned unique student identification codes (e.g., ID101). Parental and student permission was granted for the publication of student images.

**Fig. 2** A fully constructed catapult



## Procedure

The catapult activity was divided into two sessions implemented five weeks apart. Briefly, the first session included a demonstration that explained the catapult and its construction (see Fig. 2). Then, working in groups of three, the students launched ping pong balls 12 times (four each) from pre-assembled catapults and collected data on the distance travelled each time (Fig. 3), created representations of their group's data (e.g., Fig. 4 and Fitzallen et al., 2018), and analysed their data (see Watson et al., 2022). As part of their introduction to the data analysis software program *TinkerPlots* (Konold & Miller, 2015), students were then given a copy of their group's data represented as shown in Fig. 5, and asked to compare it with their hand-drawn graphs. Discussion of the gaps along the axis led to reasons why some balls had not travelled as far. Later, students were shown the combined data for the class (Fig. 6) and asked for an overall conclusion on how far the ping pong balls travelled, the consistency and variation in the results, and suggestions for improving the consistency (see Watson et al., 2022).

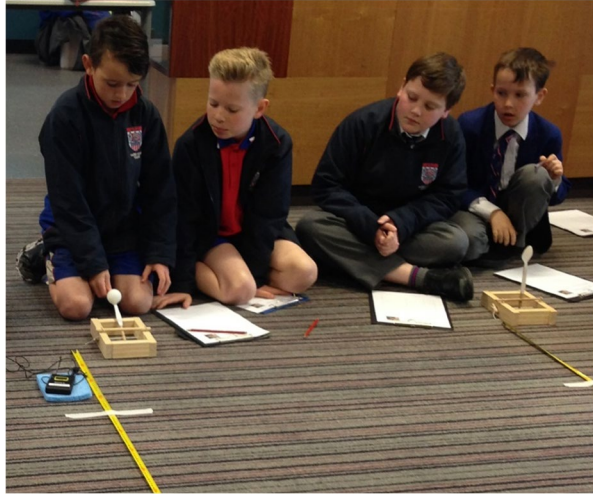
In the second session, the students tested modified catapults to determine if increasing the tension on the string improved the performance of the catapults. Baseline data were first collected using the same method as in the first session. The catapults were then modified by tightening the string holding the throwing arm, and another set of data was collected. The students then used *TinkerPlots* to create plots to answer questions about the range, variation, and typical values arising from the two trials. Finally, the students were shown the combined class data (e.g., Fig. 7) and asked to describe the catapults' improved performances.

## Data Collection

The data collected for the analysis of the ways in which students recognise, explain, describe, and employ variation consisted of transcribed responses to interviews conducted 1 month after the second session of catapult trials. This timing was determined related to the scheduling of school activities at the end of the school year. The aim was to evidence student understanding retained from the activity. It was anticipated interviewing students after the activity would likely elicit recounts of the activity events rather than expressions that conveyed their understanding of variation. Thirty-six students from the 50 students who had been present for both catapult activities were interviewed individually by one of the authors, reviewing the outcomes of the activity. The part of the interview considered here was based on interpreting *TinkerPlots* files with data for a single group and for an entire class in relation to providing evidence for the improvement of the catapults. Questions asked of the students in the interview were related to explanations of variation, including consistency, typicality, and difference between the two trials (see the [Appendix](#)). The interview employed a laptop computer with the *TinkerPlots* files available to the students who could change the plots or give instructions to the interviewer to do so. The purpose was to explore how the students used the *TinkerPlots* data representations to support the exploration undertaken, observing the way in which variation



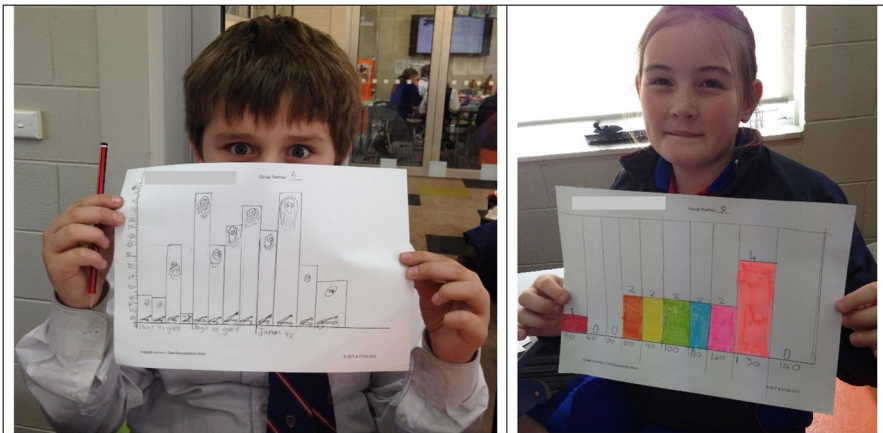
**Fig. 3** Trialling the catapults and recording data



was addressed and utilised during the interview. The part of the interview analysed for this report took on average about 11 min. The interviews were recorded on digital devices. The interview audio data were transcribed verbatim.

**Data Coding and Analysis**

As noted earlier, the learning taking place within the catapult activity was assessed from two perspectives. The formative aspects are reported in Watson et al. (2022), with the summative aspects considered here (Wiliam, 2011). As the summative assessment took place in an interview with a general range of questioning (see Appendix) rather than fixed written test questions, a qualitative analysis was appropriate.



**Fig. 4** Examples of hand-drawn plots representing group data collected

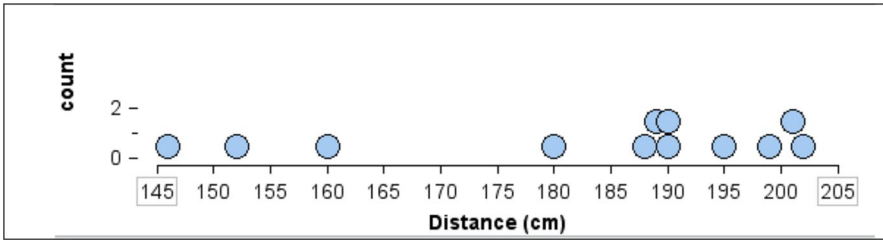


Fig. 5 Example of 12 trials from one group in TinkerPlots

A qualitative content analysis (Kuckartz, 2019) was employed first to identify the students’ explicit use of the language of variation. The aim was to identify the various words the students used to express their ideas about variation and to go beyond only identifying words from a prescribed list of words. The transcripts were then analysed using the inductive thematic analysis of Corley and Gioia (2004) as presented by Gioia et al. (2012). Although Gioia et al. (2012) conducted their research in a potentially more controversial environment of corporate management, it provides a model for other contexts where in the beginning, there may not be an agreement in the vocabulary used for the decision-making being discussed between the interviewer and the interviewee. Several aspects of Gioia et al.’s work in particular suited this study. They found that the theoretical vocabulary of the context employed in one of their studies (Gioia & Thomas, 1996) was not the language used by some of their participants in the actual interviews. Hence, the importance of beginning with identification of the vocabulary/language of the participants was critical (Research Question 1). The second stage of Gioia et al.’s (2012) methodology was to cycle repeatedly through the data to create a meaningful structure that represented the understanding and beliefs of those being interviewed. This structure first reflected what they called “informant terms,” viewed from their interviewees’ perspectives. These became first-order elements when merged into identifiable categories. These were then clustered from the perspective of the researchers into second-order themes related to the purpose of the interviews. Finally, these themes were further combined into aggregate dimensions, creating the basis for a “data structure” to use as a visual aid in resolving answers for their research question/s.

Following the procedure developed by Gioia et al. (2012), two of the authors of this report developed a data structure from the interview transcripts in three stages,

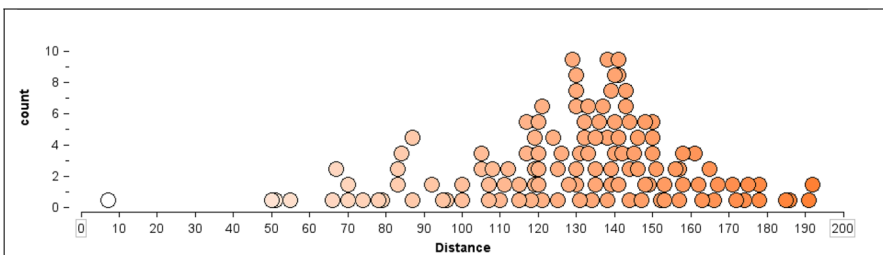


Fig. 6 Data from one class for the initial catapult trials in TinkerPlots

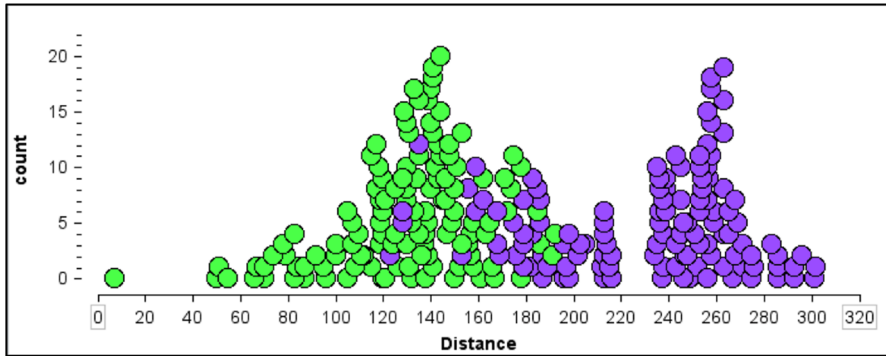


Fig. 7 Class data for comparing original and “improved” catapults in TinkerPlots

with comparison and debate at each stage. They produced many categories of response from the interviews based on terminology employed. These were successively combined to first-order elements and subsequently consolidated into second-order themes, and finally, into aggregate dimensions. This procedure of developing the data structure was based on all comments in the transcripts judged to have any reference to variation and data. Care was taken to identify and categorise the phrasing of responses with the use of derivatives of “variation,” as well as the characteristics of variation thinking suggested by Wild and Pfannkuch (1999, p. 226) (Research Question 1), and the aspects of Moore’s core elements of statistical thinking (1990, p. 135) (Research Question 2). Finally, the transcripts were further explored in a deductive fashion to report the inferences, including comparisons, made by students about the catapult activity (Research Question 3).

## Findings

As reported in Watson et al. (2022), 50 students in two Grade 4 classes took part in this STEM-focused activity linking Mathematics (including measurement and statistics), Science (the topic of force), and Digital Technology (in representing their data in a data analysis software package, *TinkerPlots* [Konold & Miller, 2015]). That report focused on the learning about variation taking place during the activity as reflected in the responses students wrote in their workbooks, responding to specific questions about the variation they were observing. This paper reports on interviews with the majority of the students 1 month after the completion of the activity. As seen in the [Appendix](#), the word “variation” was not used in the interview protocol. It was hence of interest to document the ways in which the concept was employed when answering the general questions about the activity. The “Big V” word is indeed a sophisticated word for Grade 4 students to use casually. Hence, in interviewing the students about their engagement in the activity, many terms were expected to be used for the phenomenon. As well, it was important to recognise the ways in which variation thinking occurred, as Wild and Pfannkuch (1999, p. 226) suggest. Because

of the investigation in which the students took part, there were many possibilities for students to make comments related to the questions in the [Appendix](#).

Many of the students in this study did not use the word “variation” in their responses. Initial analysis found 34 instances where a word derived from “variation” was used (Research Question 1). Further analysis of the interview data following the thematic analysis detailed by Gioia et al. (2012) identified 650 first-order elements, which were deemed to be related to variation. The elements were based on the unfolding of the activity, interaction with the *TinkerPlots* representation, and the language used. These elements were then clustered into seven second-order themes. Some of the initial comments included more than one of these seven sub-themes and were hence counted again, resulting in 966 first-order elements in total. The second-order themes identified in the data were then clustered into three aggregate dimensions related to the fundamentals of dealing with variation: contextual, specific, and general. Contextual variation included *recognizing* specific relevant aspects of the activity as it took place, for example, the second-order themes related to personal hands-on experience with the materials in the activity and with the software *TinkerPlots*. This dimension was specific to the concrete experience of variation and hence related to using the features of the software to manipulate the representations analysed as opposed to identifying features of the data from the representations, which is addressed in the other themes. The second aggregate dimension, specific variation, included *describing* differences observed in relation to the data presented in *TinkerPlots*. The second-order themes were recognizing ranges and outliers, and generally being able to read and interpret the information visible in plots. The third aggregate dimension, general variation, reflected *explaining* the message of variation in the data, particularly based on the second-order themes related to distribution and approximation, including reference to shape, proximity, and uncertainty. This data structure is shown in Fig. 8 (Research Question 2) as suggested by Gioia et al. (2012). Finally, the responses that made actual claims about the outcomes of the activity resulted in 80 explicit comparisons or inferences (Research Question 3). In total, 1080 responses were interpreted through the complete analysis.

### **Research Question 1: What Language Do Young Students Use, Derived from the Word “Variation,” When Reflecting on a STEM Activity?**

Explicit use of a word related to “variation” was not common, found in only 34 cases, making up 3.1% of the total comments analysed. These words were used in three different ways. In some cases, the words were used while physically pointing to examples in a plot to describe what was seen, for example, “they are kind of varied [pointing to right side of graph]” [ID129]. Other students described the phenomenon in words, contrasting it with the opposite phenomenon, as in “... it’s a bit more variety, it’s a bit more spread out” [ID118] or “... that one is variation [Trial 2] and that one is consistent [Trial 1]” [ID115]. Finally, some explained the cause of the phenomenon, for example in the following:

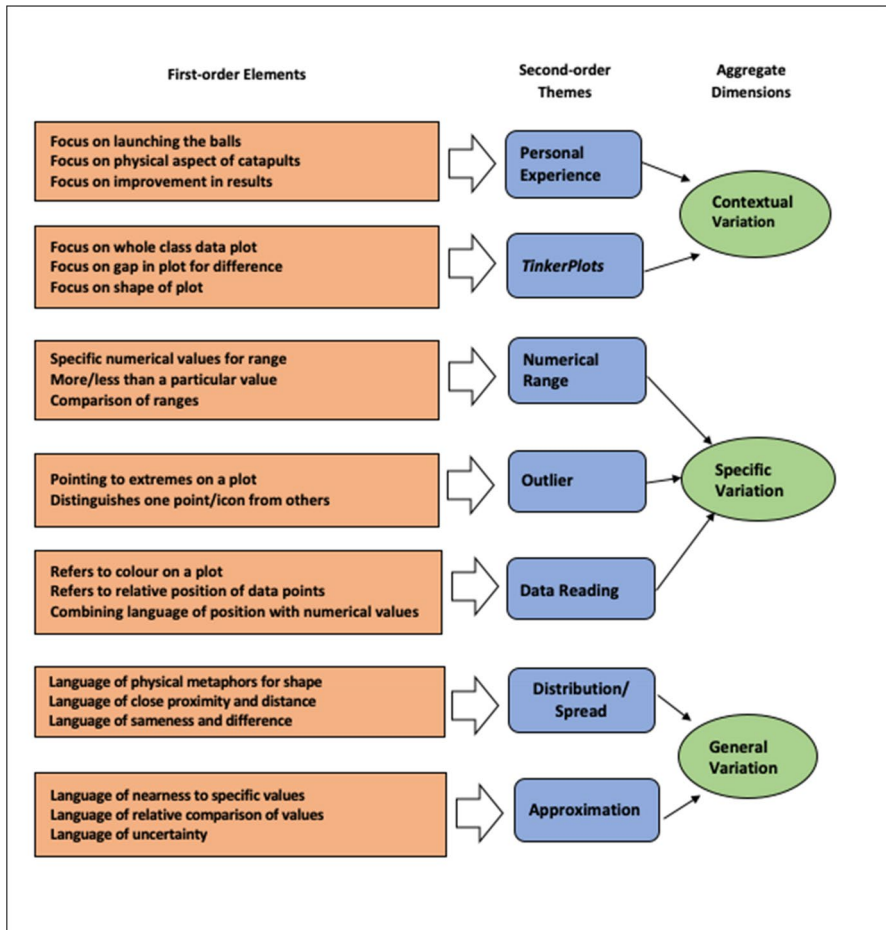


Fig. 8 Data structure for student recognition (contextual), description (specific), and explanation (general) of variation (Research Question 2)

- Well, most of them were bunched up near the 130 and 140 because they were a bit on the side but there’s a long line going up so that showed that most of them were even and that there wasn’t a lot of variety. [ID152]
- Yeah, it has more variation because everybody twisted theirs a bit differently. [ID122]

## Research Question 2: In What Ways Do Young Students Recognise, Describe, and Explain Variation When Reflecting on the Data Collection and Analysis from a STEM Activity?

As illustrated in Fig. 8, examples of the seven second-order themes resulting from analysis of the 966 first-order elements are presented in relation to the three aggregate dimensions related to the application of variation.

### Contextual Variation

Contextual variation related to description of student *recognition* of and interaction with the observable aspects of the environment, explaining either the physical setup of the catapults or features of the *TinkerPlots* representation. This occurred in 14.7% of the responses in the thematic analysis. There were two second-order themes.

#### Personal Experience

The 119 student responses (12.3%) related to the second-order theme of personal experience focused mainly on the recollection of how to use the catapult and how this affected the results. Examples included linking the colors of the icons in *TinkerPlots* displays to the twists of the string or commenting solely on the twists, e.g., "... the orange ones you can tell went higher ... and the ones that we didn't twist they were sort of like low ones, so when we twisted it, they went further" [ID104]. Other responses described the basic variation in people's actions, such as, "When they were turning it, they could have messed it up" [ID135] or "I don't reckon people did exactly the same thing" [ID139]. As well, some described the different action of firing, e.g., "You may not have put it back far enough, or you could have pulled it back too far" [ID124], or the results in terms of direction of the firing, e.g., "Because it's just the way they fired it, I reckon they've pulled it back too far and it's just gone up instead of straight out" [ID113].

Interesting in some of these responses was the informal notion of uncertainty, characteristic of encounters when making decisions with data. It was seen in language such as "they were *sort of like* low ones" [ID104], "they *could have been* messed up" [ID135], or "you *could have* pulled it back too far" [ID124].

#### *TinkerPlots*

The second-order theme related to *TinkerPlots* involved the specific manipulation of the plots used in the interviews and constituted only 2.4% of the cases identified in the thematic analysis ( $n=23$ ). Some students manipulated a plot on the screen themselves, as in discussing the separation or combination of two parts of a data set before and after the changes to the catapult.

- Because we could click the *TinkerPlots* so we could see [clicking sound] ... So, the further ones are probably the second ones [clicking sound]. [ID111]
- If you were talking about both, just drag them back together [student uses mouse pad to amalgamate the line plots]. If you drag them both together it would probably still be that area [pointing to the peak in Trial 1 data.] [ID148]

Otherwise, students focused on gaps or peaks in the data, visible on the screen, accompanied by clicking sounds indicating the utilization of *TinkerPlots* when discussing variation.

- ... and there is usually a gap between the blue and the orange [pointing]. [ID112]
- ... if you drag this one up ... [using mouse] it'll separate them, so, that one's a lot bigger. [ID113]

A few students introduced an extra feature of *TinkerPlots*, the Hat plot.

- The typical distance would be those two parts there [pointing at the two distinct clumps] ... would be almost the same. It's probably why when you put the Hat on [uses mouse pad to demonstrate] it's spread right over those two. [ID113]

## Specific Variation

Specific variation related to students *describing* the interaction with the data and occurred in 28.3% of the responses in the thematic analysis. There were three second-order themes.

### Numerical Range

The numerical range second-order theme, with 105 identified cases (10.9%), included the specific reference to one or two boundaries for one condition or the other, which may not have followed the usual ordering convention.

- From 240 to about 120. [ID104]
- Or to 240 [pointing to a gap in the data for trial 2] Or around about there. [ID129]
- Most of the orange ones from 200 to 300. [ID157]
- For the second one from 60 to 160. [ID134]
- It went from 160 to 320 cm. [ID140]

### Outlier

Although the word “outlier” was not used, there were comments and descriptions of data values satisfying its meaning ( $n = 50$ , 5.2%):

- There is one tiny weeny orange on there. [ID148]
- They have friends with them, and then his friends are all the way over there [pointing to icon]. [ID104]

- That is a lot and this one is really far away. [ID142]
- One from the first one is all the way here [pointing at data point at the start of the graph] and one from the second one is all the way up the top [pointing at the last point on the graph]. [ID116]
- Because there's more here and only three out to the left and none out to the right. [ID111]
- First was blue and second was orange, although this one was *around* there so *someone might have flinged it the wrong way* but it still counted. [ID122] Uncertainty was sometimes expressed in explaining outliers, as in this last response.

## Data Reading

Data reading ( $n=118$ , 12.2%) focused on specific language referring to aspects of the plot and positions of values within it:

- The first ones are behind 170 and most of the second trial ones are in front of it. [ID145]
- Second trial because it's bigger numbers. [ID141]
- That's the highest, 160. [ID126]
- 100 and 120 [in response to "What was the highest?"]. [ID131]
- Well, 20 was the lowest distance we got in the first trial. [ID147]

## General Variation

General variation related to *explaining* the outcomes of the experiment and perhaps including acknowledgement of uncertainty. These responses accounted for 57% of the thematic analysis. There were two second-order themes.

## Distribution/Spread

Comments on distribution/spread constituted the largest category of comments, with 375 cases identified (38.8%). Some students used relatively general language, as in "this is more in the same area and this is spread out" [ID122], whereas others employed visual language, for example, "we've got a bit of a mountain again there [pointing to the stack in trial 1], the mountain was there before so it has gotten bigger" [ID129] or "Because it's kind of in a pyramid there" [ID142]. Some language was similar to the observations of Konold et al. (2002) referring to clusters and clumps. These included the following:

- They're all bunched up so you can't see but with this one [trial 2] they are all sort of spaced out. [ID139]
- You've got all the blue ones here that are quite bunched together, so they were kind of the same, getting repeated answers but we can see up here ... I think they may be a little more spread out, but then they're still quite bunched together. [ID129]



A few students included uncertainty in their discussion of distribution, as in “I think they are both *about* the same but I reckon trial 2 is *probably* a bit more consistent because it’s not as spread out it’s a bit more bunched up” [ID118] and “*Mostly* the first trials will be behind the middle and on the middle but then the second trial *would be* after the middle ...” [ID110].

### Approximation

Approximation was shown in the language of proximity and uncertainty in 176 cases (18.2%) of the thematic analysis:

- A bit hard to tell ... the biggest group is around here [points]. [ID104]
- From 20 so let’s say 120. [ID125]
- For the trial 1, around 110, 100 and 120. [ID149]
- Maybe 260. [ID121]
- You see the blues are around here and there aren’t many up into the 200 to... the 320. [ID148]

### Research Question 3: In What Ways Do These Students Employ Variation in Making Inferences and Comparisons When Reflecting on a STEM Activity?

Making inferences and comparing groups refers to decisions related to the first and second trials (context) rather than just making observations about the data appearing in the plots that were the prompts used in the interviews (see [Appendix](#)). The justifications provided *employ* references to different types of variation. There were 80 such statements (7.4% of the total comments analysed), often including the word “because” along with reasoning based on the observable results. Of these 42 (52.5%) appeared to be justifying a conclusion or inference about the activity.

- I think the first one because it is stacked really well [pointing at peak in trial 1] but with the second trial it is all bunched out [pointing at the data in trial 2] so there is consistency so which you can see with the hat [pointing at the peak in the data trial 1] so I think I made the right decision with the first one. [ID128]
- Because in the second trials they are more spread out but they’re still close ... and in the first trials they’re more together and squished up. [ID125]
- You can kind of tell because the first trial was closer to zero [pointing towards the left-hand side of the graph] the second one was closer to the higher thing [pointing towards the right-hand side of the graph] The highest it could go. [ID105]
- I think the top one [points to the one group data] because there’s less on there and it’s more spread out than the one that has the whole classes mashed together, *it’s really hard* because there’s some spread out and some together and *they are really hard to tell*. [ID104]

Uncertainty is expressed in the final response.

The other 38 (47.5%) statements employed aspects of variation while comparing the two conditions.

- Like with that one it was probably the first trial because it's shorter and then maybe that one at the end will be the last one because we've twisted it more. [ID119]
- When we twisted it to make it stronger for it to sling, it went a lot further and it would just make sense if it was all the way up here [pointing at the orange dots], not all the way down here [pointing at the blue dots]. [ID128]
- Which is funny because the change we made should have made it more consistent, which was what happened in our group, ours was more consistent. [ID113]
- It's *hard to tell* from this one [pointing to the left of trial 2 data] to that one there [pointing to the right of trial 2 data] *maybe* that one [pointing to the right of trial 2 data] because they are kind of closer, then there's a little part, then they're all bunched up together [pointing at the left of the trial 2 data]. [ID129]

Again, uncertainty is expressed in the final response. While comparing the two trials and drawing conclusions and making inferences, notions of consistency were expressed in terms of the physical positioning or spread of the data in the graphs.

## Discussion and Conclusion

Moving from the formative analysis of workbook responses of students while completing the catapult activity (Watson et al., 2022) to the less formal environment of a summative interview a month later allowed the researchers the opportunity to explore the potentially more lasting impact of the activity on students' appreciation of the concepts introduced (William, 2011). In the interviews, the students provided expanded expressions of their understanding of variation that may not have been as evident in their written responses to explicit workbook questions posed when carrying out the hands-on activities. Following the extensive research on various aspects of students' and teachers' understandings of variation over the past 20 years, this study looked in detail at children's appreciation of variation on reflection after participating in a Science-based STEM activity. Similar to Makar and Confrey (2005), although at a different level of statistical sophistication, this study focussed specifically on the language of variation employed for describing the comparison of distributions in the catapult trials. The way in which the students reasoned about the catapult activity came through in the three aggregate dimensions that arose from the interview data. This was shown by their *recognition* of contextual variation through their personal experiences of the activity and using *TinkerPlots*, by their *descriptions* of the specific variation through their ability to discuss numerical values and plots, and by their *explanations* of general variation associated with outcomes and comparisons across the data. This extends findings gleaned about students' understanding of variation evidenced in previous research, for example, by Petrosino et al. (2003), who had students comparing the heights reached by rockets with different nose cones. This research goes further as it includes the *employing* of variation in the statement of the students' final conclusions.

It was acknowledged from the beginning of the project with the students in Grade 3 that the “Big V” word, variation, was tricky to say and remember. Not using the word variation in the interview protocol (Appendix) meant it was not prompted in initial questioning, and as observed, there were only 34 explicit usages of a derivative of the word by the students. The thematic analysis of the transcripts, which demonstrated that the concept was appreciated in three different ways while rethinking the activity, sometimes also displayed subtle hints of uncertainty, as advocated by Makar and Rubin (2009). This is seen in some of the language used in the comments presented in the previous section. For example, in discussing contextual variation, “... they were *sort of like* low ones ...” [ID104], or related to specific variation in, “someone *might have* flinged it” [ID122], and for general variation, “*Mostly* the first trials will be ...” [ID110]. Furthermore, in relation to Research Question 3, some students included their uncertainty as perceived difficulty in drawing conclusions, e.g., “it’s *really hard* because ...” [ID104], or in making comparisons, e.g., “It’s *hard to tell ... maybe ...* because *they are kind of closer*” [ID129].

Following the three stages of the methodology of Gioia et al. (2012) was helpful in characterizing the Grade 4 students’ expressions of variation during their interviews. Similar to the experience of Gioia and Thomas (1996), these students generally used vocabulary different from “variation” to describe their experiences with and understandings of the concept. It was hence considered important to document first their usage of the term before moving on to the detailed thematic analysis of their descriptions of the activity, its results, and the conclusions they reached from it. The results of this inductive analysis can hopefully contribute to deductive hypotheses in future research studies.

In relation to the overall design-based aspects of the 4-year project, the outcomes of this activity point to the continued need to include explicit verbal classroom reinforcement of the word variation and its derivatives, to expand the use of more advanced features of *TinkerPlots*, and to move to STEM contexts where different visual presentations of variation occur (e.g., introducing more variables and relationships). As the complexity of variables within contexts increases, different aspects and comparisons of variation become possible (e.g., Cooper, 2018; Shaughnessy, 2007; Vermette, 2016). Examples of these later contexts are found in Smith et al. (2019), Watson et al. (2021), and Wright et al. (2021), but many more are still needed.

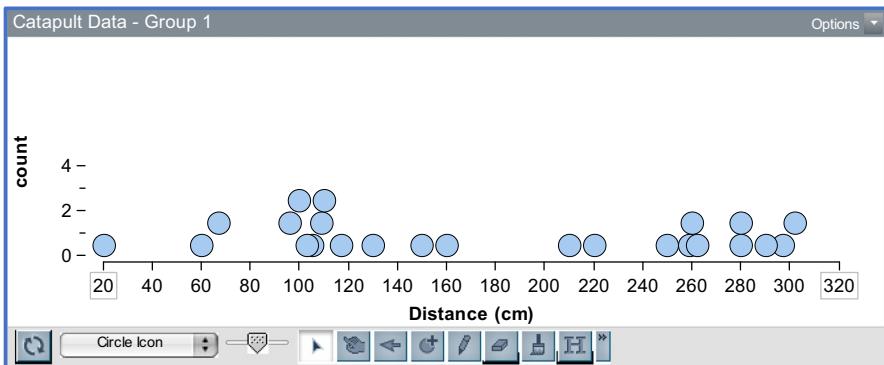
The extended activity with catapults reported here and in Watson et al. (2022) provides another example of the type of research-based practical application of STEM learning that can be implemented in the primary classroom, as called for by Rosicka (2016) and ACER (2016). Focus on student representation of data (Fitzallen et al., 2018) and on the links to the science topic of force (ACARA, 2019) shows the strength of choosing STEM contexts at this level of schooling. Given the fundamental connection of variation to both statistical enquiry and investigations across the Science and other STEM curricula (Watson et al., 2020a), this research suggests four specific aspects of variation to explore when students are immersed in a STEM-based experimental context. Based on the language used, what aspects arise from the context of the experiment? What aspects are related specifically to the data collected? Which aspects are employed when describing the physical features and progress of

the experiment? And what influences the conclusions drawn from carrying out the experiment? Watson et al. (2020a) further illustrate the importance of statistics across the STEM disciplines by demonstrating the links between the “big ideas” of the fields, the contribution statistical literacy can make to STEM literacy, and even the needs of those later choosing STEM careers, which suggest a need to increase enrolments in Science and Mathematics at the school level.

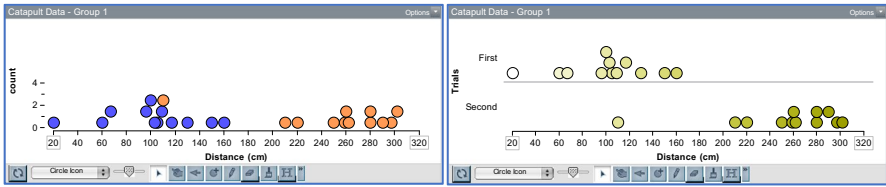
This study has illustrated that the use of the data analysis software *TinkerPlots* (Konold & Miller, 2015) empowers young students to capture the essence of variation from specific data values, spread, and distribution perspectives as described by Shaughnessy (2007). Future challenges, however, include convincing teachers the work required to introduce data analysis software into learning and teaching practices is superceded by the value it adds to student learning (Watson et al., 2021; Wright et al., 2021). Another challenge is determining ways of building student capacity to transfer knowledge gleaned from one context to other contexts. This study, as part of the larger project (Fitzallen & Watson, 2020), has brought to the fore the importance of providing meaningful contexts for the development of understanding of statistical thinking and reasoning within practical investigations and the capacity for young students to be engaged in such investigations. Research, however, is needed to find ways of ensuring students of all ages, make connections among multiple learning experiences, varying contexts, and other statistical concepts.

## Appendix. Student interview protocol

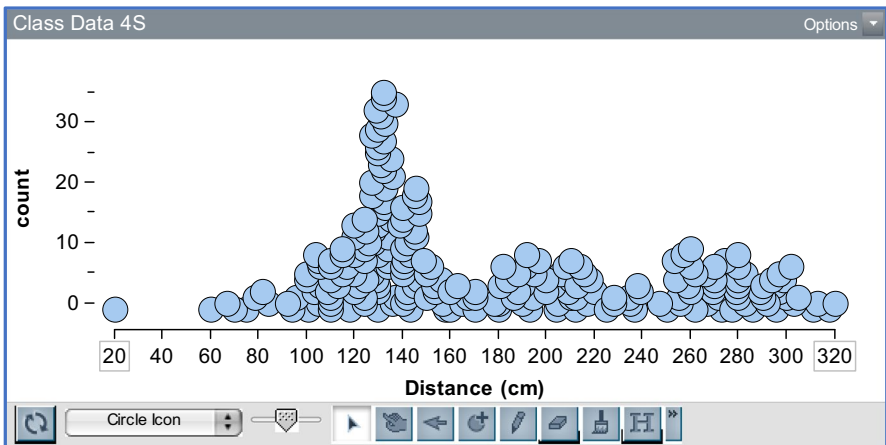
Part 1: Set up a plot from one individual’s data (24 data points) with the data stacked on a continuous scale and no attribute selected (blue dots).



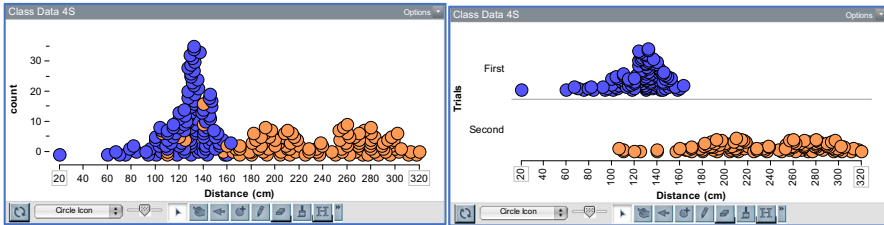
1. How are we going to tell if the catapult improves in the second trial?  
Change the plot as directed by each student or make changes to stimulate thinking and ideas.



2. How can you tell from the plot/s that the changes made improved the catapult?
3. What is the typical distance the ball travelled in the first trial before we changed the catapults?
4. What is the typical distance the ball travelled in the second trial after we changed the catapults?
5. Which data or parts of the graph helped you to decide? [Asked regularly throughout the interview]
  - a. Why did you choose this/these data points?
  - b. Where are the data consistent? or, the same? More than one place?
6. What do you think helped to get those data points to be consistent? or, inconsistent? [Provides the opportunity to link back to the context of the data collection—firing technique and catapult features.]  
 Set up a plot of the class data with a continuous scale with the data stacked on a continuous scale and no attribute selected (blue dots).



7. How are we going to tell if the catapult improves in the second trial?  
Change the plot as directed by each student or make changes to stimulate thinking and ideas.



8. How can you tell from the plot/s that the changes made improved the catapult?  
9. What do you suggest we do to the plots to look at the typical distance for each of the trials?  
10. What is the typical distance the ball travelled? First trial? Second trial? Overall?  
11. Which data or parts of the graph helped you to decide? [Asked regularly throughout the interview]
- Why did you choose this/these data points?
  - Where are the data consistent? or, the same? More than one place?

**Acknowledgements** Authors, Watson, Fitzallen, and Kelly, acknowledge the assistance of Bruce Duncan, who designed the catapult and assisted in the implementation of the classroom activities, and express extreme sadness at the untimely passing of their co-author Suzie Wright.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions. This study was funded as part of an Australian Research Council (ARC) Discovery Grant (DP150100120).

## Declarations

**Disclaimer** Any opinions, findings, conclusions, or recommendations reported are those of the authors and do not necessarily reflect the views of the ARC.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Australian Academy of Technology and Engineering. (2016). *Applied STELR*. <https://stelr.org.au/>
- Australian Council for Educational Research [ACER]. (2016). *Research conference 2016: Improving STEM learning: What will it take?*
- Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2016). *ACARA STEM Connections Project Report, June, 2016*. <https://www.australiancurriculum.edu.au/media/3220/stem-connections-report.pdf>
- Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2019). *Australian Curriculum*. <https://www.australiancurriculum.edu.au/>
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83. [http://iase-web.org/documents/SERJ/SERJ3\(2\)\\_Bakker.pdf](http://iase-web.org/documents/SERJ/SERJ3(2)_Bakker.pdf)
- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63. [https://iase-web.org/documents/SERJ/SERJ3\(2\)\\_BenZvi.pdf?1402525004](https://iase-web.org/documents/SERJ/SERJ3(2)_BenZvi.pdf?1402525004)
- Biggs, J. B. (1992). Modes of learning, forms of knowing, and ways of schooling. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 31–51). Routledge.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.
- Biggs, J. B., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education*, 33(2), 151–163. <https://doi.org/10.1177/168781408903300205>
- Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57–76). Lawrence Erlbaum.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104(9), 801–823. <https://doi.org/10.1080/00029890.1997.11990723>
- Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, 28(2), 4–15. <https://doi.org/10.3102/0013189X028002004>
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13. <https://doi.org/10.3102/0013189X032001009>
- Cooper, L. (2018). Assessing students' understanding of variability in graphical representations that share the common attribute of bars. *Journal of Statistics Education*, 26(2), 110–124. <https://doi.org/10.1080/10691898.2018.1473060>
- Corley, D. A., & Gioia, K. G. (2004). Identity ambiguity and change in the wake of a corporate spin-off. *Administrative Science Quarterly*, 49(2), 173–208. <https://doi.org/10.2307/4131471>
- Creswell, J. W. (2013). *Research design: quantitative, qualitative and mixed method approaches* (2nd ed.). SAGE.
- Early Childhood STEM Working Group. (2017). *Early STEM matters: Providing high-quality STEM experiences for all young learners*. UChicago STEM Education and Erikson Institute.
- Education Council. (2015). *National STEM school education strategy: A comprehensive plan for science, technology, engineering and mathematics education in Australia*. Author. <http://www.educationcouncil.edu.au/site/DefaultSite/filesystem/documents/National%20STEM%20School%20Education%20Strategy.pdf>
- Fitzallen, N., & Watson, J. (2020). Using the practice of statistics to design students' experiences in STEM. In B. Shelley, K. te Riele, N. Brown, & T. Crellin (Eds.), *Harnessing the transformative power of education* (pp. 74–99). Koninklijke Brill NV. <https://doi.org/10.1163/9789004417311>
- Fitzallen, N., Watson, J., Wright, S., & Duncan, B. (2018). Data representation in a STEM context: The performance of catapults. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward*. Proceedings of the 10th International Conference on the Teaching of Statistics, Kyoto, Japan, July 8–14. International Statistical Institute. [https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10\\_4B2.pdf?1531364264](https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_4B2.pdf?1531364264)
- Friel, S., & Joyner, J. (Eds.). (1997). *Teach-Stat for teachers: Professional development manual*. Dale Seymour.

- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Gioia, D. A., & Thomas, J. B. (1996). Institutional identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly*, 41(3), 370–403. <https://doi.org/10.2307/2393936>
- Groth, R. E., Austin, J. W., Naumann, M., & Rickards, M. (2021). Toward a theoretical structure to characterize early probabilistic thinking. *Mathematics Education Research Journal*, 33(2), 241–261. <https://doi.org/10.1007/s13394-019-00287-w>
- Harland, D. J. (2011). *STEM student research handbook*. National Science Teachers Association.
- Holmes, P. (1980). *Teaching statistics 11–16*. Schools Council and W. Foulsham.
- Inhelder, B., & Piaget, C. D. (1958). *The growth of logical thinking from childhood to adolescence*. Routledge and Kegan Paul.
- James, G., & James, R. C. (Eds.). (1959). *Mathematics dictionary*. D. Van Nostrand.
- Kirkpatrick, E. M. (Ed.). (1983). *Chambers 20th century dictionary* (3rd ed.). Chambers.
- Konold, C., & Miller, C. D. (2015). *TinkerPlots: Dynamic data exploration* [Computer software, Version 2.3.2]. Learn Troop.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Developing a statistically literate society?* (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town). International Statistical Institute. [http://iase-web.org/documents/papers/icots6/8b2\\_kono.pdf](http://iase-web.org/documents/papers/icots6/8b2_kono.pdf)
- Kuckartz, U. (2019). Qualitative text analysis: A systematic approach. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 181–198). Springer Open. [https://doi.org/10.1007/978-3-030-15636-7\\_8](https://doi.org/10.1007/978-3-030-15636-7_8)
- Lehrer, R., & Schauble, L. (2002). Distribution: A resource for understanding error and natural variation. In B. Phillips (Ed.), *Developing a statistically literate society?* (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town). [https://iase-web.org/documents/papers/icots6/8b3\\_lehr.pdf](https://iase-web.org/documents/papers/icots6/8b3_lehr.pdf)
- Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, 16(2), 193–205.
- Makar, K., & Confrey, J. (2005). “Variation-talk”: Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27–54. <https://doi.org/10.52041/serj.v4i1.524>
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105. <https://doi.org/10.52041/serj.v8i1.457>
- Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). National Academy Press. <https://doi.org/10.17226/1532>
- National Council of Teachers of Mathematics [NCTM]. (1989). *Curriculum and evaluation standards for school mathematics*.
- 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. <https://doi.org/10.5951/jresmetheduc.43.5.0509>
- Office of the Chief Scientist. (2013). *Science, technology, engineering and mathematics in the national interest: A strategic approach*. Australian Government.
- Peters, S. A. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal*, 10(1), 52–88. <https://doi.org/10.52041/serj.v10i1.367>
- 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. <https://doi.org/10.1080/10986065.2003.9679997>
- Reading, C. (2004). Student description of variation while working with weather data. *Statistical Education Research Journal*, 3(2), 84–105. [https://iase-web.org/documents/SERJ/SERJ3\(2\)\\_Reading.pdf?1402525005](https://iase-web.org/documents/SERJ/SERJ3(2)_Reading.pdf?1402525005).
- 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). Kluwer. [https://doi.org/10.1007/1-4020-2278-6\\_9](https://doi.org/10.1007/1-4020-2278-6_9)
- Rosicka, C. (2016). *From concept to classroom: Translating STEM education research into practice*. Australian Council for Educational Research.
- Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Proceedings of the 20th



- annual conference of the Mathematics Education Research Group of Australasia, Vol. 1, pp. 6–22). MERGA.
- Shaughnessy, J. M. (2007). Research on statistics' reasoning and learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age & NCTM.
- Shaughnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999). School mathematics students' acknowledgment of statistical variation. In C. Maher (Chair), *There's more to life than centers*. Pre-session Research Symposium, 77th Annual National Council of Teachers of Mathematics Conference.
- Smith, C., Fitzallen, N., Watson, J., & Wright, S. (2019). The practice of statistics for STEM: Primary students and pre-service primary teachers exploring variation in seed dispersal. *Teaching Science*, 65(1), 38–47.
- Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal*, 12(2), 147–169. <https://doi.org/10.1007/BF03217081>
- Vermette, S. (2016). Teachers' professional knowledge: The case of variability. *International Journal of Learning, Teaching and Educational Research*, 15(6), 48–60.
- Watson, J. M. (2009a). The development of statistical understanding at the elementary school level. *Mediterranean Journal for Research in Mathematics Education*, 8(1), 89–109.
- Watson, J. M. (2009b). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal*, 8(1), 32–61. <https://doi.org/10.52041/serj.v8i1.456>
- Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46. [http://iase-web.org/documents/SERJ/SERJ2\(2\)\\_Watson\\_Callingham.pdf](http://iase-web.org/documents/SERJ/SERJ2(2)_Watson_Callingham.pdf)
- Watson, J. M., & Kelly, B. A. (2002a). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Developing a statistically literate society* (Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa). International Statistical Institute.
- Watson, J. M., & Kelly, B. A. (2002b). Grade 5 students' appreciation of variation. In A. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 385–392). University of East Anglia.
- Watson, J. M., & Kelly, B. A. (2004). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics*, 57(1), 121–144. <https://doi.org/10.1023/B:EDUC.0000047053.96987.5f>
- Watson, J. M., & Kelly, B. A. (2005). The winds are variable: Student intuitions about variation. *School Science and Mathematics*, 105(5), 252–269. <https://doi.org/10.1111/j.1949-8594.2005.tb18165.x>
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1–29. <https://doi.org/10.1080/0020739021000018791>
- Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning*, 9(2), 83–130. <https://doi.org/10.1080/10986060709336812>
- Watson, J., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 105–137). Springer. [https://doi.org/10.1007/978-3-319-66195-7\\_4](https://doi.org/10.1007/978-3-319-66195-7_4)
- Watson, J., Fitzallen, N., & Chick, H. (2020a). What is the role of statistics in integrating STEM education? In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 91–116). Springer. [https://doi.org/10.1007/978-3-030-52229-2\\_6](https://doi.org/10.1007/978-3-030-52229-2_6)
- Watson, J., Fitzallen, N., English, L., & Wright, S. (2020b). Introducing statistical variation in Year 3 in a STEM context: Manufacturing licorice. *International Journal of Mathematical Education in Science and Technology*, 51(3), 354–387. <https://doi.org/10.1080/0020739X.2018.1562117>
- Watson, J., Fitzallen, N., & Kelly, B. (2021). *Interpolation and extrapolation in Grade 5: Exploring the data without formal mathematics* [Manuscript submitted for publication].
- Watson, J., Fitzallen, N., Wright, S., & Kelly, B. (2022). Characterizing student experience of variation within a STEM context: Improving catapults. *Statistics Education Research Journal*, 21(1), Article 9. <https://doi.org/10.52041/serj.v21i1.7>
- Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>

- Wright, S., Watson, J., Smith, C., & Fitzallen, N. (2021). What makes plants grow best? Using a data-based STEM approach to find out. *Teaching Science*, *67*(2), 37–50.
- Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 235–268). National Council of Teachers of Mathematics.

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