Hypothetical Data Analysis and Representation in Year 4

Noleine Fitzallen
University of New South Wales
n.fitzallen@unsw.edu.au

Jane Watson
University of Tasmania
jane.watson@utas.edu.au

This paper reports on students' experiences of describing and representing variation in hypothetical data. Fifty-six students (8–9 years-old) experienced collecting and working with quantitative data for two years as part of a STEM education project. The task described here was an end-of-year survey question, with three parts about a hypothetical context for surveying students in two different Australian cities: recording the data, describing the potential variation in the data, and creating a representation of what the data might look like when only a descriptive account of the context and variables were provided. The data analysis framework utilised provides a means of determining students' readiness for further development of statistical ideas.

Interest in primary school students' understanding of data, appreciation of the existence of variation in data, and representation of that variation visually has grown tremendously over the past 30 years, as national school curricula have included statistics. Often activities in the early years involved providing students with contexts and data sets considered appropriate for their year level. For example, students were presented with a set of "data cards," each including data for several common variables, which provided a hands-on starting point for representation, as students placed the physical cards on a table in a manner to represent at least one of the variables on the cards (e.g., Watson et al., 1995). This was extended as students were encouraged to create representations on paper for the variables that interested them, with quite diverse consequences (e.g., Chick & Watson, 1998). Because technology has become available more recently, much research has focused on student use of data analysis software to create representations for exploring data (e.g., Fitzallen, 2012). There continues, however, to be interest among early childhood and primary school educators and researchers in hand-drawn representations. Estrella (2018), for example, considered pre-school and Year 2 students' creations of representations as part of early transnumeration, and Leavy and Hourigan (2018) examined the initial representations and creative explanations of 5–6-year-olds when asked to "collect data" while listening to a story about selecting animals for a zoo. Watson (2018) also reported on 6-year-olds' creations of representations with concrete materials and drawings of representations on paper in chance and weather contexts. In each of these studies, data were provided in a context for students to represent in some way, in anticipation of providing a summary or answering a question.

Using data to answer statistical questions is one component of the *Practice of Statistics* (Watson et al., 2018): Pose a Question, Collect Data, Analyse the Data, and Interpret the Results. The above studies created specific opportunities for students to work through the "analyse data" component related to creating visual representations for data they were given or collected. Visual representations are an avenue for displaying the variation that occurs in data, variation that is needed for explanation of the results of the investigation. As part of the context for posing a question and collecting data, each of the studies noted above provided data for the beginning of the analysis to occur. The purpose of the practice of statistics is for it to be applied in contexts where investigators encounter novel ideas to study, which require imagining the question to ask and what the associated data will look like; in particular, identification of the variation that underlies the practice of statistics, as without variation in data, there is no need for the practice!

Earlier research conducted by Fitzallen (2012), asked Grade 4–6 students to draw a graph for any data and context of their choosing. Outcomes from that research suggested that students often address context and data separately and make few connections among relevant contexts and data that may be generated. Moritz (2004) also asked students to speculate about the data represented in

(2023). In B. Reid-O'Connor, E. Prieto-Rodriguez, K. Holmes, & A. Hughes (Eds.), *Weaving mathematics education research from all perspectives. Proceedings of the 45th annual conference of the Mathematics Education Research Group of Australasia* (pp. 203–210). Newcastle: MERGA.

graphical representations. No other studies were found that had the approach of providing a context but no data. There are times when hypothesising about the context and its potential provides the starting point for posing one's questions and deciding on actual data to be collected. English et al. (2017), however, found students experience difficulties writing statistical questions that have the potential to collect meaningful and relevant data.

In the research reported in this paper, interest was in students' capacity to hypothesise in terms of foundational elements of the practice of statistics. Students were provided a context and given the opportunity to imagine the possibilities for recording data, recognising variation, and creating a representation, but given no data with which to do so. To explore this aspect of students' development half-way through a 4-year project when the students were in the final stages of Year 4, it was decided to choose a topic believed to be very familiar to the students but not dependant on or reflected in the topics covered previously in the project. Hence the next section describes the topics students had thus far encountered.

Background on Student Experience

The students in this study were part of a 4-year project from Year 3 to Year 6 integrating data and the informal practice of statistics in STEM (Science, Technology, Engineering, Mathematics) contexts. See Fitzallen and Watson (2020) for a summary of all activities over the four years. Up until the time of the task explored here at the end of Year 4, five other experiences with data representation and variation had been encountered by the students, one in a pre-study survey, and four in classroom activities as part of the project.

In the Year 3 pre-study survey, students were asked a three-part question on data: (a) What do you think "data" means? (b) Give an example of data you have seen or collected. (c) Sketch a graph of the data (Watson & Fitzallen, 2021). The graph responses to part (c) reflected the suggestions of the *Australian Curriculum: Mathematics Version 8.4* for Year 4 (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2018) with 24% being tables/lists, 9% being pictographs, and 67% being column graphs. Nearly half of the students related their representations meaningfully to the examples they had provided in part (b), mostly related to food. This indicated their experiences with varying contexts may have been limited.

The first activity in the study involved making "licorice sticks" with Play-Doh™ two different ways, by hand and with a Play-Doh™ "machine," to introduce the concept of variation (Watson et al., 2020). Again, using the representations suggested in the curriculum, 56% of students were able to represent the three data values for each of the three people in their groups for the hand-made sticks; 81% of students could later do so for the machine-made sticks. At the end of the activity when asked to tell the story of the class investigation in a picture and in words, 21% presented two appropriately labelled graphs and 11% explained the difference in the two methods using a version of the word "variation." On the end of Year 3 survey when asked again about the Play-Doh™ activity, 43% could picture the difference of the two methods of production with 11% labelling the axes, whereas 23% described the difference in variation.

The other activity in Year 3 involved the topic of heat from the Science curriculum with students in groups of three testing the cooling rate of hot water in insulated and non-insulated plastic cups, placed in a cold-water bath (Fitzallen et al., 2017). Temperature (ºC) was the measure of change. For this activity students did not create the graphing format but gained experience in plotting data values on stylised graphs created from images of thermometers, and using the variation seen over time to explain the difference in the way in which the water cooled for the three conditions. The results indicated young students have the capacity to describe differences in variation among and between treatments from graphical representations.

At the beginning of Year 4 the students were introduced to the complete practice of statistics (Watson et al., 2018) to compare their life experiences with those of students in another city: posing and refining questions; collecting data on-line; making and analysing representations of their data in terms of variation and trend; and drawing conclusions, acknowledging uncertainty (Watson et al., 2019). The questions posed were critiqued and refined by the students with help from the teachers and researchers (English et al., 2017), and then 22 questions of six types (e.g., numerical, multiple choice, yes/no) were chosen by the students and answered by 85 students across the two cities. In terms of the representations created, 87% realised the importance of displaying all the data, with 64% including a written summary of the data in the representation. The second activity in Year 4 took place over two terms, carrying out fair tests involving testing catapults. After initial tests the force on the throwing mechanisms was increased and trials were repeated (Watson et al., 2023). In the initial trials, before entering data into the software program *TinkerPlots* (Konold & Miller, 2015), students created graphical representations of their initial trials by hand. At this point, 11% of the representations were considered idiosyncratic, with 15% being tallies, lists, or pictographs. Other representations were either bar or column graphs. Of the remaining column/bar graphs, four sub-types were identified: sequential trial ordering with distance on the y-axis, sequential trial ordering with distance on the x-axis, ordered data with distance on the y-axis, and frequency bar charts in 10 cm intervals along the x-axis. Outcomes from these activities indicated that, given the opportunity to engage in statistical investigations, young students have the capacity to create graphical representation that are meaningful to them, albeit at varying levels of understanding (Watson et al., 2023).

Following the background of students with these highly structured data-based activities and in particular their experience with the complete practice of statistics at the beginning of Year 4, the question arises about student ability to hypothesise about data, variation, and representation in a new imagined context. This leads to the following research question:

- How do students hypothesise about a new context for investigation without data being provided?

Research Approach and Participants

The research adopted a pragmatist paradigm (Mackenzie & Knipe, 2006) to explore young students' capacity to hypothesise in terms of foundational elements of the practice of statistics. The methodology was chosen to capture specifically the students' level of understanding gleaned from responses to a survey question. Following the suggestion of Ballou (2008) that open-ended questions are appropriate to gain insight into how terms are understood and how associated ideas are developed, the three-part task summarised in Figure 1 was developed. The question was included in a survey administered at the end of Year 4, reflecting to some extent the initial activity in Year 4, where students compared lifestyles in their city with another city. Note that the cities named in the task in Figure 1 are the capitals of a state and territory in Australia, and it was reasonable to assume students appreciated their locations on opposite sides of the continent with different climatic conditions and environments for native wildlife. Further, the teachers confirmed that this was a reasonable task to present to the students. Subsequently, data were collected from 56 students in two classes from a parochial school in an Australian capital city, whose parents gave permission for the data to be collected. The average age of the students was 10.5 years, and the gender split was 60% male and 40% female. The project had ethics approval from the Tasmanian Social Sciences Human Research Ethics Committee (H0015039). Each student's work was assigned a code to maintain anonymity.

A Grade 4 class in Melbourne and a Grade 4 class in Darwin were asked to fill in a survey. As part of the survey, they were asked to: ***List three of the native wildlife that are found in your state.*** (a) How would you record the data for this survey item? (b) Describe the variation you might get in the data for this survey item. (c) If you were asked to present the data, what would your representation look like? Use the space below to sketch what your representation might look like.

Figure 1. Task completed at the end of Year 4.

Data Analysis

The data coding scheme for responses to the questions in Figure 1 was informed by the SOLO model of Biggs and Collis (1989) as extended by Groth et al. (2021). The aspects of the model employed here relate to the Ikonic and Concrete Symbolic modes. The Ikonic (IK) mode (from about 18 months) precedes the Concrete Symbolic (CS) mode (from about 6 years). In the CS mode, the basis for beginning learning in school, there are potentially three levels of reasoning: Unistructural (U), where a single element relevant to the task is employed in a response; Multistructural (M), where two or more elements are employed in sequence; and Relational (R), where links are created among two or more elements relevant to the task (Biggs & Collis, 1989). In considering the IK mode, Groth et al. found it useful to distinguish two types of responses: those that are "normative incompatible," IK(ic), with the task, such as myths, superstitions, and subjective ideas out of context, or "normative compatible," IK(c), with the task, such as personal experience, imagery, or intuition in the context but not employing specific elements of the task. Distinction in the IK mode has the potential to identify IK(c) students, who are ready for transition to the CS mode (Groth et al., 2021; Watson & Fitzallen, 2021; Watson et al., 2022). Using the coding scheme described in Table 1, the second author and a research assistant coded the responses to the task separately. Agreement on coding was 84% and differences were agreed by negotiation. Tables 2, 3, and 4 provide examples of responses to each of the three questions in Figure 1, coded using the rubric in Table 1, along with the percentages of response at each SOLO level.

Table 1

Coding Rubric for the Task in Figure 1

Level	Part (a) Record data	Part (b) Variation in data	Part (c) Represent data
IK(ic)	Data unrelated to question	Not addressing variation	No evidence of context
IK(c)	Noting data in context but not recording of data	Variation not related to wildlife context	Contextual drawing without reference to data
CS(U)	Reference to a single type of representation	Single reference to difference in the context	Single aspect of data or variables represented
CS(M)	Reference to more than one type of representation	One aspect of potential difference in the data	Multiple aspects of data represented
CS(R)	NA	NA	Multiple aspects of both variables and data represented

Table 2

Levels of Response for Part (a): Record the Data

Level	How would you record the data for this survey item?	%
IK(ic)	What different foods do they eat? [ID121] I would find out if the places are for warmer or colder areas. [ID149] A lot because a lot escape in zoos. That's why. [ID154]	14.3
IK(c)	Because the wildlife can help you record. [ID159] Write down your animal's name and bring in a photo of it. [ID128] I would collect data from that survey. [ID147]	23.2
CS(U)	I would draw a table. [ID101] You could use a tally. [ID131] A graph. [ID146] I would use TinkerPlots. [ID130]	48.2
CS(M)	A graph or a tally sheet. [ID142] Picture chart with tallies. [ID139] In a column chart or a number chart. [ID111]	14.3

Table 3

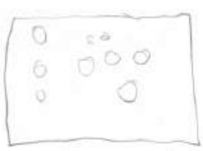
Levels of Response for Part (b): Variation in the Data

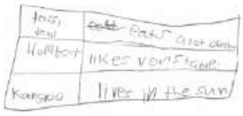
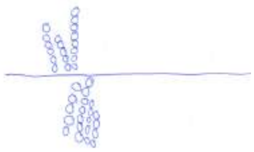
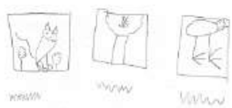
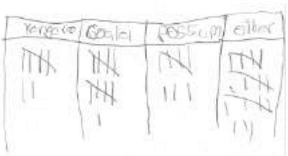
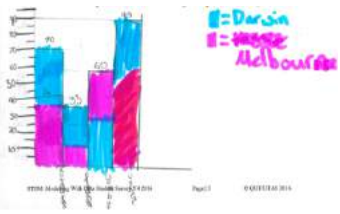
Level	Describe the variation you might get in the data for this survey question*	%
IK(ic)	What one is hotter and what is colder and how many people live there. [ID139] You would go on the computer. [ID147] I am going to make a bar graph [ID155]	13.7
IK(c)	How many different foods each one eats. [ID121] How fast they are. [ID141] Some people might not have native animals. [ID143]	27.5
CS(U)	Darwin might not have the same wildlife as Melbourne. [ID102] I think it would be quite close to each other. [ID111] No one will get the same answer. [ID131] They both might have kangaroos which would be a similarity. [ID158]	33.3
CS(M)	You might get different animals or different habitats. [ID144] Darwin might have more kangaroos than Melbourne or in Darwin they might see cockatoos while in Melbourne they don't see them at all. [ID114] You might only get 1 animal in the group or No animals in a group. [ID125]	25.5

*Five students did not answer this question.

Table 4

Levels of Response for Part (c): Represent the Data

Level	If you were asked to represent the data, what would your representation look like? Use the space below to sketch what your representation might look like.*	%
IK(ic)	 [ID147]  [ID140]  [ID160]	25.0

Level	If you were asked to represent the data, what would your representation look like? Use the space below to sketch what your representation might look like.*	%
IK(c)	 [ID106]  [ID157]  [ID164]	9.6
CS(U)	 [ID154]  [ID128]  [ID148]	15.4
CS(M)	 [ID109]  [ID105]  [ID101]	36.5
CS(R)	 [ID104]  [ID116]	13.5

*Four students did not answer this question.

Discussion

The levels of response in Tables 2, 3, and 4 illustrate the differences in the students' responses when asked to imagine data for a context that was generally familiar to them. For each part of the question, over a third of students responded in the IK mode. In these cases, having the distinction of normative compatible or incompatible was useful in distinguishing those who appreciated the context of the question, a survey on wildlife. Of these IK responses, 47% were IK(ic), and 53% were IK(c), indicating as suggested by Groth et al. (2021), the potential for over half of these students to move to the CS mode. Across the questions, 35% of responses were CS on all three, whereas 18% were all IK. Of interest is the comparison of these results with the results of the pre-study survey question from two years earlier, where students were asked specifically to name and create a graph of some data of their choice (Watson & Fitzallen, 2021), where IK responses were below 20% for all three questions. Perhaps this was related to the opportunity to choose any context to answer the question in conjunction with experience gained from hands-on activities in contexts where they collected, discussed, and represented data.

No other studies were found with results that could be compared directly with those of this study. The studies referred to in the Introduction were analysed with different methods. This study illustrates a way of identifying when student learning moves developmentally from the IK mode to the CS mode. The SOLO levels used here, and by Watson et al. (2022), has the potential to be used as a framework for structuring learning trajectories and developing assessment hierarchies. Young children need to be given opportunities to hypothesise interesting data and create representations to tell the stories related to the context of the data they imagine are in the data. Such opportunities may support them to develop statistical thinking that enhances and goes beyond the outcomes expected from regular classroom activities that usually require students to draw conclusions from data and graphical representations provided (Estrella, 2018; Leavy & Hourigan, 2018).

Practical Implications

The framework displayed in Table 1, provides a means of determining students' readiness for further development of statistical ideas and the research results show that there are benefits to giving young students opportunities to explore, create, and represent data in unconventional ways, even though some students may experience difficulties enacting elements of the practice of statistics when not provided with specific data to analyse. Such opportunities may support students to be able to apply the creative skills needed to imagine the data needed to answer statistical questions, to make projections about how the data may be collected, and to suggest various ways in which the data can be represented, which are all required to address Statistics outcomes in the *Australian Curriculum: Mathematics Version 9* (ACARA, 2022). In Year 1, for example, students are expected to "review data collected and explain how they might change the way they collect data next time." In Year 4, there is the expectation students will engage in the practice of statistics by "constructing graphs of data collected through observation during science experiments, recording, interpreting and discussing the results in terms of the scientific study."

As well as supporting learning in mathematics, opportunities abound to apply the practice of statistics in cross-curriculum activities. As students begin to apply their developing understanding of the practice of statistics in other subjects across the curriculum, particularly in planning project work, the ability to hypothesise about different contexts and how data are to be recorded and represented to show variation is going to become invaluable, particularly in the initial stages of an investigation that requires the posing of a question to explore. At each year level from Foundation to Year 7 across the Humanities and Social Sciences (HASS) curriculum (ACARA, 2022), the content descriptors for Inquiry and Skills includes "Researching," which itself says, "Locate and collect information and data from different sources," including "sources provided," "observations," "primary sources," and/or "secondary sources." Across the Science curriculum content descriptions for the same years require students to be working with data and undertaking experiments that generate data. Also, the Technologies curriculum at Years 3–4 expects students to develop the skills to "recognise different types of data and explore how the same data can be represented differently depending on the purpose" (Digital Technologies), and at Years 5–6, gather relevant data to "evaluate design ideas, processes and solutions" (Design & Technologies). Developing foundational understanding of initiating and following through with a statistical investigation in mathematics in the early years of schooling has the potential to foster student learning across many learning areas of primary and secondary education.

Acknowledgments

The research was funded by an Australian Research Council (ARC) Discovery Grant (DP150100120). The authors thank Ben Kelly for assisting with the data analysis and giving feedback on early drafts of the paper.

References

Australian Curriculum, Assessment and Reporting Authority. (2018/2022). *The Australian curriculum. Version 8.4 and Version 9.0.* https://www.australiancurriculum.edu.au/f-10-curriculum/

Ballou, J. (2008). Open-ended question. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 548–550). SAGE Publications.

Biggs, J., & 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.

Chick, H. L., & Watson, J. M. (1998). Showing and telling: Primary students' outcomes in data representation and interpretation. In C. Kanes, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times*. *Proceedings of the 21st annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 153–160). MERGA.

English, L., Watson, J., & Fitzallen, N. (2017). Fourth-graders' meta-questioning in statistical investigations. In A. Downton, S. Livy, & J. Hall (Eds.), *40 years on: We are still learning. Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia* (pp. 229–236). Melbourne: MERGA.

Estrella, S. (2018). Data representations in early statistics: Data sense, meta-representational competence and transnumeration. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking* (pp. 239–256). Singapore: Springer.

Fitzallen, N. (2012). *Students' reasoning about covariation with TinkerPlots* [Doctoral dissertation, University of Tasmania]. Tasmania: Australia.

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). Brill. https://doi.org/ 10.1163/9789004417311_007

Fitzallen, N., Watson, J., & Wright, S. (2017). The heat is on! Using a stylised graph to engender understanding. *Australian Primary Mathematics Classroom*, *22*(2), 3–7.

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.

Konold, C., & Miller, C. D. (2015). *TinkerPlots: Dynamic data exploration* [Computer software, Version 2.3.2].

Leavy, A., & Hourigan, M. (2018). Inscriptional capacities and representations of young children engaged in data collection during a statistical investigation. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking* (pp. 89–107). Springer.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research, 16*, 193–205. https://www.iier.org.au/iier16/mackenzie.html

Moritz, J. B. (2004). Reasoning about covariation. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–255). Kluwer Academic Publishers.

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

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

Watson, J., & Fitzallen, N. (2021). What sense do children make of "data" by Year 3? In Y. H. Leong, B. Kaur, B. H. Choy, J. B. W. Yeo, & S. L. Chin (Eds.), *Excellence in mathematics education: Foundations and pathways. Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 409–416). Singapore: MERGA.

Watson, J., Fitzallen, N., English, L., & Wright, S. (2020). 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.

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 International Publishing

Watson, J., Fitzallen, N., & Wright, J. (2019). Practicing statistics in year 4. In G. Hine, S. Blackley, & A. Cooke (Eds.), *Mathematics education research: Impacting practice. Proceedings of the 42nd annual conference of the Mathematics Education Research Group of Australasia* (pp. 739–746). MERGA.

Watson, J., Wright, S., & Fitzallen, N., Kelly, B. (2022). Consolidating understanding of variation as part of STEM: Experimenting with plant growth. *Mathematics Education Research Journal.* http://doi.org/10.1007/s13394-022-00421-1

Watson, J., Fitzallen, N., Wright, S., & Kelly, B. (2023). Characterizing the ways in which young students recognise describe, explain and employ variation when analysing data in a STEM context. *International Journal in Science and Mathematics Education, 21*, 943–968. https://doi.org/10.1007/s10763-022-10281-7