Data Interpretation and Representation in Middle Primary: Two Case Studies

Gabrielle Oslington
Macquarie University
gabrielle.oslington@mq.edu.au

Joanne Mulligan
Macquarie University
joanne.mulligan@mq.edu.au

Two case studies of Australian primary school students tracked changes in their data interpretation and representation over three years. Students were engaged in predictive reasoning tasks based on their interpretation of a data table showing temperature change over time. Students' explanations and graphical representations were collected at the beginning of Years 3 and 4 and the end of Years 4 and 5. The first case study was a student mathematically weaker than her peers while case study two was within the average range for her year. Despite differences in starting points, both case studies followed a similar developmental sequence of predicting, interpreting and representing, with the first case generally lagging one stage behind the second case. Similarities and contrasts between the two students are discussed.

Providing rich and complex contexts for data exploration has proved a valuable means for developing statistical literacy in primary school students. Through structured inquiry tasks with a range of possible solutions, students can make predictions and engage in meaningful investigations by making sense of the information provided (English, 2012; Fielding & Makar, 2022; Watson, 2018). In previous reports we have described changes over time of the predictions, interpretations and representations of 44 Australian primary school students when engaged in a single predictive task (Oslington, et al., 2020, 2021). In this paper we focus on the progress of two individual students attempting the same predictive task over a three-year period.

Conceptual Framework

Based on our findings, we assert that as students move through their primary school years, they exhibit an increasing level of structure in their representations and reasoning about data. By increasing structure, we mean that students start to identify and explain general properties and relationships between data sub-sets and this relational understanding will be reflected through their observations and representations. In analysing statistical development, Konold et al. (2015) described students' data observations as moving through a "loose hierarchy" where the student's focal point—or data lens—changed with maturity. Data interpretation requires describing and generalising from aggregate features of data sets, however, younger students often see data as simply a collection of points. Konold et al. (2015) identified four distinct stages or perceptual units described as 'data lenses' through which sets may be viewed: (a) idiosyncratic or unrelated to the data set (b) a single data point or points only (c) similarities between groups of data and finally (d) a wholistic interpretation where the students observed aggregate and variable properties of the set which may include data range, modal clumps, data trends and aggregation. Our primary cohort study indicated shifts in students' focus from idiosyncratic observations towards describing aggregate properties as they moved through the middle primary years (Oslington et al., 2023). We hypothesise that these shifts in data lenses may also be apparent at the individual level when examining changes in students' data interpretations and how they represented similar data sets. Data representation is an important sense making process as it allows students to visualise the structure of the data and in the early years is closely aligned to students understanding of the meaning of the data (English, 2012; Leavy & Hourigan, 2018; Mulligan, 2015). Structural features might include graphing conventions such as collinearity, equal spacing, data sequencing and coordination of bivalent data. We predict that more of these features are likely to be present in the work samples of individual

(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. 387–394). Newcastle: MERGA.

students from the later primary years, compared with those they produced in earlier years. In this paper we focus on our research question:

- How does data interpretation and data representation change in individual students between Years 3 and 5?

The Design Study

The first two iterations of this design study involving 44 primary school participants have been previously described (Oslington et al., 2020, 2021, in press). In this paper we provide a fine-grained analysis of students' shifts in data representation and interpretation through examination of two case studies. These case studies were selected post hoc as individuals who were considered representative of two contrasting ability levels: average and low achieving. The students attended an independent primary metropolitan school from the same year cohort. The school population had a high index of community socio-economic advantage (ICSEA), with 75% of families above the Australian average.

Iris (pseudonym) was 8y 0mo at the commencement of the study, and comparatively less able than her peers in mathematics and language arts. An individual learning assessment conducted at the end of Year 2 indicated that Iris was in the high-average range for IQ (Wechsler Intelligence Scale for Children—WISC-V), but approximately eight months behind peers for mathematics and 12 months behind peers in language arts (Wechsler Individual Achievement Test–WIAT-II). Iris received learning support through small group interventions with specialised teachers in Years 1, 2 and 3, and in-class learning assistance in Year 4. In the National Assessment and Literacy Program (NAPLAN) conducted in Year 3, Iris achieved mid-range Band 4 for mathematics which was slightly below the national average for other Year 3 students, and well below the school average at the top of Band 6. Sophia (pseudonym) was 7y 11mo at the commencement of Year 3. She was the more mathematically able than Iris, achieving low Band 6 for mathematics, slightly below the school NAPLAN average. Sophia had regular classroom placement for mathematics in all years and participated in literacy extension classes in Year 4 and Year 5. Neither student was achieving at the highest or lowest level for their year. Ethical consent for collection of digital recordings and work samples was obtained from participants, carers and teachers.

Highest daily temperature in the month												
Max Temp	**Jan**	**Feb**	**March**	**April**	**May**	**June**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
2010	41	38	31	31	27	21	21	25	27	29	30	31
2011	35	42	34	27	23	21	21	26	33	34	37	27
2012	30	29	27	26	23	20	20	21	24	26	27	29
2013	46	30	34	29	26	22	24	25	32	37	34	36
2014	37	32	30	30	27	23	25	23	34	34	37	32
2015	36	32	36	32	28	23	21	28	30	37	41	35
2016	39	33	31	34	28	22	26	25	25	34	33	38
2017												

Figure 1. Temperature table provided to Year 3 students at the beginning of 2018.

At each of the four data collection points, Iris and Sophia were presented with a stimulus table containing temperatures from past years (Figure 1). They used this table, along with their own remembered experiences of temperature, to complete the blank row which was the year just gone. Students then graphed any aspect of the data they chose, and were subsequently interviewed about their graph, their data observations and their prediction strategies. Data were collected by the first author, as teacher-researcher. Interview prompts included: "tell me about your graph", "did you notice anything special about the numbers?" and "how did you choose your temperatures?" Data consisting of (a) predictions, (b) interpretations (from video) and (c) representations, were collected at four time points: February of Year 3 and Year 4, and November of Year 4 and Year 5. The stimulus table was the same each year, with the exception that each iteration added an additional past year's temperatures. Interviews were recorded using a handheld iPad, and students' responses probed for clarity. Video duration varied from 1min 55s (Iris, Year 3) to 7min 15s (Sophia, February Year 4).

Results

Over the three-year span, Iris and Sophia became more accurate at data prediction, and increasingly used the statistical features of the temperature table when interpreting the data set. In addition, they also developed competence in graphing coordinate data. Table 1 lists the changes in predictions, interpretations and representations over the four data collection cycles. Predictions were considered reasonable if they fell within the 95th percentile range ever recorded for the relevant month, and the values reflected the number of reasonable predictions out of 12 months. Despite similarities, Sophia remained in advance of Iris at each iteration.

Early Year 3

Iris and Sophia's earliest attempts did not draw upon either the column or row structure of the table when predicting. Nevertheless, six of Iris's predictions were reasonable when viewed in the context of the table, as were seven of Sophia's. The students differed in their interpretation and use of the table. While Iris's explanations demonstrated an awareness that months are clustered into seasons and of seasonal change, she held misconceptions about when winter occurred. She described the first few months of the year as the hottest, and the end of the year as the coldest. Her predictive strategies did not include reference to the data table, suggesting that her data interpretation was idiosyncratic. In contrast, Sophia drew upon the data table as a source for her predictions, but at interview revealed she viewed the table as a series of single or disconnected values. For each prediction she described selecting a temperature *not* already used in a column, but present elsewhere in the table. Both students created a grid structure for their representation: Iris represented a table without values—a focus on the gridlines themselves—while Sophia included data by copying the temperature table.

Early Year 4

By early Year 4, both students incorporated data-based strategies when predicting, resulting in 11 (Iris) and 12 (Sophia) reasonable predictions. Iris observed data clusters through describing the vertical columns containing the monthly values "as being actually around the same amount". Sophia went further, describing modal values in columns as "the most common number that has happened". In addition, she utilised her knowledge of the seasons to identify cooling and warming trends across the years. The students' emerging understanding of the subtleties of the temperature table were reflected in their representations. Iris, like Sophia the year before, copied the data table—demonstrating a shift from her earlier focus on the gridlines—to a representation containing temperature data, though with transcription errors. Sophia's Year 4 representation was ambitious, consisting of a line graph of the months January to July with temperature variation on the y-axis and years 2010 to 2018 on the x-axis (Figure 2). This graph included many formal and accurate graphing elements, including a key, equal spacing on the y-axis and a temperature scale starting 20ºC. Sophia competently organised the data in a coordinate arrangement demonstrating her understanding that variation increases with temperature. Despite being able to draw this graph, however, Sophia struggled with its interpretation. When comparing the lines for June and July (which charted months with similar, stable temperatures) with the hotter, more unstable months (November through to March), Sophia described *seasonal* changes rather than variation *within a month*.

Table 1

Data Predictions, Explanations and Representations for Each Iteration

Year	Student	Predictions	Interpretation	Representation
Early 3	Iris	6	Idiosyncratic memories and a false belief about the timing of winter.	Empty grid.
	Sophia	7	Sourced numbers from the data table without using the column structure.	Copy of data table.
Early 4	Iris	11	Used the column structure to find common 10s while retaining false belief in the timing of winter.	Copy of data table.
	Sophia	12	Guided by modal temperature in columns and seasonal change.	Line graph of each month showing variability between months.
Late 4	Iris	12	Used two previous years to predict values close to temperatures, while selecting numbers "a bit different" from ones in the column.	Bar graph of all values in the table January-July.
	Sophia	12	Included multiple features including mode, an average or representative figure and seasonal pattern (winter dip).	Bar graph of all years for two hottest and two coldest months.
Late 5	Iris	11	Used two previous years to predict values close to temperatures. Predictions demonstrated continuity between seasons when compared with previous attempts.	Bar graph of two years used for temperature predictions.
	Sophia	9	Multiple features including seasons, and impact of bushfires, drought and stimulus values. Continuity between seasons when compared with previous attempts.	Line graph of 2010-2017.

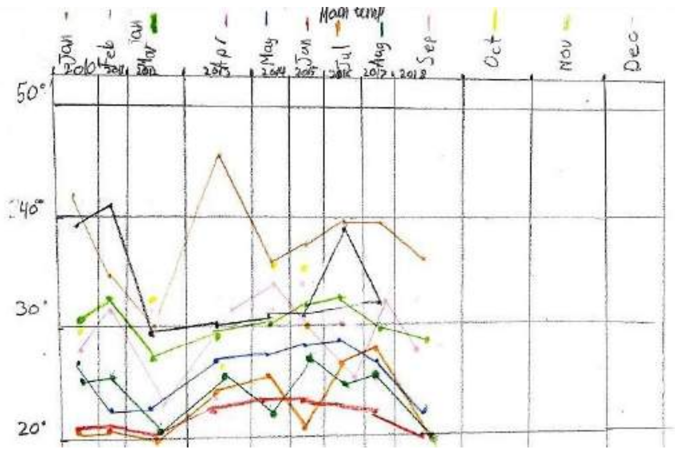

Figure 2. Line graph by Sophia early Year 4.

Late Year 4

By the end of Year 4, both Iris and Sophia predicted reasonable values for every month. Iris's unit of analysis remained the column structure, and she described selecting temperatures similar to the previous two years. Consequently, all her predictions were all within one or two degrees of the previous two temperatures. She justified this strategy by referring to climate change i.e., only the past two years were reliable measures. Iris didn't draw upon the row structure of the graph, nor explicitly link predictions to seasonal change. In contrast, Sophia described multiple data features. At first, she looked for modes: "If there was any kind of a repeat…" it was preferred, and then explained looking for the "average". When probed to describe average, Sophia explained: "what most of them was closest to." Sophia also observed seasonal changes across the row structure of the table, describing a pattern as "…kind of making a dip. At first there are higher temperatures, then it goes lower, and then it goes up."

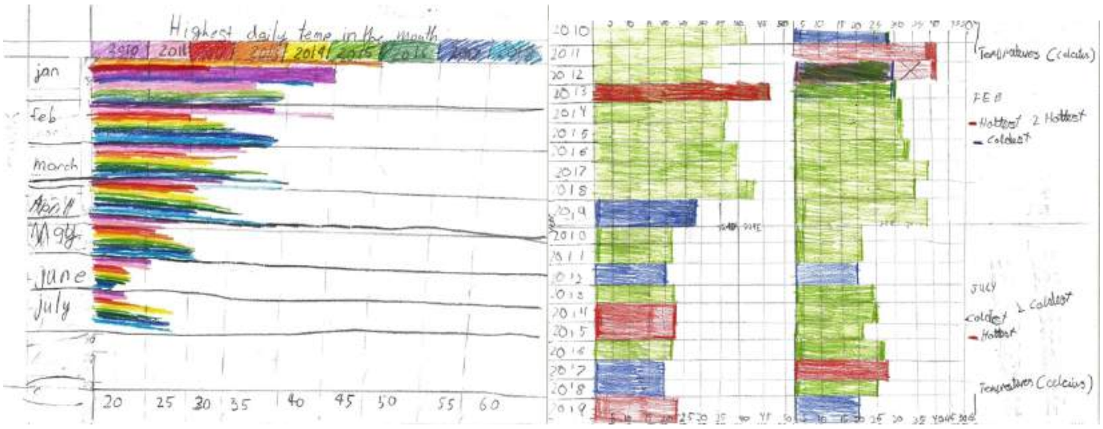

Figure 3. Graphs by Iris (left) and Sophia (right) from late Year 4.

Iris and Sophia both drew bar graphs for their late Year 4 representations, and it was Iris this time who attempted to include all data points. Her graph included formal features such as data labels, a key, a heading, appropriate colour coding, a scale focusing on the range of interest, and almost equal spacing (Figure 3, left). Iris described the graphing process as "confusing", and its planning and construction required making a graph different to any sort she had seen before. Sophia, in contrast, restricted her data representation to four temperature sequences, each one a single graph: January, February, June and July (Figure 3, right). Selecting some values over others led to a simpler, easier to interpret representation when compared to Sophia's former attempts. She included the formal graphing elements of key, labels, equal spacing and intervals of 5 degrees on the temperature scale. However, because Sophia again used the year, rather than the month as the independent variable, her graphs demonstrated differences between months and not seasonal change that occurred over the calendar year.

Late Year 5

In Year 5, Iris's prediction strategy again referenced the two most recent years, also seen in her graph. This graph accurately representing the temperature changes, with a seasonal dip and a consistent scale confined to values above $10^{o\ C}$(Figure 4). Iris's predictions all fell within the historical range except for December which was 45^{o}C. Sophia's predictions also included several overestimations relative to the historical data set, i.e., 48^{o}C for January and 43^{o}C for February. Despite these overestimations, the Year 5 predictions for both students showed continuity over the months i.e., they started as warm in January, declined to a winter dip, and then increased smoothly from August to a hot summer. Sophia accurately marked the seasons on her table and ensured that adjacent months followed appropriate seasonal trends. Eastern Australia experienced some very serious bushfires in January 2020, and Sophia's overestimations were directly linked to her memories of this event. She added that by March, the onset of Autumn was moderating maximum temperatures. The other monthly predictions were linked to their respective season, with the exception of October, where she also noted that her selection (37^{o}C) was also the mode.

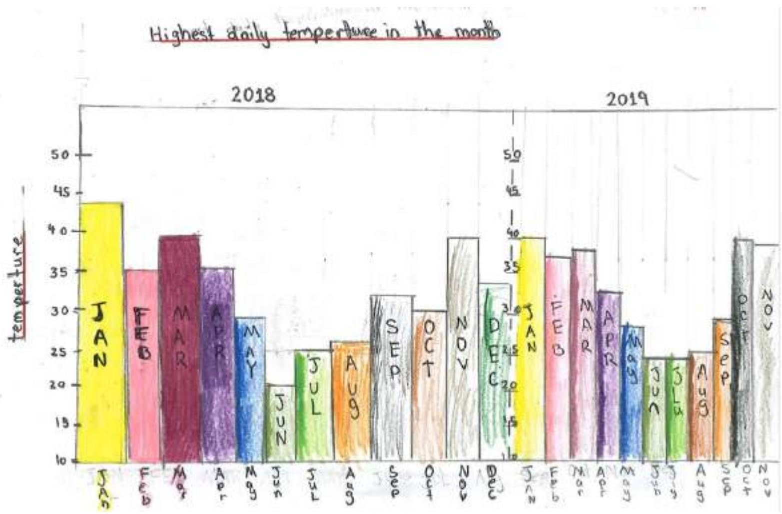

Figure 4. Graph of two previous years of maximum temperatures by Iris in Year 5.

When representing, Sophia returned to her early Year 4 strategy of using a line graph to represent every data point, although this time with more technical accuracy (Figure 5). Sophia chose a line graph "because it is easier to compare data than a bar or a column graph". While it was her intention to graph every year, she only completed 2010-2017 before running out of time. Her graph had an unusual orientation with temperature on the *x*-axis, and she started her scale from zero. She described her representation as "a line graph of the highest temperatures in each month. It shows the differences between the years in temperature, and it also showed the dip in temperature for when it becomes winter." This graph enabled Sophia to express an informal generalisation regarding variability in the monthly maximum temperatures: "They are basically different at the start and end of the year, but they all come quite close in winter".

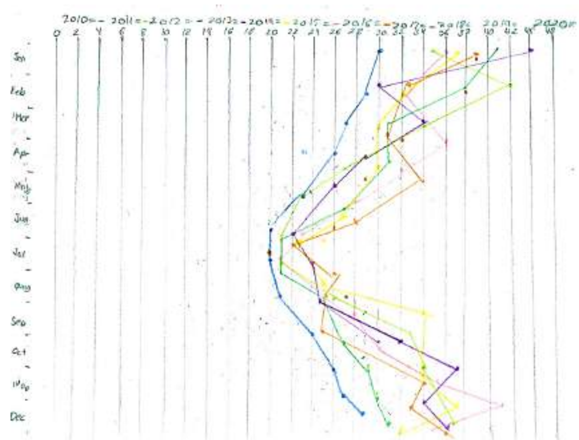

Figure 5. Line graph of monthly temperature changes by Sophia in Year 5.

Discussion

Despite differences between the two cases, common elements emerged in their predictions, interpretations and representations. Iris's Year 3 interactions focused upon non-data components of the task, including her false belief about the timing of winter, even while remaining engaged in the task. Sophia from Year 3, used data-based strategies: i.e., sourcing values already in the temperature table and seeking missing temperatures in columns. The idea that all numbers "should" be included equally (making a flat data distribution) has been reported in other studies of young students, for example Year 2 students predicting lost milk teeth (Ben-Zvi & Sharett-Amir, 2005). It may be associated with the younger child's perception of fairness. Reasonableness of temperature predictions for both students improved significantly between Year 3 and the second data collection point in early Year 4. As they progressed through the years, their observations moved from "non-data based", or the first data lenses perspective (Iris only), to noticing "multiple and relational components" of the table (Konold et al., 2015). While there was a general trend towards more sophisticated data viewing, the lenses used for prediction were not always the same as those in the representations. For example, in early Year 4, Iris copied the whole data table paying attention to individual case attributes (second data lens). However, when she predicted, Iris noted the temperatures "as being actually around the same amount", i.e., recognising similarities between groups of data (third lens). Similarly, in late Year 4, Sophia observed relational components, i.e., seasonal change as a pattern moving from left to right demonstrating an aggregate data view. Her representation, however, was limited to the similarities in four sequences (third lens).

Similarly, for each data collection point, students' interpretations lagged the success of their predictions. Both Sophia and Iris identified strategies such as coldest temperatures at the end of the year, using the same 10s value or selecting a missing number in the table. These strategies were frequently applied inconsistently, and potentially moved from intuitive to explicit reasoning while the interview was in progress. Sophia articulated this in her late Year 4 interview where she explained that for some predictions, she "just got the sense of them". The students' representations followed a pattern, with Sophia approximately one year ahead of Iris at each point. Iris's Year 3 representation of the grid indicates she was paying attention to the physical grid lines, rather than the numerical values. Reading a data table, such as the stimulus table in the study, assumes that students have mastered the column and row construction and appreciate the construct as a spatial array of squares. Research by Battista et al. (1998) with Year 2 students demonstrated that this spatial array is not intuitive for many students. Sophia's focus in Year 3 and Iris's in Year 4 was on the whole data set. This reproduction of the table seems to stem from two factors. The first was simply not knowing what to graph. Prior experience with graphing included gathering information and organising into lists, tables, and picture graphs. For Sophia in Year 3 and Iris in Year 4, these prior experiences were not sufficiently flexible or ingrained to transfer when graphing something new. Second, neither knew what to include or exclude. English (2012) and Mulligan (2015) describe the challenge of data representation as a selection process, thus deciding which features to emphasise over others. The resistance to discarding information continued even as their graphing skills developed. Sophia's early Year 4 and Iris's later Year 4 representations both attempted to include all temperatures, resulting in messy graphs, difficult to interpret. In Sophia's case, her representation hindered, rather than supported her understanding of variation within months, and Iris was unable to read the basic feature of temperatures cooling in the winter. Sophia returned to an 'all values' representation again in Year 5, although this instance, her representation enhanced her interpretation of the data set.

This paper contributes to the growing awareness of students development of predictive reasoning and meta-representational competence in the middle years of primary school. The cases described here propose that students may move through stages when interpreting data tables and constructing graphs. By construction and visualisation of data sets through freehand drawings, students have the opportunity to notice and internalise key structural elements such as equal spacing, scale and coordinating axes. Developing meta-representational competence—or the capacity to represent and restructure data—prior to the introduction of formal graphing is recommended to avoid a procedural approach where students learn to graph without conceptual understanding. Further research into this process could inform educators of the optimal stage at which to intervene with more formal pedagogical approaches to developing statistical concepts and graphing.

References

Battista, M. T., Clements, D. H., Arnoff, J., Battista, K., & Van Auken Borrow, C. (1998). Students' spatial structuring of 2D arrays of squares. *Journal for Research in Mathematics Education. 29*(5), 503–532

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? *Reasoning about distribution: A collection of current research studies. Proceedings of the 4th international research forum for statistical reasoning, thinking and literacy.* SRTL-4. Brisbane: University of Queensland. https://www.academia.edu/976792

English, L. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics, 81*(1), 15–30.

Fielding, J., & Makar, K. (2022). Challenging conceptual understanding in a complex system: Supporting young students to address extended mathematical inquiry problems. *Instructional Science, 50*, 35–61.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

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 (Eds.), *Statistics in early childhood and primary education* (pp. 89–108). Singapore: Springer. https://doi.org/10.1007/978-981-13-1044-7_6

Mulligan, J. (2015). Moving beyond basic numeracy: Data modeling in the early years of schooling. *ZDM Mathematics Education, 47*(4), 653–663.

Oslington, G., Mulligan, J., & Van Bergen, P. (2020). Third-graders' predictive reasoning strategies. *Educational Studies in Mathematics*, *104*(1)*,* 5–24.

Oslington, G., Mulligan, J., & Van Bergen, P. (2021). The development of predictive reasoning in Grades 3 through 4. In Y. H. Leong, B. Kaur, B. H. Choy, S. L. Chin, & J. B. Yeo (Eds.), *Excellence in mathematics education. Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 305–312). Singapore: MERGA.

Oslington, G., Mulligan, J., & Van Bergen, P. (2023). Shifts in students' predictive reasoning from data tables in years 3 and 4. *Mathematics Education Research Journal*. https://doi.org/10.1007/s13394-023-00460-2

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.