THE INFLUENCE OF VARIATION AND EXPECTATION ON THE DEVELOPING AWARENESS OF DISTRIBUTION

JANE M. WATSON
University of Tasmania
Jane.Watson@utas.edu.au

ABSTRACT

This study considers the evolving influence of variation and expectation on the development of school students' appreciation of distribution as displayed in their construction of graphical representations of data sets. Three interview protocols are employed, presenting different contexts within which 109 students, ranging in age from 6 to 15 years, could display and interpret their understanding. Responses are analyzed within a hierarchical cognitive framework. It is hypothesized from the analysis that, contrary to the order in which expectation and variation are introduced in the school curriculum, the natural tendency for students is to acknowledge variation first and then expectation.

Keywords: *Statistics education research; Interviews; School students; Graphs*

1. TERMINOLOGY

Like many words used in statistics, "distribution" has a more or less sophisticated meaning depending on the adjective placed in front of it. Among the synonyms used for distribution in the Chambers dictionary (Kirkpatrick, 1983) are dispersal, range, allotment, and classification. These are in turn based on the word "distribute" meaning variously "to divide among several … to disperse about a space … to spread out" (p. 364). These descriptions are useful starting points in exploring children's experiences with graphing distributions. Moore and McCabe (1993) progress to describe distribution in terms of variation, which they treat as an undefined term, and variable, which is "any characteristic of a person or thing that can be expressed as a number" (p. 2): "The pattern of variation of a variable is called its *distribution*. The distribution records the numerical values of the variable and how often each value occurs" (p. 6). They go on to say that distributions are best displayed graphically. These basic representations are often called frequency distributions to distinguish them from theoretical distributions based on continuous curves.

For the school-age students interviewed in this study, distributions are likely to represent collections of data from relatively small data sets that are shown graphically in stacked dot plots, bar graphs, or histograms. The idea of a theoretical distribution such as the normal distribution is not part of their vocabulary or experience. For some younger students idiosyncratic representations may satisfy the less restrictive constraints of the Chambers definition but still show range, spread, and classification. Bakker and Gravemeijer (2004) for example described Grade 7 students' early work with case-value plots as the beginning of exploration of characteristics of distributions.

Although Moore and McCabe (1993) treated "variation" as an undefined term, Reading and Shaughnessy (2004) distinguished between "variability," as the

Statistics Education Research Journal, 8(1), 32-61, http://www.stat.auckland.ac.nz/serj

characteristic of an entity that is observable, and "variation," as the describing or measuring of that characteristic. This is the distinction used in this study because for school students it is the act of describing or representing variability that appears in the graphs created by them. The term "expectation" is chosen in contrast to "variation," usually reflecting the meaning of the expected value (e.g., mean) of a probability distribution. This translates to the familiar terms middle or average for frequency distributions. It may also however in some contexts refer more colloquially to the expected shape of a distribution, for example showing a particular trend.

2. OVERVIEW OF THE PROBLEM AND ITS IMPORTANCE

Although not told that they are beginning to learn about distributions, children in early childhood classrooms create pictographs by recording the favorite fruit of members of the class or the modes of transport used to get to school. Throughout the school years more complex forms of representation are introduced until perhaps at first year university level students, in some countries, are expected to understand the theoretical underpinnings of the normal, binomial, Poisson, exponential, and other distributions. Although students may be able to create various types of graphs for different sets of data, as required to meet curriculum objectives, a larger objective in terms of the goal of statistical literacy when students leave school is to be able to tell a story from a context with a distribution that displays variation, clustering, middles, and surprises. This may or may not involve a conventional text-book type of graph. Of interest from an educational perspective is the development that takes place in students' abilities to create representations that are effective in displaying the variation in data sets that will best tell the stories in the appropriate contexts.

In parallel with the introduction of various increasingly complex graphical forms, the data handling curriculum introduces measures of center, measures of chance, and later measures of spread. These are typically the arithmetic mean, the counting-of-favorable-outcomes approach to probability, and the standard deviation. The first two are associated with the statistical concept of expectation, whereas the third is associated with the concept of variation. The complexity of the calculations required for the standard deviation means that it is not introduced until the final school years and it has been suggested by Shaughnessy (1997) that the associated concept of variation traditionally has not received very much explicit attention until then. Whether this apparent differentiation in emphasis on the two ideas of expectation and variation has an influence on students' developing ideas of distribution is unknown. The purpose of this study is to explore students' efforts in graphing distributions for evidence of these two concepts.

3. BACKGROUND

The relationship of school students' understanding of variation and expectation and their understanding of distribution has been slow to emerge in the literature, following an initial focus on graphing skills. Asking students to create representations for contexts without specific data has provided a window on developing understanding of the relationship.

Historically the study of students' creation and interpretation of graphical representations has mainly been related to the conventional production of school-taught graphical forms, usually based in algebra (e.g., Kerslake, 1981; Leinhardt, Zaslavsky, & Stein, 1990) but sometimes in relation to data (e.g., Curcio, 1987; Curcio & Artzt, 1996; Friel, Curcio, & Bright, 2001). In viewing the school-level conception of distribution as "graphing," there has been considerable attention to students' abilities to create various graphical types, with emphases for example both on what types should be taught when (e.g., Friel et al., 2001) and on appropriate scaling, labeling, and directionality of plots (e.g., Leinhardt et al., 1990; Mevarech & Kramarsky, 1997). Until recently, however, explicit consideration of variation in relation to graphical representations has not been a feature of research. The pleas of Green (1993) and Shaughnessy (1997) brought variation generally to the attention of statistics educators interested in student understanding of the chance and data curriculum at the school level.

Explicit attention to variation included a focus on how specific features of graphs influence decision making, for example in comparing two data sets presented in graphical form (e.g., Watson, 2001, 2002). The relationship of variation to the statistical concept of distribution is close but intuitively variation is a term covering all sorts of observed change in phenomena whereas distribution is a more formal notion based on graphs that is built into the later years of the school curriculum (National Council of Teachers of Mathematics [NCTM], 2000). The work of delMas and Liu (2003) illustrated this in relation to the understanding of standard deviation and spread at the early tertiary level, whereas Petrosino, Lehrer, and Schauble (2003) showed that relatively formal ideas about spread and difference could be introduced as early as Grade 4. Ben-Zvi and Amir (2005) explored emerging ideas of distribution with three Grade 2 students in considering data on the loss of "baby" teeth. They found, for example, that when speculating about data (and implicitly distributions) the students were reluctant to suggest repeated values. Considering questions about data explored by elementary students, Russell (2006) found students who focused on individual values, particularly the mode, as well as those who saw "clumps" of data values, or thought about "middles" in an intuitive sense. She made specific suggestions for moving students to an "aggregate" view of data distributions as described by Konold, Higgins, Russell, and Khalil (2003). Further evidence of such a developmental pattern was presented by Friel, O'Connor, and Mamer (2006) who observed student explorations of sugar content in cereals and of heart rates. In both cases comparing distributions of data sets was an integral part of the investigation. Looking more explicitly at the expected shape of distributions, Shaughnessy (2006) described middle and high school students' decisions about "real" or "fake" data, finding various strategies for decision making. These included a focus on outliers, on the whole range of possible outcomes, on the likely range of outcomes, and on the distance from a fixed point, usually the expected center. The work of Watson and Kelly (e.g., 2002a) indicated that general understanding related to variation, and at times specifically related to distributions of outcomes (e.g., Watson & Kelly, 2004a), could be improved with instruction at the school level.

Although variation in data creates distributions, there are two other aspects of statistical settings that are likely to have an impact on what a graph looks like. One aspect is the presence of some underlying expectation that can be observed in the distribution, for example a peak in the center of a symmetric distribution or the uniform nature of single die outcomes. In a theoretical distribution such expectation determines the shape of the distribution, for example the proportion of "successes" in a binomial distribution or the constancy of a uniform distribution. A second aspect in an actual empirical situation is that there is likely to be variation from the theoretical distribution itself. Hence the person creating a graph may have to consider the variation from expectation that creates a distribution (or trend), as well as the variation from the expected distribution. The question of how much variation from an expected distribution is considered realistic in a given situation depends to a large extent on the graph-drawer's experience with similar contexts in the past. This can make the creation of representations from verbal descriptions quite complex.

Depending on previous learning experiences, representations may be based on traditional graphical forms or may be quite unique. The latter may be difficult for others to interpret, even for experts (e.g., Roth & Bowen, 2003). Calls to allow students to create their own graphs (e.g., Curcio & Artzt, 1996) then place pressure on researchers to interpret the meaning of graphs if the students are no longer available to explain what they have done. Initial choice of what data values, or type of data values, to represent, may not in the end suit the story expected to be told.

Asking students to create graphs of variables based only on verbal descriptions has been the basis of occasional studies in mathematics education. Swan (1988) for example was interested in tasks such as showing in a graph how the price per ticket varies with group size for a fixed total cost. Mevarech and Kramarsky (1997) and Moritz (2002) considered tasks representing the situation of the amount of time a student studies and the level of grade that is obtained. Moritz (2000) also considered student representations of growing taller with age but stopping at age 20. The impression of researchers is that such tasks are more difficult than straightforward representation of data values, perhaps due to the need to appreciate context and visualize a trend or association rather than remember rules for creating axes and plotting points.

The relationship of the order in which expectation and variation are emphasized in the school curriculum and the order in which students develop an appreciation of the two concepts was explored by Watson (2005). She used quotes from students from preparatory grade (6-year-olds) to high school to hypothesize that students' intuitions develop in the reverse order to that suggested in the data handling curriculum. The youngest students for example were able to suggest variation, with different numbers of red lollies in different groups of 10 drawn from a container with 50% red lollies in it (e.g., 4, 5, 1, 3, 6, 8) but unable to predict expected numbers clustered as suggested by the proportion of reds in the container. Predictions were likely to be based on favorite numbers or the size of the student's hand. By Grade 7 most students were able to provide predictions based on half of the lollies being red and reasonable variation of values around this (e.g., 5, 3, 6, 4, 5, 4). The current study presents a detailed analysis of the same data set with respect to graphical representations to support further the hypothesis. At the same time the beginnings of a more sophisticated idea of distribution are documented.

4. RESEARCH QUESTIONS

The research questions for this study are based on three tasks in different contexts that required students to create representations of data sets.

1. What levels of sophistication are shown in terms of the acknowledgement of variation and expectation in the creation of graphical representations of distributions of data sets?
2. Does there appear to be a trend for higher levels of performance with later grades?

5. METHOD

5.1. TASKS

Three interview protocols are the basis of the exploration in this study. As part of the larger projects in which these interviews were embedded, several hundred students completed surveys based on concepts in the chance and data curriculum (Watson, 2006). The interviews took place in order to focus on understanding that could be displayed with extra time and in-depth questioning (Burns, 2000, pp. 582-3). In particular, aspects of expectation and variation were explored in relation to the distributions created by students in completing the tasks.

In each case students were asked to create a representation of a data set or situation. Each context was different, giving the opportunity to compare and contrast the attempts at creating distributions to tell the story in data. The first, BOOKS, was based on the creation of pictographs given concrete materials in an interview setting (Watson & Moritz, 2001; see Appendix A). Students were given cards depicting books and people and asked to represent the specified numbers of books people had read (e.g., “Lisa read 6,” “Danny read 3”) on a table top. The names of children and numbers of books read were supplied by the interviewer and questions of interpretation and prediction were asked after the representation had been created. The data presented showed a tendency for girls to read more books than boys. The second task, WEATHER, was based on the description of average temperature: “Some students watched the news every night for a year, and recorded the daily maximum temperature in Hobart. They found that the average maximum temperature in Hobart was 17° C” (Watson & Kelly, 2005; see Appendix B). After initial questions, including predictions for maximum temperatures for six days of the year, students were asked to describe the daily maximum temperature for Hobart over a year in a graph. The third task, LOLLIES, was based on an experimental situation where students were asked to imagine a container with 100 lollies mixed up in it: 50 red, 20 yellow, and 30 green (Kelly & Watson, 2002; Reading & Shaughnessy, 2000; see Appendix C). They were asked to imagine the outcomes from pulling out 10 lollies and to suggest the number of red lollies in the 10 from six such trials. After other questions and six experiments from an actual container as described, they were asked to draw a picture of the imagined outcomes of 40 such experiments.

5.2. SAMPLE

The student work chosen for analysis in this study was combined from responses in two different studies. Students in Grades 3 to 9 were chosen to be interviewed based on interesting or unusual responses to the in-class survey. Teachers advised on the suitability of the students to articulate their views to the interviewers. Parental and student permission was granted for the interviews. Some students completed more than one task. The preparatory students (P) were 6-year-old students described in Watson and Kelly (2002b) chosen by their teacher as high achieving in number skills and happy to talk to visitors. Again parental permission was obtained. They were asked all three protocols. For the WEATHER and LOLLIES protocols, the same students in Grades 3, 5, 7, and 9 answered both. A summary of the number of students in each grade completing each task is given in Table 1.

Table 1. Number of students interviewed for each task by grade

	Grade					
Task	P	3	5	7	9	Total
BOOKS	7	6	8	14	8	43
WEATHER	4	18	18	15	15	70
LOLLIES	7	18	18	15	15	73

5.3. ANALYSIS

Two criteria are the basis of the analysis reported in this paper. One criterion is the framework from the work of Biggs and Collis (1982; Biggs, 1992; Pegg, 2002a, 2002b) in cognitive psychology. Their Structure of Observed Learning Outcomes (SOLO) model suggests five levels of performance that may be assessed in relation to a task that is set with the expectation of success in the mode of cognition of students during their years of schooling. These levels and references to characteristics of responses shown are given in Table 2 (see also Watson & Moritz, 2000).

Table 2. Summary of SOLO level expectations for tasks[a]

Name	Elements	Conflict (should it arise)
Level 0: Prestructural	No elements related to task employed in response	No recognition of conflict/contradictions
Level 1: Unistructural	Single element of task employed in response	No recognition of conflict/contradictions
Level 2: Multistructural	Multiple elements employed in response, usually in sequence	Recognition of conflict/contradictions but inability to resolve adequately
Level 3: Relational	Multiple elements employed in a coordinated, integrated fashion in response	Resolution of conflict that arises in task
Level 4: Extended abstract	Response goes beyond Relational level to introduce other elements not in the initial task but relevant to its extension	May suggest potential for further conflict and resolve or give alternatives

[a]Summary adapted from Biggs & Collis (1982), Pegg (2002a), and Watson & Moritz (2000).

The other criterion for analysis is related to the statistical appropriateness of the responses given. For these tasks this has to do with creating a representation that displays aspects of expectation as expressed in the task as well as appropriate variation. This should result in some kind of distribution that tells the story of the task set. For BOOKS, the pictograph should tell a frequency story of the number of books each child has read. The children’s names provide cases against which case values are recorded. This is a case value graph (Konold & Higgins, 2003) of the type discussed by Moritz (2000, 2002), Pfannkuch, Rubick, and Yoon (2002), and Chick (2004) as an introduction to considering frequency. The names may be placed in no special order, in alphabetical order, or ordered by the frequency associated with each. For the WEATHER task it is likely that a time series graph is drawn representing either daily maxima or monthly averages of daily maxima. In the case of daily values this is a transition from a case value graph. Using the frequency of days whose maximum temperature is each value in the range, say 9 to 34 (totaling 365), would produce a frequency distribution. The LOLLIES task also suggests representation of case values, this time with respect to 40 draws of 10 lollies from a container. Each draw results in a number of red lollies varying from 0 to 10. If these are recorded as case values from 1 to 40 serially, a representation similar to a time series graph is created. Counting and recording frequencies for each of these 40 outcomes in categories 0 to 10 produces an approximation to a random probability distribution. Data are ordered in 11 groups and related to a theoretical premise (the binomial distribution). The task for LOLLIES hence appears the most difficult statistically of the three tasks.

Case value plots may not look alike but they are characterized by the display of count or measurement values (e.g., number of books read, maximum temperature, number of red lollies) for individual identifiable cases (named child, day of the year, numbered student). When the cases are strictly ordered (e.g., successive days of the year, successive years, or successive trials) the graph appears as a time series graph (e.g., of maximum temperature, of number of red lollies). A different format emerges when the graph changes to recording the frequency distribution of the variable of interest rather than successive case values. Here the range of possible values of the variable is plotted (usually on the horizontal axis) (e.g., 0 to 7 books read, minimum to maximum daily maximum temperature, 0 to 10 red lollies) and frequencies are recorded vertically (e.g., number of children who read X books, number of days when the maximum temperature was X degrees, or number of times X red lollies were drawn).

The clustering of responses to the three tasks (Miles & Huberman, 1994, p. 248) with the SOLO framework as an implicit structure led to descriptions of the levels that, while reflecting the inclusion of the more relevant elements, also identified variation as the key initial element. Variation was then linked in more appropriate and structurally complex fashions to the data before the element of expectation was introduced. These more explicit labels for the levels are introduced in Table 3 and indications of typical responses for each graph creation task are given. Coding of representations was based on these levels. It was completed independently and confirmed by two researchers, one of which was the author.

Table 3. Redefined levels for tasks in this study with examples

	BOOKS	WEATHER	LOLLIES
Level 0: Idiosyncratic – No indication of variation or expectation	Indistinguishable piles of books on top of people	Drawings of wind or a weather map	Drawings of lollies and children
Level 1: Unstructured variation	Books and children spread about	Summer/Winter Tables of temperatures	Lists of numbers of lollies
Level 2: Variation shown by value	Children in a line with books in perpendicular lines	Successive dates with temperatures	Successive draws in a series, or frequency with variation only
Level 3: Initial acknowledgement of expectation	Children ordered by books read (least to most) or reference to middle	Seasonal change	Acknowledgement of middle
Level 4: Integration of variation and expectation	Prediction based on middle/mean, and distinguished variation	Seasonal and daily change	Distribution centered on five

The representations presented in this paper are not randomly selected but purpose-chosen as typical of the levels of response identified from the two data sets, as well as illustrating aspects of variation and expectation displayed. They demonstrate proposed hierarchies in terms of structure and appropriateness. Following the presentation of clusters of responses for the three tasks, a summary is presented across tasks.

6. RESULTS

6.1. BOOKS

Although the BOOKS task asked students to create a case value pictograph, several questions gave students the opportunity to highlight aspects of the data set and the variation and expectation present in it (see Appendix A). In the final data set presented to students for example, there were two children who had each read four books. Students were also asked to show how the pictograph would look after all children had been to the library and selected another book. Of interest in the pictographs presented is the degree to which variation is catered for in the distributions created by the students.

At Level 0, the piling of books on top of or beside the pictures of the children appears to preclude any description of variation in the number of books children had read. Examples of this approach are shown in Figure 1 and although all students displayed one-to-one accuracy in counting, this is not visible in the display, and the responses are considered prestructural or idiosyncratic with respect to representing variation.

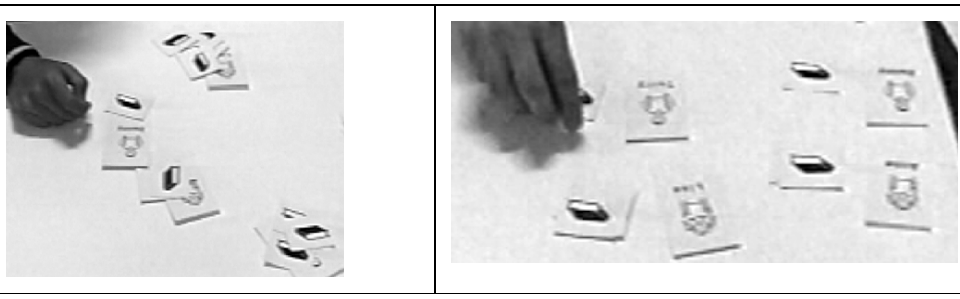


Figure 1. No visible (or very little) variation shown in representation (Level 0)

The pictographs shown in Figure 2 display variation in the number of books read, either through a scattered representation of both children and books or through a more ordered representation of books for still scattered children. Showing a single aspect of variation these responses are considered to be Level 1.

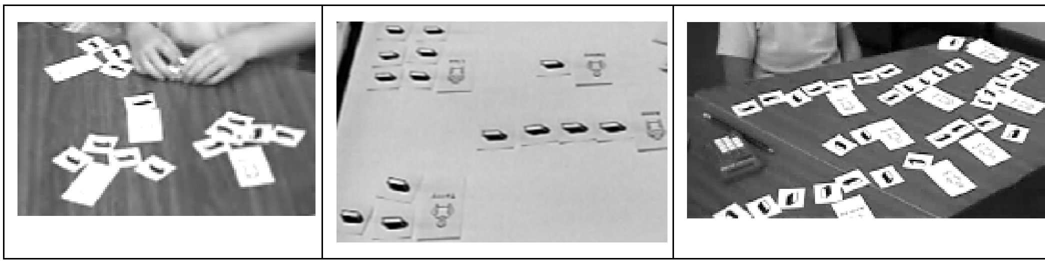


Figure 2. Variation clear but unstructured (Level 1)

In Figure 3 the children are placed either vertically or horizontally along the edge of the pictograph in order to line up the books in a grid format. In the lower left pictograph the additional books from the library are displayed on the far side of the representation, whereas in the pictograph on the lower right, two children have been placed by the four books that they each had read. Difficulty occurred for this student, however, because the additional library book was represented twice rather than once (shown by placing two

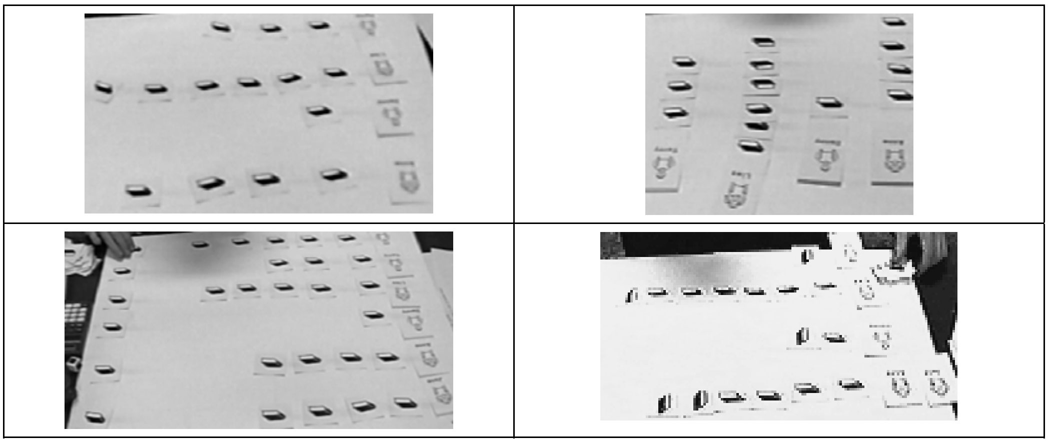


Figure 3. Variation shown in rectangular format (Level 2)

books at right angles at the end of the row). These responses are judged to be Level 2 in taking into account at least two aspects of representing difference among the children in the number of books they had read.

Up to Level 2, students did not look at the shape of the data in making predictions about how many books a new student to the class might have read (see Appendix A). Many younger students refused to answer the question, some saying they could not do so because they did not know the student and some because they did not want to make a guess. Other students provided values based on the gaps in the displays they had created.

At Level 3, responses indicated an intuition about expectation within the displayed variation, either through rearranging the pictograph or making informal reference to the middle. The representations in Figure 4 order the case values so that variation is more easily gauged and the range of values from minimum to maximum is clear. In the pictograph on the far right the children who had each read four books are again placed side by side but this time the extra library book is placed on the other side of the children and only one book is used for the two children with four books. When asked how many books the new students Paul or Mary might have read, some students suggested informal references to middle.

Instructor: Suppose Paul comes along … how many might he have read?
Student: About 3 because it is in the middle of all the other numbers.
Instructor: … Mary? … similar or different?
Student: Probably the same.
Instructor: You would be pretty sure Paul had read 3?
Student: No, you wouldn't know, you would just guess.

This type of response was also classified as Level 3 in moving toward an expected value.

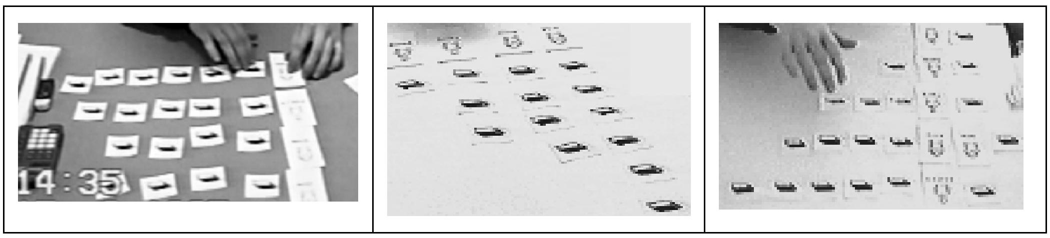


Figure 4. Ordered variation in the display (Level 3)

To be classed as Level 4 a response had to address expectation and variation explicitly in the prediction question about how many books a new student to the class, Paul or Mary, might have read. Sophisticated responses employing both variation and expectation are illustrated by the following response.

Instructor: Suppose Paul comes along … how many might he have read?
Student: … About 4 probably.
Instructor: Why?
Student: Because if you add them all up and then divide them, roughly that's what you get. It is about that anyway.
Instructor: … Mary? … same or different?
Student: I suppose you could add up the girls and the boys and keep them separate.
Instructor: Why?
Student: Because the girls are obviously more interested in reading.

Noticing the variation between boys and girls and separating the estimates was typical of Level 4 responses. Table 4 summarizes the responses by grade and level for the 43 students who responded to the BOOKS interview protocol. Only students from Grade 7 began to consider expectation in their distributions and/or predictions.

Table 4. Summary for BOOKS protocol (n = 43)

	Grade				
	P	3	5	7	9
Level 0	1	3	2	1	1
Level 1	3	1	3	1	1
Level 2	3	2	3	5	2
Level 3	0	0	0	4	2
Level 4	0	0	0	3	2
Total	7	6	8	14	8

6.2. WEATHER

For the WEATHER protocol (Appendix B) students were asked within the context of a statement about the average daily maximum temperature to draw a graph of the temperature in Hobart for a year; no grid or framework was provided. Later students were asked to judge three other representations, as shown in Appendix B. The appendix also shows the questions asked before students were asked to draw a graph of the maximum daily temperature throughout the year. It was expected that students would be familiar with variation in the weather context, especially trends associated with the seasons, but this was not introduced by the interviewer.

Some students drew pictures rather than graphs, some of these depicting variation and others not. Those shown in Figure 5 are static in nature, although telling something about daily maxima. These representations are Level 0. A Grade 3 student explained the center drawing in Figure 5 as, "A stick person. Probably be trees. Blowing a little bit. Probably be like a hot day with a little bit of a breeze," whereas a Grade 5 student drew the weather map on the right of Figure 5, describing it in the following terms:

Student: [draws a square] So you may have a map of Tasmania … Hobart is here; Swansea is here; Strahan is here and Launceston is there. [puts dots on page] You may say that Hobart is 11 and Strahan may be 15 and Launceston may be 20 and Swansea may be 13.

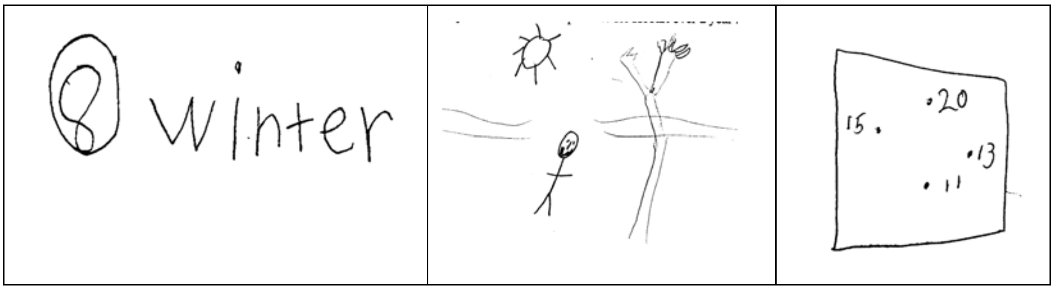
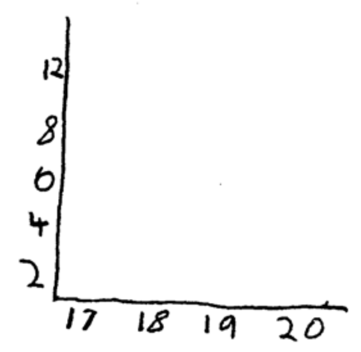


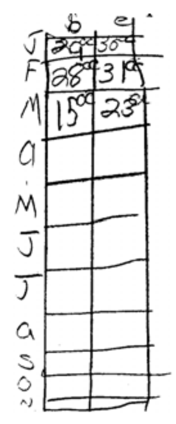
Figure 5. Static weather pictures (Level 0)

The two representations in Figure 6 indicate variation from season to season or for weather. These responses are considered to be at Level 1.

Figure 6. Variation shown in weather pictures (Level 1)

Also at Level 1, quite a few students could draw a framework for dealing with the task, indicating that they were attempting to show variation, but then had little idea of how to organize the story they wanted to tell. A start is shown in Figure 7. Explanations of some of the attempts to record values were difficult to follow and, although suggesting variation, there was no link to any expectation or trend. An example is given on the right of Figure 7.





S: That's start, oh well beginning and the end [writes on sheet]
I: So you have got the beginning of the month and the end of the month. Is that what you are doing?
S: Yes. Start off with … so I would do like start with the highest … [writes]
I: So what are you putting in there? You are doing a table?
S: Yes. Like 29°C is the highest the start of January and the end would be about 30.

Figure 7. Frameworks to indicate variation but no data or a few data values (Level 1)

Two types of graphs were employed to begin to structure the display of variation in the temperatures at Level 2. Some suggested a frequency approach for various temperatures, as shown in Figure 8, choosing various temperatures, apparently randomly, for reporting. The student who drew the representation on the left for example said, “That [first column] shows that there’s 4 days which bring 17.” Others were based on time throughout the year. These were more likely to display a trend in variation, as shown in Figure 9, although some did not progress far enough to do so.

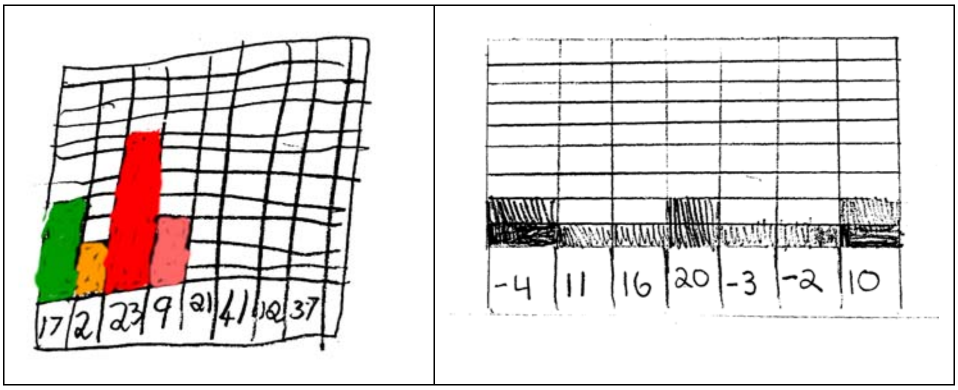


Figure 8. Frequency graphs (Level 2)

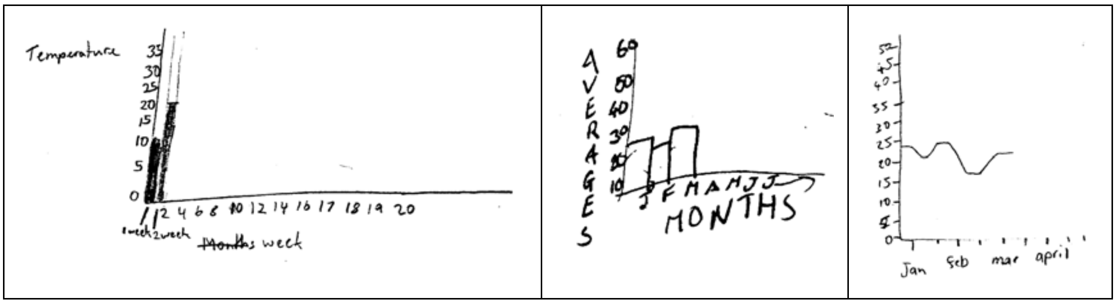
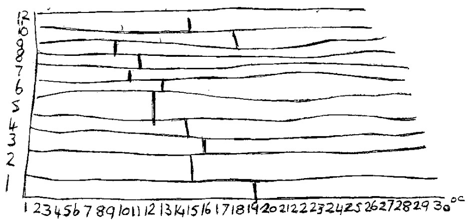


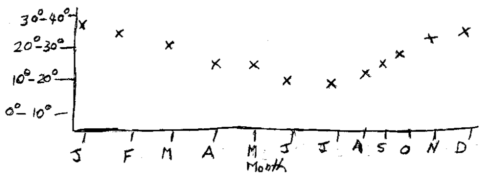
Figure 9. Beginnings of temperature graphs (Level 2)

At Level 3 the graphs showing seasonal change were represented variously as continuous lines, vertical lines for short periods, line graphs, and bar graphs for months or seasons. Examples showing two methods of display, along with the students’ explanations, are given in Figure 10. These responses reflected the intuitive expectation of the weather context.

Although there was mention of seasonal difference and change in the extracts accompanying the graphs in Figure 10, there was no discussion of daily variation when the graphs were drawn. Two examples that include short term variation as well as seasonal expectation are shown in Figure 11. These are judged to be Level 4 responses in the ability to focus on both variation and expectation.

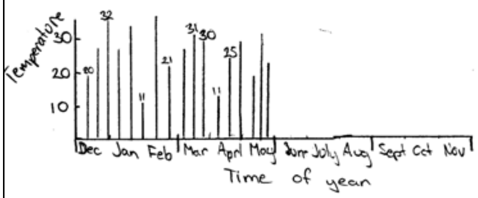


I: *So if those are the averages for the months – now can you tell me why they are going up and down a bit?*
S: Colder and hotter.
I: OK – so where's the coldest part of the year?
S: In the middle [points to graph].
I: Sort of in the middle there, somewhere.
I: And it gets warmer toward the end.
S: Yes, Yes.

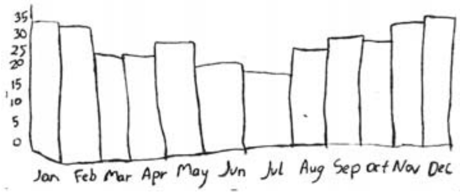


S: I think I will have a little line graph.
I: Can you explain it to me.
S: Well January is usually the hottest month and so the average what the temperature is.
I: The average of the temperature and that [graph] represents that for the month.
S: Yes each cross.
I: And why does it go down?
S: Because in the middle of winter it is generally colder than it is in the middle of summer.

Figure 10. Graphs with seasonal variation (Level 3)



I: What are each of these lines here?
S: They are just a–throughout the–this is so you get a view of all the different temperatures that it can range from and just like it might be up, it might be a hot day one day and it might start going colder and it might get hot again and then as it goes down. It is going to start getting colder around June and July and then it is going to start coming back up [points to graph].
I: What does each line represent?
S: A week.



I: So what have you done there?
S: It's the highest in January, February, and December cause that's the middle of summer... The coldest would be around here in winter. In around these sections, it's around middling.
I: It's interesting you've got May a little bit higher here…
S: Yea, it could change. There'd be a lucky day sometimes. It could just go up over.
I: So are these temperatures, are they what, maximums, or averages or…?
S: Yeah, maximum averages.

Figure 11. Temperature graphs with seasonal expectation and daily variation (Level 4)

Table 5 summarizes the levels of response for each grade for the WEATHER task for the 70 students who responded in the interviews. For this sample of students only two Grade 7 students, whose responses are shown in Figure 11, reached Level 4 in appreciating both expectation and variation in their responses.

Table 5. Summary of levels by grade for the WEATHER protocol (n = 70)

	Grade				
	P	3	5	7	9
Level 0	4	6	2	0	0
Level 1	0	8	7	3	0
Level 2	0	4	7	9	8
Level 3	0	0	2	1	7
Level 4	0	0	0	2	0
Total	4	18	18	15	15

6.3. LOLLIES

The complete interview for the LOLLIES task, as seen in Appendix C, provides the background for the question relating to imagining the outcomes of 40 trials where 10 lollies are drawn each time (with replacement) and the number of reds counted. After an initial request to draw a graph with no support provided, a blank set of axes was provided to students who did not initially produce a distribution of outcomes centered about five to help them think about distributions. The axes did not assist some students. The levels of response for the initial representations are presented first, followed by those for students shown axes.

Similar to the WEATHER task, Figure 12 shows how some younger students interpreted the task by sketching the context for the drawing of lollies from a container. These are examples of Level 0 responses. A Grade 3 student explained the representation in the center of Figure 12 as follows, pointing to each part of the drawing:

> Student: Well they are 4 tables and they are the boxes with the lollies in them and they are two sheets of paper on each row so they can write down their answers and that's just a person who watches, sits there and collects the pieces of paper from each one. And then there's a row of 10 people [vertical lines].

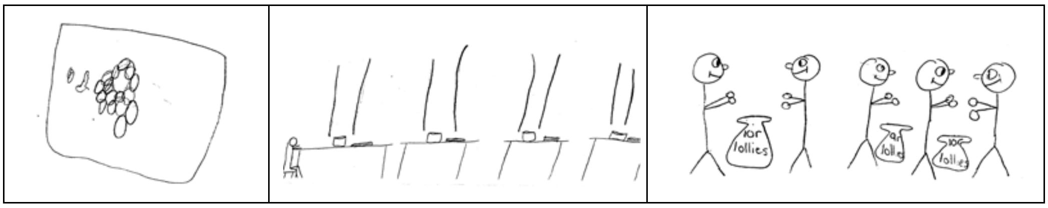


Figure 12. Sketches of contexts for drawing lollies (Level 0)

Other students across the grades initially provided numerical values for the outcomes. Responses suggesting individual values, rather than enough to indicate some variation, are shown in Figure 13 and are also assigned Level 0.

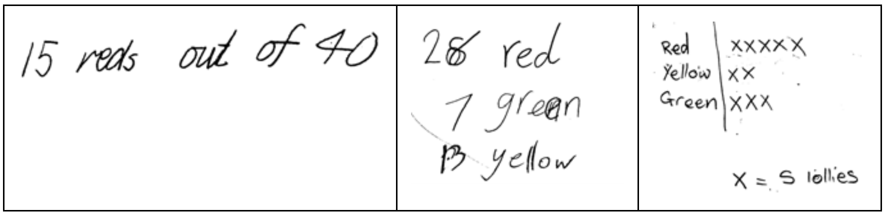


Figure 13. Individual outcomes from draws (Level 0)

Responses explicitly suggesting variation are shown in Figure 14. A Grade 5 student explained the table on the right of Figure 14, as “Well it isn’t exactly forty people, there was like a group and they each wrote down their answers in a line across the top.” These responses are placed at Level 1.

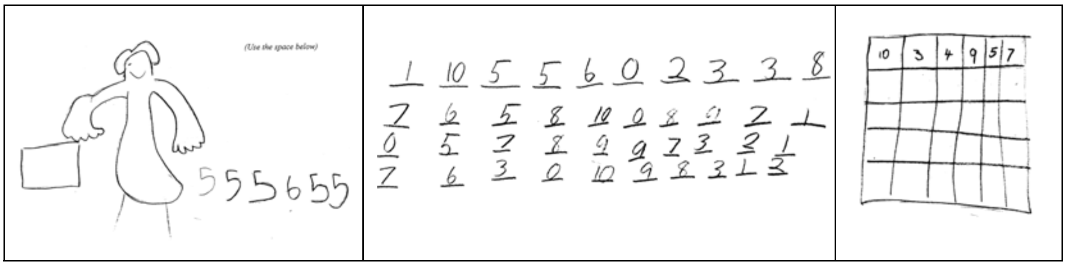


Figure 14. Multiple outcomes from draws (Level 1)

The representing of outcomes for the 40 draws in a time series format was used by some students and often they were stopped from completing all 40 due to time constraints in the interview. The spread of the suggested number of reds was often quite large and occasionally very small. Two examples are shown in Figure 15 (both from Grade 7); these representations are placed at Level 2.

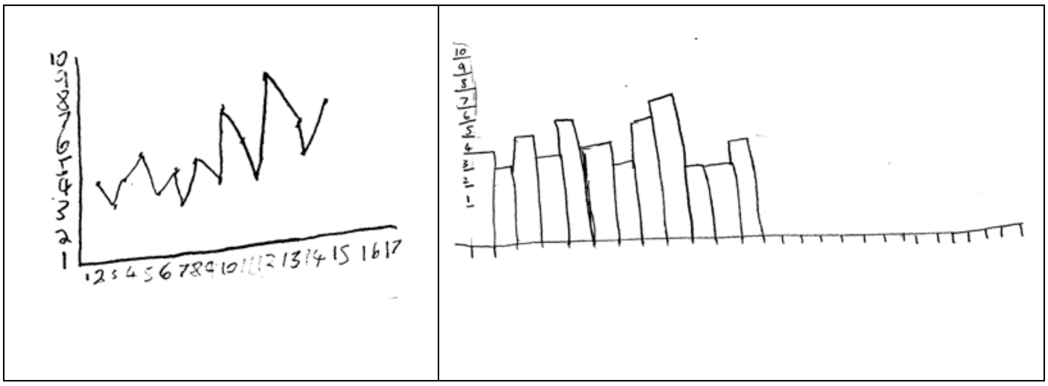
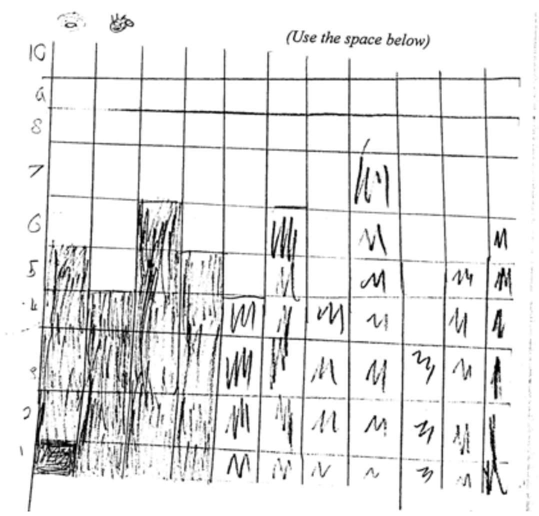


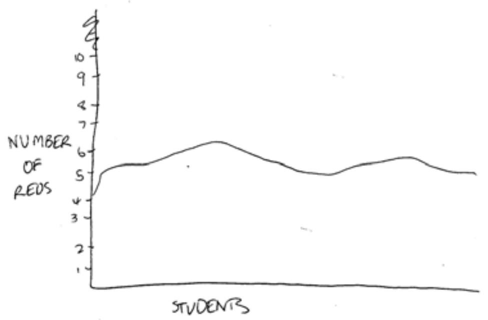
Figure 15. Time-series-like graphs (Level 2)

Some graphs of the time-series type showed a realistic degree of variation about a middle value, as do the two in Figure 16 by Grade 3 and Grade 9 students, with the accompanying explanations. These were judged Level 3 in appreciation of both variation and an intuitive notion of center.

Without axes being provided, only four students produced a prototype of a typical frequency distribution. These responses were judged to be at Level 4 and are all shown in Figure 17.

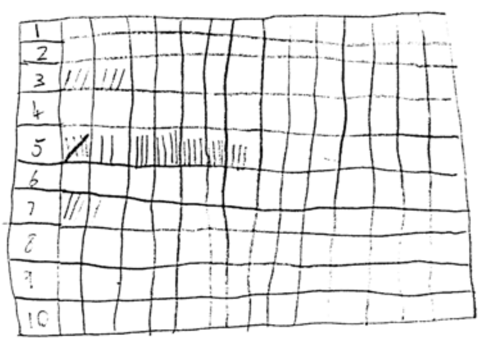


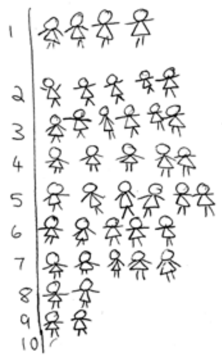
S: How many reds they have got.
I: So that is a person? Each one of these columns is a person is it?
S: Yes. And the coloring in is how many reds they got.
I: Out of 10.
S: This is like the end [sheet].
I: So that person got 1 red so far. The first person gets about?
S: Five
I: Next person?
S: Four.
I: Do one more yeah … Yes you have got the right idea. I can see what you mean by showing it. So you would do that like that for 40 people though wouldn't you. So you think they all would get around what?
S: Around 6 and 5, around that, 6, 5, 4 and 7.

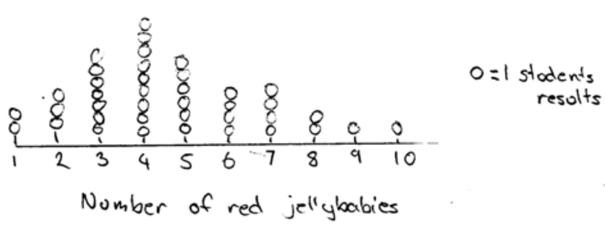


S: I drew a graph and they had the number of reds up the sides. The students along the bottom. And I drew a line going around 5, it goes up to 6 and down to 4 sometimes.

Figure 16. Time-series-like graphs with appreciation of center (Level 3)







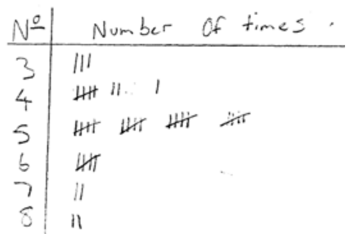
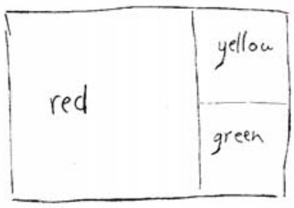


Figure 17. Frequency distribution graphs (Level 4)

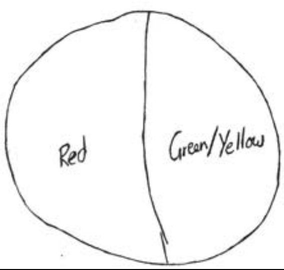
Table 6 shows the levels of response for each grade for the initial representations drawn for the LOLLIES protocol. Young students had some difficulty with appreciating the task in its original form. Only two graphs produced by Grade 7 students appeared to represent expectation to the exclusion of variation in the initial graph. The representations and the students' explanations are shown in Figure 18. These used an area model for probability and were judged Level 0 with respect to this model. They were the only two responses initially to represent expectation rather than variation, the reverse to the hypothesis of this study. It may be that classroom instruction influenced these representations as the two students were from the same class.

Table 6. Summary of initial levels by grade for the LOLLIES protocol (n = 73)

	Grade				
	P	3	5	7	9
Level 0	4	14	7	4	2
Level 1	3	3	6	3	4
Level 2	0	0	3	7	5
Level 3	0	1	0	0	3
Level 4	0	0	2	1	1
Total	7	18	18	15	15



S: Drawing about [outline of a rectangle] … drawing like a space just like drawing a square [draws a rectangle] – it is not really square.
I: That's all right, just a sketch.
S: And probably about this much of it red and this much [yellow] and green.
I: *Now what does this [left section] represent?*
S: The red lollies. Probably about two thirds.
I: *About two thirds of what would come out would be red.* [nods] *And about how much yellow and green?*
S: Actually I was trying to make that – they would probably be a little bit less than a quarter.
I: A little bit less than a quarter.
S: Yes.
I: For each of them or together?
S: For each of them I think.



S: Oh, well they could do it in a pie graph or something and have the reds and then the other section is whatever else.
I: So the reds would be … right, OK. Would it be one pie graph for the whole class do you think?
S: [pause] Yes could be.
I: Do you want to do me a sketch of what you think it might look like.
S: [draws pie graph]. There's a bit more than half.
I: Do you think that would be the reds?
S: Yes.

Figure 18. Graphs representing expectation only (Level 0)

The four students who drew the graphs in Figure 17 were not presented with the axes format as they had produced an equivalent form on their own. Some other students were not presented with the graph format with axes because of confusion with the task, time constraints, or perceived lagging interest in the protocol.

Fifty-four students (all Prep, 13 Grade 3 and 9, 11 Grade 5, and 10 Grade 7) who produced lower level initial graphs were shown the graph format with axes. The Prep students were given a complete "boxed" grid where they could color in boxes if they desired. Of the 54 students, 3 Prep, 10 Grade 3, 8 Grade 5, 9 Grade 7 and 8 Grade 9 improved their levels of response. Only one student in each of Grades 3, 7, and 9 produced a response using the axes that could not be deciphered and was assigned a Level 0 category, when a higher level response had been produced earlier. The levels of response by grade using the axes are shown in Table 7.

Table 7. Summary of levels by grade for the LOLLIES protocol with axes provided (n = 54)

	Grade				
	P	3	5	7	9
Level 0	1	2	0	1	1
Level 1	5	2	1	0	0
Level 2	1	6	10	1	3
Level 3	0	3	0	6	7
Level 4	0	0	0	2	2
Total	7	13	11	10	13

Using the axes provided, six students in Grade 3 and eight students in Grade 5 improved from Level 0 or 1 to Level 2. Some graphs took into account the total of 40 people but others ignored this aspect of frequency. The graphs shown in Figure 19 show variation but not expectation in the center. Similarly 2 Grade 3, 6 Grade 7, and 5 Grade 9 students improved their responses to Level 3 with the axes, by indicating an expectation for values around 5. Two of these are shown in Figure 20.

Two students in each of Grade 7 and Grade 9 improved their responses to Level 4 when presented with axes. These graphs show an appropriate shape for the distribution of outcomes except that the variation is too great and the responses ignore the values on the vertical axis. Two are shown in Figure 21, the first apparently attempting to record the 40 outcomes. A few other students produced distributions that were centered on values greater than 5. These, because they acknowledged a center, although not the appropriate one, were allocated to Level 3.

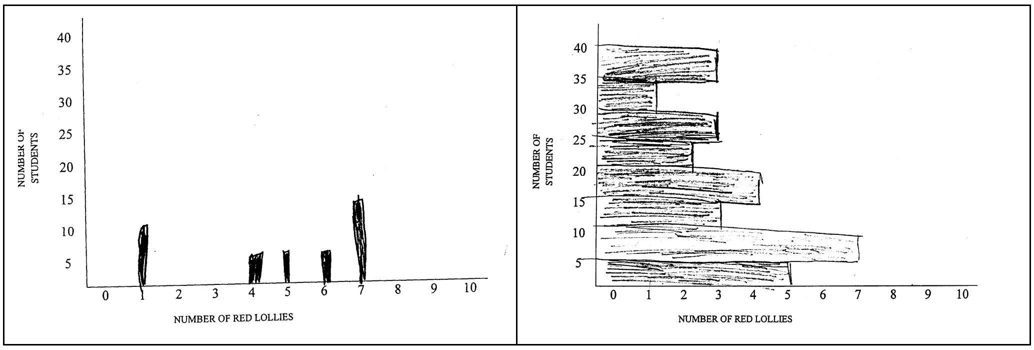


Figure 19. Frequency graphs with axes (Level 2)

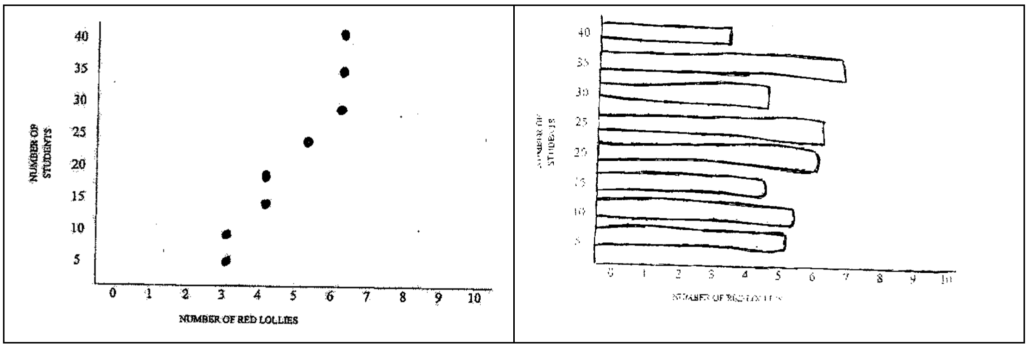


Figure 20. Frequency graphs with axes and a central expectation (Level 3)

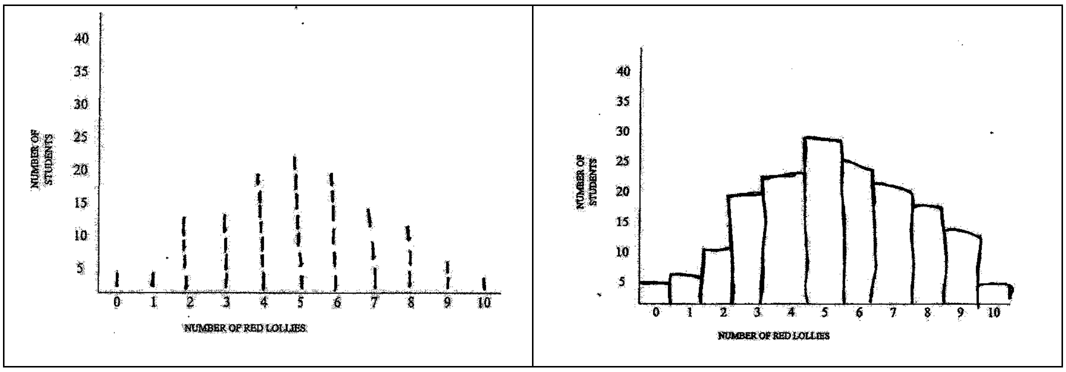


Figure 21. Frequency distributions with axes (Level 4)

6.4. SUMMARY

A summary of levels of response across the three tasks is given in Table 8. The difference in complexity of the tasks means that it is inappropriate to equate performance across tasks. What is of interest is the similarity in structure in the observed representations created by students. On all three tasks five hierarchical levels of sophistication in representing variation are seen, with expectation being acknowledged and represented at higher levels. There were only two instances, for the LOLLIES task, where students appeared to represent expectation in terms of probability, rather than variation. This may be associated with classroom instruction.

The trend for higher level performance with increasing grade likely reflects experiences in the classroom with ideas of average, probability, and graphing. An appreciation of variation in these contexts, however, appears to be established for most students by the middle years.

7. DISCUSSION

7.1. VARIATION, EXPECTATION, AND DISTRIBUTION

What issues are involved when distributions are being judged in relation to the appropriateness of the variation and expectation displayed? In the light of the growing interest in variation in recent years, Shaughnessy (2007) suggests eight different aspects of variation that arise in various statistical contexts. Most of these can be observed in graphical representations: (i) variation in particular values such as outliers, (ii) variation over time, (iii) variation over an entire range, (iv) variation within a likely range, (v) variation from a fixed value such as a mean, (vi) variation in sums of residuals, (vii) variation in co-variation or association, and (viii) variation as distribution. Aspects (vi) and (vii) are beyond the scope of this study but (viii) is of interest in the sense of the creation of a distribution that displays the variation inherent in the creator's mind. Shaughnessy's description of (viii) focuses on variation "between or among a set of distributions" (p. 985), which goes one step further than the notion of variation inherent in a single distribution as observed in this study. Expectation fits into Shaughnessy's list at several places as a counter point to the concept of variation. Certainly expectation is a determinant of the "likely range" in aspect (iv) and the mean, or other fixed values such as proportion of red lollies, in aspect (v). It also underlies the last three aspects. It seems clear that different kinds of tasks, as presented here, require acknowledgement of different aspects of variation.

For the BOOKS task, the variation is present in the given data values and the interest is in how students choose to represent this. In some sense the lower levels of response observed for this task fall outside of Shaughnessy's (2007) categories. The appreciation of individual values, however, points to an initial requirement of representing variation. The appreciation of the entire range of values was shown by a few students who mentioned it in the context of predicting how many books the new student might have read. Aspect (v), involving an appreciation of center, comes into play in some of the predictions and responses that acknowledge uncertainty in the prediction and appear to link the expectation with variation. Although the more sophisticated presentations in Figure 4 appear to satisfy statistical norms, the earlier representations are important in demonstrating the progression made by students in understanding the nature of the task. If progressions are recognized it should be easier for teachers to assist students in moving from less appropriate to more appropriate representations.

With the WEATHER and LOLLIES representations, students have a more complex task in representing variation because it is not presented to them in explicit data values. Only an expected value is presented at the start. An appreciation of variation in the context hence becomes important when students draw their graphs. For these two tasks students all seem to appreciate that the maximum temperature will not be the same every day and that the number of red lollies drawn from the container will not be the same every time. The prediction of six values or of a distribution for maximum temperature or number of reds, always shows variation, although sometimes it is wider than appropriate. The most appropriate graphs in a statistical sense show both a distribution, representing seasonal trend in maximum temperatures or likelihood of obtaining red lollies, and "random" variation about the distribution (see Figures 11 and 17). These two tasks certainly illustrate the first four of Shaughnessy's (2007) aspects of variation. Some responses include unusual values; some show variation in time for the weather or in sequential student draws for the lollies; some indicate variation over an entire range, particularly for the lollies task but also sometimes for temperatures; and some show appreciation of a limited likely range for both temperatures and numbers of lollies drawn. Although it may be considered implicit, the graphical representations that vary about a value of 17^{o}C on the vertical axis or peak at 5 red lollies, are showing an appreciation for Shaughnessy's aspect (v). There is also the additional aspect (ix) which reflects the contextual model that produces the distribution represented: the seasonal trends in temperature and the theoretical sampling distribution for the lollie draws. Again there appear to be several steps or stages in students' increasing appreciation of the variation and its link to expectation in the context of the overall tasks. An understanding of these

Table 8. Percent of responses for each task at each hierarchical level

	BOOKS					WEATHER					LOLLIES (initial)					LOLLIES (given axes)				
Grade	P	3	5	7	9	P	3	5	7	9	P	3	5	7	9	P	3	5	7	9
Idiosyncratic	14	50	25	7	12	100	33	11	0	0	57	78	39	27	13	14	15	0	10	8
Unistructured Variation	43	17	37	7	12	0	39	39	20	0	43	17	33	20	27	71	15	9	0	0
Variation shown by value	43	33	37	36	25	0	22	39	60	53	0	0	17	47	33	14	46	91	10	23
Initial acknowledgement of expectation	0	0	0	29	25	0	0	11	7	47	0	6	0	0	20	0	23	0	60	54
Integration of expectation and variation	0	0	0	21	25	0	0	0	13	0	0	0	11	7	7	0	0	0	20	15
n	7	6	8	14	8	4	18	18	15	15	7	18	18	15	15	7	13	11	10	13

will help teachers plan interventions to assist students in progressing to higher levels of representation and explanation.

Although the production of graphical representations should take place within the larger setting of a complete statistical investigation, for example as described by Chick and Watson (2001), Lehrer and Romberg (1996), Petrosino et al. (2003) and Friel et al. (2006), studying the products of such investigations is likely to provide insight into students' thinking during the process. These are particularly relevant to the inferences drawn. As well, issues related to the context within which a task is set are important. Are students more familiar with the weather than with pulling lollies unseen from a container? Is the pictograph task too elementary to interest older students? The three tasks were chosen specifically to provide both a range of complexity and data based on two processes: scientific measurement and artificial probability sampling. The use of contextual knowledge was seen most often in the WEATHER protocol, where students told of their experiences of Hobart's weather (Watson & Kelly, 2005); but also younger students used their contextual knowledge of reading to suggest imaginary reasons why Paul or Mary might have read a suggested number of books. In the BOOKS protocol, some students stated assumptions of context that could underlie their prediction for Paul and Mary. It is likely that it was limited contextual experience with pulling lollies from containers that contributed to the wider than realistic distributions drawn by students (e.g., Figure 21). The issue of the influence of contextual knowledge on students' inferences in data handling situations is beginning to attract research attention (e.g., Langrall, Nisbett, & Mooney, 2006; Mooney, Langrall, & Nisbet, 2006) and should be expanded to include the attention paid to its influence on variation and expectation in the creation of graphical representations.

The use of the term "distribution" in the title reflects the statistical perspective in relation to what is expected by the time students move into senior secondary study. It is unlikely that students will use the word before then. They will however hopefully draw many graphs that show appropriate variation associated with the contexts of tasks set. If they learn the importance of the words "variation" and "expectation," this will be an important part of the vocabulary for their later statistical lives.

7.2. LIMITATIONS

Some of the limitations of the study result from combining data sets for tasks that were not all completed by the same students. Although it is possible to compare and contrast representations by the students who completed the LOLLIES and WEATHER tasks, it has not been done for this paper (see Watson, Callingham, & Kelly, 2007). Although some students explained well their thinking while creating representations during the interviews, others said very little. It is possible that further probing might have produced more complete explanations.

It may be considered that the questions in the protocols, particularly the WEATHER and LOLLIES tasks, encouraged a consideration of variation. Each set up the potential for comparing varied data values against an initial expectation. In the WEATHER protocol, this was provided by the statement about the average daily maximum temperature in Hobart being 17^{o}C. In the LOLLIES protocol, expectation was provided in the statement that the bowl of lollies contained 50 red, 20 yellow, and 30 green lollies. The first was a more straightforward and familiar statement of expectation in context for most students. In both cases, however, students were presented with both concepts, expectation and variation, at the start of the protocol and hence had the opportunity to build both into their responses. In the BOOKS task, the opportunity for display of understanding of the two concepts appeared as pictographs were discussed or predictions requested. Although creating pictographs that displayed variation, some students did not recognize this as a characteristic that could be discussed. In other protocols presented to these students, not associated with graphical representations, expectation was the feature of the prompting questions (Watson & Kelly, 2004b, 2006) not a contextual referent as in the WEATHER and LOLLIES protocols.

7.3. IMPLICATIONS FOR RESEARCH AND THE CLASSROOM

This study adds to the growing body of evidence about the increased complexity of appreciation of variation and expectation in statistical contexts throughout the school years. In two of the contexts presented here, specific data values were not presented to students; in the other context a very small data set was presented. It may hence be claimed that the contexts were not realistic and students were not encouraged to use knowledge they may have learned in the classroom to deal with quantitative data sets. Whether this may have influenced the apparent delay in demonstration of ideas associated with expectation is unknown. The two contexts that did not contain specific data sets, however, may represent scenarios more likely to be encountered in out-of-school situations. They may perhaps present evidence of how likely or otherwise it is for students to transfer their knowledge to less specifically data-based environments.

In a related study involving six protocols, including the WEATHER and LOLLIES tasks but also others more specifically aimed at probability as expectation, Watson et al. (2007) observed a parallel development of concepts related to variation and expectation. Based on a Rasch analysis of hierarchically coded responses, six levels were identified, ranging from no acknowledging of either variation or expectation to an establishing of links between the two in comparative settings employing proportional reasoning. It appears that the use of more tasks, some specifically addressing expectation, prompts students to display their developing appreciation earlier. In the current study the tasks, especially the BOOKS protocol, were quite open-ended, allowing students to display understanding they felt to be relevant rather than to be prompted to recall averages or probabilities.

This study holds open the question of the natural development of ideas of variation and expectation, free of teaching intervention or specific prompting during interviews. It appears to support the view that ideas of variation develop naturally before those of expectation. Watson (2005) produced a similar argument based on descriptive anecdotal examples and further suggested that the school curriculum does not reflect this development. It may be that currently the curriculum does not support students' natural inclination to focus on variation and instead forces attention on expectation in the form of averages and probabilities first. It would appear that if it is desired for students to develop both concepts together then the curriculum needs to reflect the two ideas and their interaction from the start. It is likely that David Moore (1990, 1997) would support this revision in thinking about the curriculum given his view of variation as the foundation concept underpinning the field of statistics.

The tasks used in this study illustrate a wide range of contexts within which students can be asked to create representations of distributions. The importance of considering both data sets and data-free scenarios is seen, as well as the importance of choosing contexts where students have some intuition about the variation present. In school settings it may be possible to combine such graphing tasks with other tasks in science, social science, or health where variation appears in the topic being studied. The consideration of both subject matter understanding and the ability to create distributions, leads to the implications for assessment.

If a hierarchical progression of observed levels of graph production can be agreed upon, for example as suggested in Table 3, it will then be possible to create rubrics for assessment based upon them. These can then be combined with other rubrics of subject matter performance for authentic cross-disciplinary assessment, as desired in many of today's schools. The results of this study may not show that students get close to the formal idea of distribution by Grade 9 but they indicate what a complex process is involved. The outcomes suggest that in making progress much explicit classroom discussion is required along the way.

ACKNOWLEDGMENTS

The data presented in this paper were collected as part of two Australian Research Council Grants (A79231392 and A00000716). An earlier paper, based on these data and other multi-variate data, was presented at the Fourth International Forum on Statistical Reasoning Thinking and Literacy in Auckland, New Zealand in 2005. Jonathan Moritz and Ben Kelly conducted some of the interviews and contributed to the data analysis.

REFERENCES

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Dordrecht: Kluwer Academic Publishers.

Ben-Zvi, D., & Amir, Y. (2005). How do primary students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about Distribution: A collection of studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4).* [CDROM, with video segments]. Brisbane, Australia: University of Queensland.

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). London: Routledge.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Burns, R. B. (2000). *Introduction to research methods* (4th ed.). London: Sage.

Chick, H. (2004). Simple strategies for dealing with data. *Australian Mathematics Teacher, 60*(3), 20-24.

Chick, H. L., & Watson, J. M. (2001). Data representation and interpretation by primary school students working in groups. *Mathematics Education Research Journal, 13*, 91-111.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*, 382-393.

Curcio, F. R., & Artzt, A. F. (1996). Assessing students' ability to analyze data: Reaching beyond computation. *Mathematics Teacher, 89*(8), 668-673.

delMas, R. C., & Liu, Y. (2003). Exploring students' understanding of statistical variation. In C. Lee (Ed.), *Reasoning about variability: Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy* [CD-ROM]. Mt. Pleasant, MI: Central Michigan University.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124-158.

Friel, S. N., O'Connor, W., & Mamer, J. D. (2006). More than "Meanmedianmode" and a bar graph: What's needed to have a statistical conversation. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance* (pp. 117-137). Reston, VA: National Council of Teachers of Mathematics.

Green, D. (1993). Data analysis: What research do we need? In L. Pereira-Mendoza (Ed.), *Introducing data analysis in the schools: Who should teach it?* (pp. 219-239). Voorburg, The Netherlands: International Statistical Institute.

Kelly, B. A., & Watson, J. M. (2002). Variation in a chance sampling setting: The lollies task. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific: Proceedings of the twenty-sixth annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 366-373). Sydney, NSW: MERGA.

Kerslake, D. (1981). Graphs. In K. M. Hart (Ed.), *Children's understanding of mathematics: 11-16* (pp. 120-136). London: John Murray.

Kirkpatrick, E. M. (Ed.). (1983). *Chambers 20th century dictionary*. Edinburgh: W & R Chambers.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), A *research companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T. L., Russell, S. J., & Khalil, K. (2003). *Data seen through different lenses*. Amherst, MA: University of Massachusetts.
[Online: www.umass.edu/srri/serg/papers.html]

Langrall, C., Nisbet, S., & Mooney, E. (2006). The interplay between students' statistical knowledge and context knowledge in analyzing data. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/2A3_LANG.pdf]

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, *14*(1), 69-108.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs and graphing: Tasks, learning and teaching. *Review of Educational Research*, *60*(1), 1-64.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, *32*, 229-263.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mooney, E., Langrall, C., & Nisbet, S. (2006). Developing a model to describe the use of context knowledge in data explorations. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/2A4_MOON.pdf]

Moore, D. S. (1990). Uncertainty. In L. S. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2), 123-165.

Moore, D. S., & McCabe, G. P. (1993). *Introduction to the practice of statistics* (2nd ed.). New York: W. H. Freeman.

Moritz, J. B. (2000). Graphical representations of statistical associations by upper primary students. In J. Bana & A. Chapman (Eds.), *Mathematics education beyond 2000: Proceedings of the twenty-third annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 440-447). Sydney: MERGA.

Moritz, J. B. (2002). Study times and test scores: What students' graphs show. *Australian Primary Mathematics Classroom*, *7*(1), 24-31.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Pegg, J. E. (2002a). Assessment in mathematics: A developmental approach. In J. M. Royer (Ed.), *Mathematical cognition* (pp. 227-259). Greenwich, CT: Information Age Publishing.

Pegg, J. E. (2002b). Fundamental cycles of cognitive growth. In A. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 41-48). Norwich, UK: University of East Anglia.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131-156.

Pfannkuch, M., Rubick, A., & Yoon, C. (2002). Statistical thinking and transnumeration. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific: Proceedings of the twenty-fifth annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 567-574). Sydney: MERGA.

Reading, C., & Shaughnessy, M. (2000). Student perceptions of variation in a sampling situation. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 89-96). Hiroshima, Japan: Hiroshima University.

Reading, C., & Shaughnessy, M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Dordrecht: Kluwer Academic Publishers.

Roth, W-H., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction, 21*(4), 429-473.

Russell, S. J. (2006). What does it mean that "5 has a lot"? From the world to data and back. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance* (pp. 17-29). Reston, VA: National Council of Teachers of Mathematics.

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 twentieth annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6-22). Waikato, NZ: MERGA.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: Information Age Publishing.

Swan, M. (1988). Learning the language of functions and graphs. In J. Pegg (Ed.), *Mathematics Interfaces: Proceedings of the twelfth biennial conference of The Australian Association of Mathematics Teachers* (pp. 76-80). Newcastle, NSW: The New England Mathematical Association.

Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, *47*, 337-372.

Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, *51*, 225-256.

Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: Theory, research and practice* (Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia, Melbourne, pp. 35-42). Sydney: MERGA.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

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*, 83-130.

Watson, J. M., & Kelly, B. A. (2002a). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/2a1_wats.pdf]

Watson, J. M., & Kelly, B. A. (2002b). Emerging concepts in chance and data. *Australian Journal of Early Childhood*, *27*(4), 24-28.

Watson, J. M., & Kelly, B. A. (2004a). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics*, *57*, 121-144.

Watson, J. M., & Kelly, B. A. (2004b). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics and Technology Education*, *4*, 371-396.

Watson, J. M., & Kelly, B. A. (2005). The winds are variable: Student intuitions about variation. *School Science and Mathematics*, *105*, 252-269.

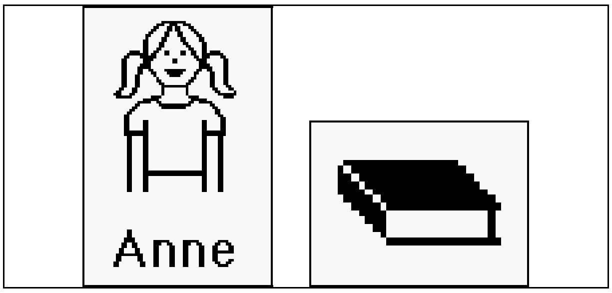
Watson, J. M., & Kelly, B. A. (2006). Expectation versus variation: Students' decision making in a sampling environment. *Canadian Journal of Science, Mathematics and Technology Education*, *6*, 145-166.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, *31*, 44-70.

Watson, J. M., & Moritz, J. B. (2001). Development of reasoning associated with pictographs: representing, interpreting, and predicting. *Educational Studies in Mathematics*, *48*, 47-81.

JANE M. WATSON
Faculty of Education
University of Tasmania
Private Bag 66
Hobart, Tasmania 7001
Australia

APPENDIX A: INTERVIEW PROTOCOL FOR THE BOOKS TASK



Sample of cards used as materials for representation.

We have some cards here, to represent some children, and some cards for the books they have read.
(Show information sheet: Anne read 4 books, and Danny read 1, and Lisa read 6, Terry read 3.)
Now suppose that Anne read 4 books, and Danny read 1, and Lisa read 6, Terry read 3.
Representing (Part 1)
Can you use the cards to show the information?
Why did you do it that way?
Interpreting (Part 1)
If someone came into the room, what could they tell by looking at your picture?
Representing (Part 2)
Suppose Andrew read 5 books. Can you show that Andrew read 5 books?
Suppose Jane read 4 books. Can you show that Jane read 4 books?
Now, suppose Ian hasn't read any books. Can you show that Ian hasn't read any books?
Now, suppose everyone went to the library and read one more book each. Can you change your picture to show that they all read one more book each?
Interpreting (Part 2)
If someone came into the room, what could they tell by looking at your picture now?
Can you tell who likes reading the most? How?
Can you tell how many books they've read all together?
Who do you think is most likely to want a book for Christmas? Why do you think that?
Predicting
Suppose Paul came along, and we didn't know how many books he had read. What would be you best estimate/prediction/guess of how many books he might have read?
Now suppose Mary came along. What would be your best estimate/prediction/guess of how many books she might have read?

APPENDIX B: INTERVIEW PROTOCOL FOR THE WEATHER TASK

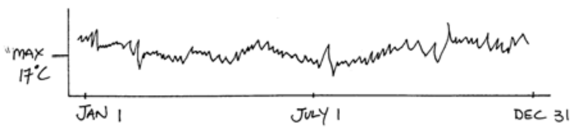
1. Some students watched the news every night for a year, and recorded the daily maximum temperature in Hobart. They found that the average maximum temperature in Hobart was 17°C.
    a) What does this tell us about the temperature in Hobart?
    b) Do you think all the days had a maximum of 17°C? - Why or why not?
    c) (What do you think the maximum temperature in Hobart might be for 6 different days in the year?)* ______, ______, ______, ______, ______, ______
    d) Why did you make these choices?

    e) For the whole year, what do you think the highest and lowest daily maximum temperature in Hobart would be? highest maximum _____ lowest maximum ____
    f) For the month of January, what do you think the highest and lowest daily maximum temperature in Hobart would be? highest maximum _____ lowest maximum ____
    g) For the month of July, what do you think the highest and lowest daily maximum temperature in Hobart would be? highest maximum _____ lowest maximum ____

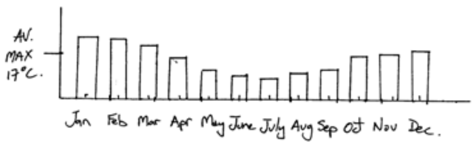
2. How would you describe the temperature for Hobart over a year in a graph?

3. Here are some ideas from other students. What do you think of them?

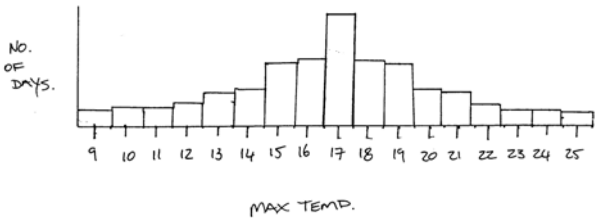
a)



b)



c)



APPENDIX C: INTERVIEW PROTOCOL FOR THE LOLLIES TASK

1. Suppose you have a container with 100 lollies in it. 50 are red, 20 are yellow, and 30 are green. The lollies are all mixed up in the container. You pull out 10 lollies.

a) How many reds do you expect to get?
b) Suppose you did this several times. Do you think this many would come out every time? Why do you think this?
c) How many reds would surprise you? Why do you think this?

2. Suppose six of you do this experiment.

a) What do you think is likely to occur for the numbers of red lollies that are written down?
______, ______, ______, ______, ______, ______ Why do you think this?

3. Look at these possibilities that some students have written down for the numbers they thought likely.
(a) 5,9,7,6,8,7 (b) 3,7,5,8,5,4 (c) 5,5,5,5,5,5 (d) 2,3,4,3,4,4
(e) 7,7,7,7,7,7 (f) 3,0,9,2,8,5 (g) 10,10,10,10,10,10
Which one of these lists do you think best describes what might happen? Why do you think this?

4. Suppose that 6 students did the experiment. What do you think the numbers will most likely go from and to?
From __________ (lowest) to __________ (highest) number of reds. Why do you think this?
Now try it for yourself: ______, ______, ______, ______, ______, ______
Given the results, do you want to change any of your previous answers?

5. Suppose that 40 students pulled out 10 lollies from the container, wrote down the number of reds, put them back, mixed them up.
a) Can you show what the number of reds look like in this case? (Use the blank space below)
b) Now use the graph below to show what the number of reds might look like for the 40 students.