STUDENT DESCRIPTION OF VARIATION WHILE WORKING WITH WEATHER DATA

CHRIS READING
University of New England
creading@metz.une.edu.au

SUMMARY

Variation is a key concept in the study of statistics and its understanding is a crucial aspect of most statistically related tasks. This study aimed to extend and apply a hierarchy for describing students' understanding of variation that was developed in a sampling context to the context of a natural event in which variation occurs. Students aged 13 to 17 engaged in an inference task that necessitated the description of both rainfall and temperature data. The SOLO Taxonomy was used as a framework for analyzing student responses. Two cycles of Unistructural-Multistructural-Relational levels, one for qualitative descriptions and the other for quantitative descriptions, were identified in responses. Implications of the extended hierarchy for describing understanding of variation for research, teaching and assessment are outlined.

Keywords: Describing variation; SOLO Taxonomy; Inference task; Secondary students

1. INTRODUCTION

The analysis of variation, that is, the irregularities in data, is critical to the study of statistics (Wild & Pfannkuch, 1999, p. 235). Despite this critical nature of variation, not much is known about how students perceive variation. A prior review of the literature has shown that, despite the importance of variation, most research examines the understanding of central tendency and that research on understanding of variation is limited (Shaughnessy, Watson, Moritz, & Reading, 1999). In fact, the work by Shaughnessy et. al (1999) is one of the first attempts to unpack, in a systematic way, what is happening in students' understanding of variation. Given that variation is critical to the study of statistics, more research needs to be undertaken to better understand how students view and describe variation. This study was undertaken to develop a hierarchy to assess students' understanding of variation. The results of this study are expected to assist researchers and teachers by providing a tool for describing the level of statistical sophistication in the description of variation.

1.1. STUDENTS' PROPENSITY TO DISCUSS VARIATION

When dealing with data, consideration needs to be given to both measures of central tendency and measures of variation. So, which of these are students more likely to use if not prompted when working with data? Research has shown that when engaged in reducing data, although some students base their responses on measures of variation, many more students use measures of central tendency (Reading & Pegg, 1996, p.190). This investigation involved Australian secondary school students for whom most data reduction learning experiences deal with finding 'mean, mode and median', hence it is not surprising that so few bother with measures of variation. On the other hand, in Australian schools students are presented with few learning experiences that involve making inferences from data and generally are not given specific instruction as to how to engage in such activities. Thus, responses to tasks that involve making inferences are less likely to reflect approaches imposed by teachers. In fact, analysis of secondary school student responses to open-ended questions involving

Statistics Education Research Journal 3(2), 84-105, http://www.stat.auckland.ac.nz/serj

making inferences from data showed that more students based an inference discussion on measures of variation than measures of central tendency (Reading, 1998, p. 1430).

The conflicting student inclinations in these two different experiences, engaged in reducing data and engaged in making inferences about data, suggest that students may have a propensity to consider measures of variation when dealing with data but unless this is given a chance to develop, such a propensity may eventually be overcome by the ‘push from teachers’ to discuss measures of central tendency. Even more importantly though, if teachers concentrate their efforts on working with measures of central tendency, then students will be denied the opportunity to experience situations where they can begin to understand variation and to develop any propensity they may have to reason about variation. Already researchers are recognizing the need to develop learning situations where students can be encouraged to develop the notion of variability. One such approach is Bakker’s (2003) ‘growing samples’ activity that allows students to investigate the shape of distributions as a basis for developing a better understanding of variability. However, the present research was not designed to determine whether early attempts at inference by students are more likely to be based on measures of central tendency or variation but, to consider aspects of reasoning about variation that do become apparent when students make inferences.

1.2. CONSIDERATION OF VARIATION

The study of measures of variation in schools, such as the standard deviation, has developed notoriety with teachers as being particularly cumbersome, resulting in many teachers having difficulty developing the concept with students or avoiding it altogether. This is unfortunate given that many students show, at least in some contexts, a natural propensity to base discussion of data on measures of variation rather than central tendency (Reading, 1998). In order to be better prepared to equip students with an understanding of variation, teachers need to understand how students reason about variation and also to have a means for assessing how students reason about variation.

Concern over lack of attention to variation has prompted researchers to investigate in more detail students’ understanding of variation. Some studies were undertaken following dissatisfaction with the responses of Grade 4 students in the USA to a National Assessment of Educational Progress (NAEP) test item (Shaughnessy et al., 1999; Reading & Shaughnessy, 2000). In this extended response item, students had to predict the number of red gum balls in a sample of ten obtained from a gum ball machine, and then explain their reason(s) for choosing that number. This task allowed students to demonstrate their understanding of *centrality* (expected outcome according to formal probability calculation) but not *variability* (in outcomes across repeated trials). Shaughnessy et al. (1999) redesigned this task and analyzed pencil and paper responses to gain useful information about students’ conceptions of variation in a similar sampling situation based around a candy bowl rather than a gum ball machine. These modified investigations have been extended in various contexts (Torok & Watson, 2000). In particular, Shaughnessy and Ciancetta (2002) allowed students to experience the variability in results, with ten trials of a spinner task, before predicting the outcomes.

When outlining the foundations of ‘thinking statistically’ Wild and Pfannkuch (1999, p. 226) identify ‘consideration of variation’ as one of the fundamental types of thinking. They list four components of consideration of variation: noticing and acknowledging variation; measuring and modelling variation for the purpose of prediction, explanation or control; explaining and dealing with variation; and developing investigative strategies in relation to variation. Reading and Shaughnessy (2004) have also suggested two additional components: describing variation; and representing variation. To best investigate students’ reasoning about variation it is necessary to delve into as many as possible of these components, and examine how students describe the variation they observe and endeavour to interpret and/or use for inference.

2. DEVELOPMENTAL HIERARCHIES

The increasing popularity of research into cognitive frameworks to assess students’ understanding of phenomena when responding to learning or assessment activities has provided the impetus for the creation of developmental hierarchies in stochastics. Following is an introduction to a particular model for explaining developmental growth and then a summary of aspects of existing developmental hierarchies that particularly address variation.

2.1. THE SOLO TAXONOMY

Research into Developmental-Based Assessment (DBA), the assessment of students based on the quality of their understanding and learning (Pegg, 2003), has contributed to the increased acceptance of developmental frameworks. This approach to assessment, which focuses on the mental structure of understanding, differs from outcomes-based assessment which focuses on what students are expected to know. This paper focuses on the Structure of the Observed Learning Outcome (SOLO) Model, an approach to assessment which rests on an empirically established cognitive developmental model (Pegg, 2003).

The neo-piagetian SOLO Taxonomy (Biggs & Collis, 1991) consists of five modes of functioning, with levels of achievement identifiable within each of these modes. The two modes relevant to the present research are the ikonic mode (making use of imaging and imagination) and the concrete symbolic mode (operating with second order symbol systems such as written language). Although these modes are similar to Piagetian stages, an important difference is that with the SOLO Taxonomy earlier modes are not seen as replaced by subsequent modes and in fact are often being used to support growth in the later modes.

A series of levels of increasing cognitive development has been identified within each of these modes. The three levels relevant to this study are: *unistructural* responses - with focus on one element, *multistructural* responses - with focus on several unrelated elements, and *relational* responses - with focus on several elements in which inter-relationships are identified. These three levels form a cycle of cognitive growth, from unistructural, through multistructural, to relational responses, that occurs within a mode. For example, when describing a geometric figure, students may focus on an element such as a ‘property of the figure’. Unistructural responses would describe one property of the figure, perhaps focusing on the lengths of the sides. Multistructural responses would address more than one property, perhaps the lengths of the sides and sizes of the angles. Relational responses would identify links and deal with a relationship between the properties, perhaps stating that adjacent angles being right angles would imply pairs of parallel sides. The relational level response in one cycle is similar to, but not as concise as, the unistructural response in the next cycle. Early applications of SOLO only described one cycle of levels within each mode, but more recently researchers have identified more than one cycle of levels within a mode (Pegg, 2003, pp. 244-245). This taxonomy is particularly useful because of the depth of analysis that can be achieved when interpreting students’ responses.

2.2. DEVELOPMENTAL HIERARCHIES FOCUSING ON VARIATION

Neo-Piagetians have provided a foundation of cognitive frameworks on which to base developmental hierarchies in probability (e.g., Jones, Langrall, Thornton & Mogill, 1997) and in statistics (e.g., Mooney, 2002). SOLO has already been employed to explain statistical thinking frameworks (e.g., Jones et al., 2000; and refined by Mooney, 2002) and is more recently being used as the basis for development hierarchies related to variation (e.g., Watson, Kelly, Callingham & Shaughnessy, 2003). Mooney (2002) developed four SOLO based ‘levels’ in each of four processes. Variation is only mentioned in one of the four processes, ‘organizing and reducing data’, and the relevant descriptors for that process (Mooney, 2002, pp. 36-37) are reproduced in Table 1. A series of studies reported by Jones, Mooney, Langrall and Thornton (2002) was used to validate the SOLO-based levels.

Table 1. Partially reproduced Statistical Thinking Framework (Mooney, 2002)

	Organising and Reducing Data
Levels	**Focus of Responses**
1 - Idiosyncratic	Is not able to describe the spread of the data in terms representative of the spread.
2 - Transitional	Describes the spread of the data using invented measures that are partially valid.
3 - Quantitative	Describes the spread of the data using a measure from a flawed procedure or a valid and correct invented measure.
4 - Analytical	Describes the spread of data using a valid and correct measure.

Watson et al. (2003) used the Torok and Watson (2000) hierarchy levels, in conjunction with SOLO, as a starting point for the analysis of responses to a bank of assessment items, culminating in the description of four levels for the understanding of statistical variation (Watson et al., 2003, p. 11) described in Table 2. Although these four levels were developed to measure understanding of variation they do not explain how students actually describe the variation.

Table 2. Developing Concepts of Variation (Watson et al., 2003)

Levels	Focus of Responses
1 - Prerequisites for variation	Working out the environment, table/simple graph reading, intuitive reasoning for chance.
2 - Partial recognition of variation	Putting ideas in context, tendency to focus on single aspects and neglect others.
3 - Application of variation	Consolidating and using ideas in context, inconsistent in picking most salient features.
4 - Critical aspects of variation	Employing complex justification or critical reasoning.

Reading and Shaughnessy (2000) interviewed 12 students regarding the sampling task used earlier by Shaughnessy et al. (1999) and identified non-sophisticated discussion of variation in this context. These interviews were further analyzed and two hierarchies for understanding of variation, one for description and the other for causation, were developed (Reading & Shaughnessy, 2004) based on students' perceptions in the sampling situation. Only the Description Hierarchy is relevant to the present study and the four levels of this hierarchy are summarized in Table 3.

Table 3. Description Hierarchy (Reading & Shaughnessy, 2004)

Levels	Focus of Responses
D1 - Concern with Either Middle Values or Extreme Values	Describe variation in terms of what is happening with either extreme values or middle values. *Extreme Values* are used to indicate data items that are at the uppermost or lowest end of the data, while *Middle Values* indicate those data items that are between the extremes.
D2 - Concern with Both Middle Values and Extreme Values	Describe variation using both the extreme values and what is happening with the values between the extremes.
D3 - Discuss Deviations from an Anchor	Describe variation in terms of deviations from some value but either the anchor for such deviations is not central, or not specifically identified as central.
D4 - Discuss Deviations from a Central Anchor	Describe variation by considering both a centre and what is happening about that centre.

The Reading and Shaughnessy (2004) Description Hierarchy was based on student responses to a sampling task but did not use SOLO as a conceptual framework. The present study was designed to use, and modify or extend if necessary, this hierarchy to code responses given in a weather-related inference task and to consider SOLO as a suitable conceptual framework to explain the hierarchy. A weather-based task was chosen because weather is a phenomenon which involves variation that everyone experiences, hence it can provide students with a meaningful context for data description and inference.

The main research question was how students describe variation during an inference task. This necessitated investigation of three related research questions: Is the hierarchy developed for analyzing students' descriptions of variation in a sampling situation (Reading & Shaughnessy, 2004) also applicable for coding responses with data descriptions given when making inferences from weather-related data, in which there is natural variation? If this hierarchy is suitable, does SOLO offer a broad framework for explaining the hierarchy? If SOLO is a suitable framework, can cycles of levels be identified within the SOLO modes? The findings would contribute to the refinement of conceptual models developed in earlier research, and could assist researchers and teachers by providing a developmentally-based hierarchy.

3. METHOD

This section describes an exploratory study that involved posing to students in Grades 7, 9 and 11 a weather-related inference task with two separate segments. The following describes the task students faced, procedure, participants, and analytic approach and associated issues.

3.1. WEATHER ACTIVITY – STUDENT TASK

The Weather Activity presented students with a Scenario, see Figure 1, based around choosing the most suitable month for a proposed Youth Festival to be held in the students' own town. It was stressed to students that they did not have to worry about any other aspects of the celebration, only the weather. The activity was implemented over a period of time and incorporated both data description and inference components.

WEATHER ACTIVITY SCENARIO

XXXXXX is to introduce a new celebration into the calendar. *Youth Alive* will celebrate the youth of the city and be held at an outdoor venue. Although not all details have been decided concerning the activities to be held on the day, a decision needs to be made as to a date for the celebration so that it can be slotted into the calendar. Organisers have expressed concern as to the effect that XXXXXX's often unpredictable weather could have on such a celebration. You have been commissioned to submit a report to the committee describing XXXXXX's weather and to suggest a suitable month for the celebration. Other factors will be taken into consideration to decide exactly which day in the month *Youth Alive* will be held.

Figure 1. Weather activity scenario

The activity was designed to have two separate segments, the first segment based around rainfall data and the second segment based around temperature data. Before each segment began, each student was randomly allocated one particular month of data to consider. The data used by the students in the task consisted of rainfall figures (daily millimetres) for 36 months in the first segment, and temperature figures (daily minimum and maximum temperatures in degrees celsius) for 36 months in the second segment. The use of 36 different months ensured that each student in the class had different data. The weather was chosen for the three years 1998 through to 2000 because the activity was undertaken in 2001. Examples of the data as presented to the students appear in the Appendix. It should be noted that the monthly data provided to each student exhibited different patterns and different variability.

Within each segment students had an individual task followed by a group task that were both open ended. First, each student was asked to individually describe the weather in his or her month in a written response. The students were told that these descriptions were to be used in the next step to compare with the descriptions provided for other months by a small group (about four) of classmates and decide on the most suitable month of the year for the festival from amongst those months within the group. Later, the students worked together in these groups, comparing their data descriptions, and developing a written group response that both described what they chose as the most suitable month for the festival, out of those they compared, and explained their reasoning.

The use of open-ended tasks meant that students had the freedom to adopt criteria or attend to issues they considered necessary. No specific instruction was given to discuss variation, despite the fact that description of variation was the focus of the research. This approach was taken so that students were free to discuss variation how and when they saw the need. This methodological approach has been utilized by other researchers. Watson et al. (2003) designed items to allow students to be free to demonstrate their understanding of variation and Ben-Zvi's (2003) open learning activity gave no specific direction to discuss variability despite the fact that analysis of the data was to include how students reasoned about variation.

3.2. PROCEDURE

The weather activity task was implemented in classes, during normal teaching time by a research assistant. Allocation of data sets to students was random but allocation of students to groups was not. The random allocation of data meant that students did not necessarily receive the same month for the temperature segment as they had used for the rainfall segment. The allocation of students into groups for the group activity was made by the teacher, based on knowledge about good working relationships from previous group work. If students were unsure about the task and needed a prompt, then it was suggested that they should look for any pattern in the data or any key features that may be useful.

The weather activity was planned to spread over a number of weeks to suit the school schedule. The individual and group work for each segment was planned to occur during one standard lesson time-slot for the class. The temperature segment lesson, however, did not immediately follow the rainfall segment lesson. Between the two segments of the activity there was a teaching episode, presenting a statistics-related section of the curriculum. These teaching episodes were requested by the teacher to align the weather activity with the students' learning experience. The episodes were implemented by the class teacher and involved demonstrating to the class statistical tools that might be useful when describing the data. Grade 7 students were introduced to stem-and-leaf plots and the summary statistics: maximum, minimum, average, and range. Grade 9 students were introduced to box-and-whisker plots, from a development base of stem-and-leaf plots with which they were already familiar. In Grade 11 an entire unit of work on statistics, including stem-and-leaf plots and box-and-whisker plots, was implemented between the two segments of the activity.

The weather activity provided both individual and group written responses for analysis. However, only the analysis of the individual responses will be reported here. For discussion of the group responses see Reading and Lawrie (2004). A wrap-up activity planned as the last segment of the weather activity, having individual students make a final decision about the most appropriate month with all data discussions available to them from all groups, was not completed because there was insufficient time due to the intervention of other events in the school. However, this did not detract from the usefulness of the responses provided in each of the rainfall segment and temperature segment as the students did not know at that point in time that the wrap-up activity would not be completed.

It is acknowledged that students could have developed their understanding regarding variation during the group discussion part of the rainfall segment and the teaching episode, in ways which could affect the quality of their later responses to the first (individual) part of the temperature segment. However, as the research was attempting to refine a hierarchy for coding responses and not to assess a student's performance at any particular instant or to compare performance before and after instruction, the possible improvements in quality of response in fact would provide a richer array of data for refining the hierarchy. This sequence of activities enabled the researchers to capture the reasoning of a heterogeneous set of students at different points during their exposure to data with natural variation.

3.3. PARTICIPANTS

This research targeted students in Grades 7, 9 and 11 (aged 13 to 17) in a secondary school in a rural city in northern New South Wales, Australia. Students from one class in each of the three grades were included in the study. Classes were selected so as to include students with average mathematical ability. Only two teachers from the school were involved, as one teacher had charge of two of the classes. The actual number of students who completed each of the individual steps of the research activity was not consistent, as attendance in each class varied over the particular days when activities were presented. The breakdown of students participating in the weather activity is presented for the rainfall segment in Table 4 and for the temperature segment in Table 5.

Table 4. Participants in the rainfall segment of the weather activity

	Grade 7	Grade 9	Grade 11	Total
Male	15	17	7	39
Female	6	11	9	26
Total	21	28	16	65

Table 5. Participants in the temperature segment of the weather activity

	Grade 7	Grade 9	Grade 11	Total
Male	16	15	9	40
Female	5	9	10	24
Total	21	24	19	64

3.4. ANALYTIC APPROACH AND ASSOCIATED ISSUES

The purpose of the analysis of the individual responses was to determine the applicability of the Reading and Shaughnessy (2004) hierarchy, developed for a sampling situation, to coding responses to an inference task and refine the hierarchy if needed. Coding of the written discussions in the responses was undertaken in three stages. First, the responses were coded independently by the researcher and the research assistant, based on the Reading and Shaughnessy (2002) hierarchy in Table 3. Second, level descriptions were revised based on any newly identified descriptions of features of variation and the hierarchy was expanded by developing new levels based on responses not suitably accommodated. Such a revision process for coding hierarchies has been utilized by a number of researchers (e.g. Mooney, 2002; Langrall & Mooney, 2002; Watson et al., 2003). Finally, all responses were recoded independently by both the researcher and research assistant based on the new hierarchy. The recoding produced an 85% agreement and then discussion was used to resolve disputed codings. Such discussions also helped to refine the clarification of each level in the hierarchy. Before describing the hierarchy in detail some relevant aspects of student performance on the task are outlined.

Students were given access to graph paper but were not specifically required to produce a graph. While a number of students in each grade (especially in Grade 7, as shown in Table 6) chose to draw a graph as part of their response, only three responses (one Grade 9 and two Grade 11) actually referred to the information in the graph as part of their written explanation. All three responses were given during the temperature segment, after the teaching episode that involved graphing.

Table 6. Student creation of graphs

	Grade 7	Grade 9	Grade 11
Rainfall	67% (14/21)	21% (6/28)	26% (5/16)
Temperature	78% (13/21)	8% (2/24)	0% (0/19)

Almost all students included aspects of both a description and a prediction in their individual responses, rather than just a description as requested in the individual task. This compulsion to predict may reflect a need to give a purpose, predicting, to justify having to give a description. The discussion includes aspects of the responses irrespective of whether the student gave a prediction as well as description, or just gave a description as requested. Noticeable in the responses was whether the explanations were based on the given data and how much of the data were used.

Although many students mentioned factors external to the data when justifying decisions, most of them also referred to the data. Some external factors mentioned were of a personal and less relevant nature, such as the occurrence of a birthday, while others were of a less personal and more relevant nature, such as prior knowledge of local weather or similar events that have been held in the past. Encouragingly 72% of responses from Grade 7 students, 96% from Grade 9 and 97% from Grade 11 made at least some reference to the supplied data indicating that most students in the higher grades appreciated the need to use the data provided as a basis for the written explanations.

When describing the weather for the month some students used just some of the data by choosing to focus on a particular part of the month while others incorporated all of the month. Those focusing on part of the month generally chose specific day(s) or a consecutive sequence (block) of days. Many students focused on such a block when discussing the rainfall. This focus appears to have been influenced by the rain/no rain (dichotomous) nature often attributed to the rainfall variable. Focus on the whole month was more typical for the responses dealing with the temperature data and justifications for decisions often dealt with quoting one or more simple statistic(s), such as the maximum, minimum or average, for both sets of data as if they were one.

The references to features of variation in the data varied considerably in length and quality and were both qualitative and quantitative in nature. The following details the results of the analysis of the responses.

4. RESULTS

This study aimed to develop a way of assessing students' descriptions of variation, by extending the Reading and Shaughnessy (2004) Description Hierarchy presented earlier in Table 3. In the first step of the analysis, attempts were made to code the responses using the four levels D1, D2, D3 and D4 of this original hierarchy. Most responses were found to fall within the D1 and D2 levels, describing variation using *Extreme Values* or *Middle Values* or both. It soon became apparent, however, that while the written discussions were clearly falling into these levels, some were expressing the features in words only while others were expressing the features numerically. Those using words only, with no numeric descriptions of features of the variation, were labeled qualitative responses while those that did include numeric features were labeled quantitative. Thus the revision and expansion of the Reading and Shaughnessy (2004) hierarchy focused on developing two distinct groupings of responses based mainly on the previous D1 and D2 levels, one grouping based on qualitative descriptions of variation and the other based on quantitative descriptions. These two types of responses are analyzed separately in sections 4.1 and 4.2 below. Further discussion of the comparison of the Reading and Shaughnessy (2004) hierarchy and the proposed groupings of coded responses can be found in the Discussion section later on.

For easy reference in the discussion, an identification tag has been assigned to reproduced responses, based on consecutive appearance. The tags begin with R, followed by a grade number (7, 9 or 11) and then a specific student number. For example, R1102 is the second response from a Grade 11 student to appear in the discussion. Any response, or part thereof, that is reproduced directly is shown in italics. No grammatical corrections have been made to the responses. When interpreting any reproduced responses it must be remembered that each student was dealing with data for a different month.

4.1. QUALITATIVE RESPONSES

The responses in this first grouping use word-based expressions, rather than numerical expressions, to describe the variation in the data. Some of these responses describe the variation by the use of general terms or phrases to describe the nature of the changes identified while others are more specific in their qualitative descriptions. Next is a discussion of the structure of responses at each of the three levels, unistructural, multistructural, and relational.

Unistructura1 Responses

The unistructural responses give one qualitative description to summarize an impression of the variation and can be grouped into two types, magnitude-related and arrangement-related. The magnitude-related terms, typical examples given in Table 7, are used in an absolute sense to give an indication of how the magnitude of the numbers is changing. Some terms suggest little change, while others suggest more change. The arrangement-related terms, typical examples given in Table 8, are used in a relative sense to give an impression of the position of the data elements relative to other data elements. Some terms suggest an inability to decide on any basis for the arrangement while others suggest a regular, describable arrangement. The use of the term *distributed*, see Table 8, is noteworthy. The students involved would most likely not have met the term 'distribution' in a formal statistical sense and these references may be in a more general sense of things being arranged.

Table 7. Magnitude-related phrases used in unistructural responses

Suggesting Little Change	Suggesting More Change
slightly on and off	*least predictable*
reasonably steady	*a bit unpredictable*
most consistent	*seem to be more mixed around*
pretty much consistent	*a bit erratic*
no sudden variations	*very unpredictable*
pretty regular	

Table 8. Arrangement-related phrases used in unistructural responses

Undecided on Arrangement	Decided on Arrangement
no pattern	*spread out*
no particular pattern	*scattered through*
no real pattern	*evenly spread*
	even balance
	evenly distributed
	distribution is limited

Multistructural Responses

The multistructural responses make use of more than one qualitative statement when describing the variation and fall into two categories, *Limiting* and *Sequential*. The first category, *Limiting*, comprises responses that deal with the data by setting a general limit on the values, often indicated by too much or too little. Such responses were more common for description of the temperature and typically summarize the data using the term 'too', such as *too cold* or *too hot*, for example R1101 and R901.

(R1101) *September wouldn't be good because its too cold in the morning…*

(R901) *It could be a good month to have it because it doesn't get too hot.*

Such responses were not as common for rainfall, but R701 and R702 are examples, with R702 qualifying the *just about nothing* description of the rain by giving the total.

(R701) *I think they should have it in february because theres not much rainfalls …*

(R702) *I think my month is the best because it rains just about nothing at all. This months complete rainfall is 8.6 mm.*

The second category, *Sequential*, comprises responses that deal with the data item by item or by grouping like data items in a qualitative way. Such responses were more common for description of the rainfall and typically summarize the data into like blocks of days. These blocks tended to be wet days (rain) and dry days (no rain), reflecting the dichotomous nature imposed on the rainfall variable by students. Response R902 has generalized this blocking while R1102 is more specific about the blocks. Another form of description of this rain/no rain dichotomy was by pairing, such as in R903. The less common sequential temperature responses block off the days in the month based generally on a higher versus lower temperature dichotomy, as typified by R904.

(R902) *There are a lot of dry days then a couple of wet days then a lot of dry days again.*

(R1102) *In the first 10 days of the month would be good as there is no rainfall here and then it continues as 4 days with rain, 5 days clear, 2 days with rain, 7 days clear, 2 days with rain.*

(R903) *Usually the rain is in pairs. After a high column little or no rain is after it.*

(R904) *In the minimum temperatures there seems to be a pattern of a few higher temperatures and then a few lower temperatures and so on.*

Relational Responses

The relational responses give a qualitative description of the variation suggesting that both limiting and sequential aspects have been considered and linked to give an overall description. An example is R905, which gives a general limit but then goes on to discuss blocks in a sequential manner. Such linked responses, however, were uncommon for those who gave qualitative descriptions.

(R905*) January '98 seems to be a pretty average month in terms of rainfall. Not too much and not too little. The rain seems to fall pretty regularly, but the amounts are not much. I think January would be a good month to hold "Youth Alive". The main pattern seems to be a short spell of dry days (3-5 days) and then 1 or 2 wet days but as the rain is pretty light and not a large amount falls, I think this month would be pretty good.*

4.2. QUANTITATIVE RESPONSES

The responses in this second grouping use numerical values, often simple statistics, to describe the variation in the data. Next is a discussion of the structure of responses at each of the three levels, unistructural, multistructural, and relational.

Unistructual Responses

The unistructural responses discuss one quantitative feature when describing the variation and fall into two distinct categories, one based on a description of the *Extreme Values* of the data and the other based on *Interior Values*. The *Extreme Values* responses describe the extreme values of the data explicitly by referring to the minimum and/or maximum or implicitly by referring to the highest and/or lowest. Responses mentioning the maximum and/or minimum explicitly are easily identifiable, so the following examples particularly demonstrate some of the more implicit references. Few of these more quantitative responses gave only one extreme for the data, thus it was rare to find the minimum without the maximum and vice versa. *Extreme Values* responses were much more common with the temperature data, typified by R906, than rainfall, typified by R907.

(R906) *August 98 was a relatively cold month, the highest temp being only 17.9 degrees Celsius and the lowest being a freezing -6.9 degrees Celsius.*

(R907) *I can see that the highest rainfall was 45.2 ml and the lowest was 0.0ml.*

Many responses, such as R906, gave the highest and lowest data values for the temperature by quoting the highest figure for the Maximum Temperature and the lowest figure for the Minimum Temperature, as if the two separate variables were being treated as one temperature variable. It is possible, though, that students only considered the top of the maximums relevant to the task at hand and the bottom of the maximums not so relevant. Similarly, the bottom of the minimum temperatures could be considered more relevant than the top of the minimums. The response R703 gave, not just the *most* but also, the second highest rainfall for the month, before reverting to describing the total rainfall for blocks of days. Other ways of expressing the minimum or maximum implicitly included *doesn't go below* (see R704), *decreasing past* and *exceeding* (see R908).

(R703) *The 4th had the most rain with 31 mils second was the 5th with 29.8 mils so…*

(R704*) March would be a pretty sweet month to have this thingy in because it doesn't go below 5 degrees and usually about 30 degrees Celsius at peak temperature.*

(R908) *February doesn't seem to have a pattern except that it seems to have a fairly warm to hot climate with temperature either exceeding 30 degrees or decreasing past 4 degrees.*

A natural progression for those giving both the maximum and minimum was to describe the range. Some responses, such as R705, actually expressed the maximum and minimum in a *from... to...* form, thus supplying a maximum and minimum and implying a range. Other responses explicitly mentioned the range, either for just one of the variables, as in R1103, or for both, as in R1104.

(R705*) In between the 3rd and 12th would be a good time to have the thing with it been warm but not to hot. The max temps where 21 degrees Celsius to 29 degrees Celsius and the min temps where 5 degrees Celsius to 15 degrees Celsius.*

(R1103) *The maximum temp in March was 28.7 degrees Celsius the minimum was 5.6 degrees Celsius. The range in Max. temp was 12.3 degrees Celsius. I think this month would be good to hold the youth fest in because it stays fairly warm throughout the month.*

(R1104) *The maximum temperature average is 16.6 degrees Celsius which is cool but not too cold temperature. I don't think this max temp would be ideal for the youth fest. The max temps range from 14 - 19.8 degrees Celsius so the max highest is what I would be wanting. The min deg C ranges a lot from 10.5 - -0.5, this is cold weather and wouldn't suit a festival…*

The *Interior Values* responses describe the interior values of the data by referring to blocks of rainfall or temperature. Those responses mentioning the blocks were generally descriptions for rainfall. Sometimes the responses referred to blocks in general, as in R1105, while others were more specific about the number of days or the exact dates when they occurred, as in R1106. Some other ways of referring to the blocks included as patches (see R1107) and periods (see R706).

(R1105) *It appears that after a larger rainfall of 36mm it rains slightly on and off for the following week before another large rainfall. It also rains a few days before the heavy rainfall sort of like a build up and dies down at the end of the 2nd heavy rainfall.*

(R1106) *There seems to be rain nearly every 5 days for 1 - 3 days either side of the 5th day. The 18th seems to be the best day because it is in the middle of 15 and 20 and it is the middle day of a six day dry spell.*

(R1107*) In June 99 the temprature is cold. For winter there is a warm patch. The temprature then drops for around four days in the middle. It increases towards the end. In the min column below zero tempratures came in patches apart from one.*

(R706) *From the 18th to the 24th was the longest period with out rain. From the 8th to the 18 was the longest period with rain almost non-stop, with 26.6 millimetres. So I think that the best time to have an outdoor event in July would be from sometime between 18th and 24th. It rained 15 days and didn't rain 16 days for the month.*

Multistructural Responses

The multistructural responses discuss more than one feature of data when describing the variation and usually combine elements identified for discussing extremes with elements identified for discussing interior values. Only sixteen responses were coded at this level, all describing the temperature data and just one from a Grade 7 student. While the quality of the description of the extreme values does not vary much in these responses, the quality of the description of the interior values does. Extreme values are usually discussed as maximum and/or minimum and in some cases an actual range is given. When discussing the interior values some responses give an overview while others discuss specific data values.

Typical of those responses giving an overview of the interior values, while also mentioning the extreme values, are R909 and R910. The first gives the statistically unsophisticated *rise and fall* overview of the interior values, along with the maximum and minimum. The second states what the temperature will get down to (i.e., the minimum) but then describes each of the two temperature data sets by giving *seems to follow a bit of a pattern* as an overview of the interior minimum temperatures and *stays pretty much constant* as an overview of the interior maximum temperatures.

(R909*) In my month I can see that the highest temp was 15 degrees Celsius. The lowest temp record was -6.7 degrees Celsius. I believe this would be a bad month to hold the festivle because it is too cold. The temp pattern seems to rise and fall throughout the month.*

(R910) *The minimum temperature seems to follow a bit of a pattern. The temp. gets down to -9 degrees Celsius. The maximum temperature stays pretty much constant, it isn't affected much by the really cold minimum temperatures.*

Only five responses could be considered to have gone into more detail about the interior values while also mentioning the extreme values. Two such responses are presented here. R1108 does this by discussing, in a statistically unsophisticated way, the patches of warmer or colder weather in more detail sequentially through the month. R1109 attempts to consider a relationship between the two variables Minimum Temperature and Maximum Temperature.

(R1108) *This month is in the middle of the summer, so most of the temperatures are in the late 20's earlie 30's. The lowest max temperature is 20.4 degrees at the start of the month. The highest max temperature is the second last day of the month, temperature is 31.3 degrees. There seems to be groups of high min temperatures of 10 degrees plus, 4 or 5 high ones and then 1 or 2 low min temperatures, Where as with the max temperatures the temperature builds up for example, 22 degrees, 26 degrees, 26 degrees, 27 degrees, 29 degrees, 30 degrees, 24 degrees and then suddenly drops 5 degrees or 6 degrees.*

(R1109) *October - 00 would be a good month to hold the Youth Fest, because the max. deg of temperature varies between 11.3 and 28.3, and the min deg of temperature lies between -3.1 and 14.3. The max. temperature is high at the beginning of the October month, it slowly rises then gradually drops mid October, at this period the min degree temperature is around it's best, again the max. temp rises and drops towards the end of the month. At the same time, the min-temp rises when the max temp is remaining constant (20/21). Therefore, if the Youth fest is to be held in October, 00, it should be on a day that is included in the constant temperature pattern.*

Relational Responses

The relational responses attempt to tie together the extreme and interior values and suggest immature notions of deviations in the data values. R911 considers the day to day deviation for one 24 hour period, while R1110 considers the day to day deviation on a couple of the days and R707 discusses what appear to be, but are not obviously, 'averaged' deviations from day to day.

(R911) *The highest temperature is 20.3 degrees Celsius and the lowest temperature is -6.3 degrees Celsius. The maximum degrees Celsius ranges 13.2 degrees Celsius. The minimum temp ranges 15.7 degrees Celsius. I don't think this month would be good for the youth event as it is to cold. The temperature jumps quite a bit in places. One day the min temp was -0.9 and the next it was 7 degrees Celsius.*

(R1110) *I don't think that Jul 98 would be a very convenient time to hold the youth festival because although the weather is reasonably steady it is often very unpredictable. For example on the 4th of July the temperatures rose almost 6 degrees over night yet only a few days later on the 7th it dropped another 5.8 degrees over night. There is no real pattern here like I said the weather seems to be very unpredictable. The information shows that it is a month with moderate cool weather. The average range of the temperatures between the min and max degrees for a certain day is on the 20th when the range is 15.9 degrees Celsius and on the 28th when the range is a mere 1.1 degrees Celsius. The minimum degrees for any day is on the 3rd with -5.9 degrees Celsius.*

(R707) *I think that my month would be unsuitable to hold "Youth Fest" because the high temperatures are on average around 2 or 3 degrees different everyday. The same happened with the Minimum temperatures. They were also very cold with -3, -5 degree.*

5. DISCUSSION

This study focused on refining the Reading and Shaughnessy (2004) hierarchy based on responses from weather-related inference tasks. SOLO was used as a framework to support the refined hierarchy and two cycles of levels of cognitive growth were identified. While the Reading and Shaughnessy hierarchy was useful as a starting point, it was not detailed enough to accommodate the range of responses that were given by students. The students in the present study were engaged in a different task, involving inference from data with real variation rather than a sampling task in a probability context. Also, there was a richness in the contexts from which responses were collected, both before and after the group work and the teaching episodes. This section first addresses the three research questions proposed earlier, in the light of the results. Following that, the newly developed hierarchy is compared to hierarchies developed by other researchers and finally some limitations of the study are considered.

5.1. DESCRIPTION OF VARIATION HIERARCHY

The three research questions are now addressed. First, the specific refinements used to produce the refined hierarchy are outlined. Next, it will be argued that SOLO provides a suitable explanation for this hierarchy. Finally, the notion of two cycles of levels identified within one SOLO mode, as has been found by other researchers, is confirmed for these responses.

Refinement of the Reading and Shaughnessy (2004) hierarchy

The first research question asked whether the hierarchy developed for analyzing students' descriptions of variation in a sampling situation (Reading & Shaughnessy, 2004) was also applicable for coding responses with data descriptions given when making inferences from weather-related data, in which there is natural variation. Although the Reading and Shaughnessy (2004) hierarchy proved useful as a foundation for the coding, a more detailed structure was needed to account for the array of responses given by students. Table 9 links each of the two groupings of the hierarchy proposed by the analysis in this study to the original Reading and Shaughnessy levels on which they were based. The three levels of the first grouping, based on qualitative feature were considered by the researchers to be less statistically sophisticated versions of the D1 - *Extreme or Middle Values*, and D2 - *Extreme and Middle Values* of the Reading and Shaughnessy hierarchy. The responses described in this qualitative grouping help to give an insight into early considerations of features of variation that later develop to become the more easily recognizable numeric descriptions of variation. The three levels in the second grouping, based on quantitative features of variation, more closely align with the D1 and D2 levels of the Reading and Shaughnessy hierarchy.

Table 9. Refined hierarchy linked to the Reading and Shaughnessy (2004) hierarchy

Refinement for Description of Variation Hierarchy	**Link to Reading &Shaughnessy (2004) Hierarchy**
Qualitative Responses	Expressed in words only, with no descriptions of numeric features of variation
unistructural - one qualitative feature of variation	Like the D1 responses, *Extreme Values or Middle Values*
multistructural - more than one qualitative feature of variation	Like the D2 responses, both *Extreme Values and Middle Values*
relational - link qualitative features of variation	Links the *Extreme Values and Middle Values* features of variation
Quantitative Responses	Expressed with numeric features of variation
unistructural - one quantitative feature of variation	Equivalent to the D1 responses, *Extreme Values or Middle Values*
multistructural - more than one quantitative feature of variation	Equivalent to the D2 responses, both *Extreme Values and Middle Values*
relational - link quantitative features of variation	Links the *Extreme Values and Middle Values* features of variation, may suggest notion of deviation and hence be heading towards a D3 response.

The responses in the qualitative grouping are considered to be less statistically sophisticated than the responses in the quantitative grouping. Although students' qualitative descriptions show that they have been able to notice and acknowledge variation, they have not been able to apply a measure to their description. It should be noted that the use of the term *Middle Values* in the Reading and Shaughnessy hierarchy, meant to refer to the values not occurring at the extremes, was being misinterpreted by users of the hierarchy as referring to measures of central tendency. To avoid further confusion the terminology was changed from *Middle Values*, as used by Reading and Shaughnessy, to *Interior Values* in the refined hierarchy. The term *Middle* has still been used in Table 9, consistent with the Reading and Shaughnessy hierarchy but the term *Interior* is used in later descriptions of the refined hierarchy. The expression 'Like' is used in the explanations for the qualitative responses because these responses were describing the same sort of features as described in the Reading and Shaughnessy D1 and D2 levels but not in the same way, i.e., they did not contain the numerically described features of variation that D1 and D2 contained. The expression 'Equivalent' is used for the quantitative responses because these responses included features of variation described in the same numeric fashion as those in the D1 and D2 levels.

No responses were found in the present study that specifically discussed deviations from an anchor, central or non-central, and hence could be considered as equivalent to those identified by Reading and Shaughnessy as D3 - *Discuss Deviations from an Anchor* or D4 - *Discuss Deviations from a Central Anchor*. However, there were two responses in the present study, R1110 and R707 at the relational level, which may be considered transitional to being coded as D3 because of the attempt to describe the deviations.

Thus, in response to the first research question, it was possible to refine the Reading and Shaughnessy (2004) hierarchy by identifying responses equivalent to those in the D1 and D2 levels and by also identifying responses that were structurally similar to D1 and D2 responses but expressed in the less statistically mature qualitative form. Additional research is needed with more statistically sophisticated responses that those given in the present study to be able to refine the D3 and D4 levels, where deviations become the focus of the discussion.

Other researchers, too, have reported finding similar approaches to dealing with variation as those identified here. For example, delMas and Liu (2003) investigated students' formation of ideas when they were first learning about factors that affect standard deviation. Of interest are strategies they identified students using when attempting to move bars in a graph to produce maximum or minimum standard deviation. One strategy, 'equally spread out', focusing on equal separation of the bars in the graph, is similar to the descriptions in the *Interior Values* focused qualitative unistructural responses identified in the present study, while the 'far-away' strategy, focusing on getting the bars as far away from each other as possible, is similar to the variation descriptions focusing on *Extreme Values*.

SOLO as a theoretical framework for the refined hierarchy

Having established that this hierarchy is suitable, the second research question asked whether SOLO could offer a broad framework for explaining the hierarchy. Discussion now focuses on explaining how the taxonomy was used to explain this cognitive growth as a distinct cycle of unistructural (U), multistructural (M) and relational (R) levels (see section 2.1 and Pegg, 2003, p. 243). Table 10 summarizes the application of the SOLO Framework to the six levels. In the qualitative responses, the first three levels now labeled as the first cycle, identification of the element of interest as 'a feature of the variation of the data described qualitatively' allows the three levels within that category, to be explained as unistructural, multistructural and relational. The unistructural (U1) responses contain one such element, the multistructural (M1) responses contain more than one such element and the relational (R1) responses link these elements. This cycle has some qualitative descriptions that are more *Sequential* in nature while others are more *Limiting*. One key to better defining what is happening in this first cycle might be to look to other research that identifies intuitive notions, such as that by Makar and Confrey (2003) who found that pre-service teachers were using 'informal' terms when comparing dotplots but in the process were discussing non-simplistic concepts. Responses that suggested consideration of clustering, as opposed to modal clumping, and the terms used by these prospective teachers may help to unravel the often-unclear terminology used by younger students and add to the definition of levels in this cycle.

Table 10. Refined description of variation hierarchy

First Cycle **Qualitative Responses**	**element - qualitative feature of variation of data**
U1 - unistructural - one qualitative feature of variation	magnitude related - in an absolute sense to give indication of size of change, e.g., *pretty much consistent* or arrangement related - in a relative sense to give position, e.g., *spread out pretty evenly*
M1 - multistructural - more than one qualitative feature of variation	limiting related - set limits on the data values, e.g., *doesn't get too hot* and/or sequential related - deal with data item by item, e.g., *lots of dry days then a couple of wet days then a lot of dry days again*
R1 - relational - link qualitative features of variation	link the general limit with the discussion of blocks sequentially, e.g., *seems to fall pretty regularly but the amounts are not too much .. main pattern seems to be a short spell of dry days (3-5days) and then 1 or 2 wet days but rain is pretty light and not a large amount falls…*
Second Cycle **Quantitative Responses**	**element - quantitative feature of variation of data**
U2 - unistructural - one quantitative feature of variation	based on extreme values - discuss maximum, minimum, range or interior values - refer to blocks or patches of days
M2 - multistructural - more than one quantitative feature of variation	based on extreme values and/or interior values, e.g., refer to range but also to the rise and fall of temperatures throughout the month
R2 - relational - link quantitative features of variation	linking of extreme values and interior values may suggest immature notions of deviations, e.g., discussions including day-to-day deviations or 'averaged' deviations from day-to-day

In the quantitative responses, the last three levels now labeled as the second cycle, identification of the element of interest as 'a feature of the variation of the data described quantitatively' explains the three levels, unistructural (U2), multistructural (M2) and relational (R2) of this cycle. This cycle of levels includes responses that clearly deal with *Extreme Values* while others deal with *Interior Values*. The importance of investigating these notions of evaluating dispersion were also borne out by the research of Lann and Falk (2003), who evaluated strategies used by statistically naive tertiary students, as they compared sequences of data for greater variability. Some criteria for making decisions were common such as the Range and Interquartile Range, while others were more intuitive and less easy to unravel. Responses using the Range would identify as having an *Extreme Values* focus, in the refined hierarchy, while those using the Interquartile Range would be classified as *Interior Values*. Lann and Falk (2003) also attempted to analyze the justifications given for selected responses but found that the analysis of these explanations was not such an easy task. Results from their, yet to be investigated, considerable number of 'no definite diagnosis' responses, may also add to the story in the second cycle but is more likely to assist in unraveling the mystery of what students really mean when they give responses such as those in the first cycle.

Two Cycles of SOLO Levels identified

Having established that SOLO proved useful as a suitable framework, the third research question asked whether cycles of levels can be identified within the SOLO modes? Two distinct cycles of the unistructural-multistructural-relational levels have been identified. Both these cycles are part of the concrete symbolic mode (Pegg, 2003, p.242) where a person thinks through the use of the symbol systems, both language and numeric, as used by the literate. Pegg (2003, p. 245) provides a useful diagrammatic representation of the link between coexisting cycles of levels within the concrete symbolic mode. As pointed out earlier the existence of more than one cycle of levels of cognitive development within one SOLO mode of cognition has now been observed by other researchers and so it is not unexpected that two cycles of levels would be observed in this study. Of particular interest with the first cycle in the present study is the strong emphasis on visual elements in the descriptions of variation. This would be expected because, as Pegg (2003, p. 244) points out, this first cycle in the concrete symbolic mode provides an interface to the less cognitively developed ikonic mode of operation, where actions are internalized as images. In fact, some responses demonstrated that students revert to the ikonic mode, based on personal experience, such as their own knowledge of festivals and the town's weather, when trying to justify their evaluation of the suitability of the month for the event.

The nature of the responses as described in the refined Description of Variation Hierarchy, within each of the two cycles, demonstrates a developmental cognitive progression from the first to second cycle. Those responses at the first multistructural level specifically coded as *Limiting* appear to be precursors to the *Extreme Value*s responses at the second unistructural level, while those coded as *Sequential* appear to be precursors to the *Interior Values*. As the terminology used by students progresses through the levels of the two cycles it appears as if the students are adjusting the focusing lens on a microscope. The higher the level of response achieved the finer the detail provided about the variation that exists. Even finer detail is expected to unfold in future research during analysis of responses to other tasks and from more advanced students.

5.2. COMPARISON TO OTHER DEVELOPMENTAL HIERARCHIES

The Description of Variation Hierarchy as refined by this study, see Table 10, provides a greater depth of explanation of the focus of student responses on variation than the previously developed hierarchies, by Mooney (2002), see Table 1, and Watson et al. (2003), see Table 2. This has been achieved by identifying cycles of levels, unistructural, multistructural and relational, within the SOLO mode of cognitive growth for the less statistically sophisticated categories in both of these hierarchies. Is the greater depth of explanation within the refined hierarchy consistent with references to spread in the Mooney (2000) hierarchy? Though variation is only acknowledged through references to spread in the *Organising and Reducing Data* process of Mooney's (2002) framework, similarities in descriptors can be found with the refined hierarchy proposed in Table 10. Mooney's *Transitional* responses, with their 'invented' measures are similar to the proposed first cycle of qualitative responses. However, the proposed second cycle of quantitative measures has not distinguished between invented and valid measures as Mooney's *Qualitative* and *Analytical* levels have done. There were not sufficient responses at, or above, the relational level of the quantitative responses in the present study to develop the hierarchy further at these higher levels of cognition.

Is the greater depth of explanation in the refined hierarchy consistent with the hierarchy developed by Watson et al. (2003)? The ability to describe variation is essential to demonstrating the achievement of levels of understanding developed by Watson et al (2003). The qualitative responses identified in the first cycle of the refined hierarchy in Table 10 are typical of descriptions given in responses at Level 1 - *Prerequisites for Variation* of the Watson et al. hierarchy. The quantitative responses identified as second cycle of the refined hierarchy are typical of descriptions given in responses at Level 2 - *Partial Recognition of Variation*. As previously mentioned responses at, or above, the relational level of the quantitative responses were lacking and if such responses are collected in the future they may provide cycles of levels of description of variation equating to the upper two levels of the Watson et al. Hierarchy, Level 3 - *Applications of Variation* and Level 4 - *Critical Aspects of Variation*. Thus the refined hierarchy proposed by the present study has provided a greater depth of explanation to the lower cognitive levels of both the Mooney (2002) and Watson et al. (2003) hierarchies.

5.3. INTERPRETATIONS AND LIMITATIONS

Any consideration of the findings reported above needs to take into account students' interpretations of the task and two noticeable limitations of the study relating to student motivation and the profile of the data supplied for the task. Student interpretation issues focus around the initial intent of the activity, the different approaches to the two data variables and lack of recognition of the benefit of using visual representations. The way most of the students interpreted the task, though not contrary to, was not exactly what was initially intended. The natural urge to predict the most suitable month, even before being asked to do so, suggests that for students to give more meaningful descriptions of data they need a context and a sense of purpose. In this case, the students were given a context, using rainfall and temperature data from their own town for the preceding three years, and a purpose, to decide on the suitability of a particular month for the scheduling of a Youth Festival.

The students' familiarity with weather and with their expectations of the need for suitable weather for the festival may have contributed to the differing approaches that students took to describing the variation in the data for rainfall and for temperature. Consideration of the data for just part of the month was more common when describing the rainfall data, where blocking of 'rain' and 'no rain' days was often the focus. For temperature, use of the data for all of the month was more common and extremes of temperature became the focus of the better responses. Rainfall almost took on a dichotomous nature in that interest centred on whether it 'rained or not', while temperature maintained a more continuous nature with the number of degrees being considered to be of enough importance to be discussed.

The use of a realistic context, though considered more meaningful, appears to have precluded students from recognizing an opportunity to make use of skills newly acquired in the classroom. Few students beyond Grade 7 drew a graph to help describe the data and only three students referred specifically to their graphs in their explanations. Even the inclusion of a teaching episode, to introduce a new graphing technique to the students, did not result in any noticeable increase in the use of graphs to aid the inferences. It is possible that if students had been encouraged, or actually required, to draw a graph of the data then visual cues may have assisted them to give more detailed descriptions of the variation.

Student motivation was clearly evident early on in the task, but waned as the activity progressed. Well intentioned attempts to provide a realistic context for the inference task were obviously successful to the point of creating another problem. Some students thought the Youth Alive festival was really going to take place and Grade 7 were particularly disappointed when they found out that this was not so.

Apart from the differing student interpretations of the nature of the two variables, rainfall and temperature, the profile of the data also differed in the amount of information supplied. One set of data was given per month for rainfall (daily millimetres) and two sets of data per month (daily maximum degrees centigrade and daily minimum degrees centigrade) for temperature. The two sets of data for temperature proved more of a complication for students than had been anticipated. In many cases students dealt with this issue by using only one set of data or the other, or by combining all the data into one set with maximum temperatures and minimum temperatures together.

These various interpretation issues and limitations were not considered to detract, however, from the wealth of information contained in the responses. This was especially so given that the coding of the responses was not to be used as a quantification of the best of students' capabilities but more as an indication of what descriptions of variation are used by students as they respond to the particular weather-related inference task.

6. IMPLICATIONS

6.1. IMPLICATIONS FOR RESEARCH

Several implications for research arise from this study. First, the refining of the hierarchy has demonstrated that levels devised for description of variation in sampling task responses have proven useful as a starting point for analyzing responses in an inference task and that SOLO can provide a suitable framework for such a hierarchy. Descriptions of lower level responses in the original hierarchy have been expanded and levels have been created for responses that are less statistically sophisticated than those in the original hierarchy. The strength of this refinement of a previously developed hierarchy now needs to be tested by applying the developed cycles of levels to the coding of responses posed in statistical tasks based in other contexts. Another implication of this study is that more statistically sophisticated responses need to be analyzed to identify the structure of possible cycles of levels that may exist above the two cycles proposed. It is expected that research with more advanced students will reveal some detail of more sophisticated development. The delMas and Liu (2003) research is a clear indication of the reasonableness of this expectation. A strategy they found being used by college students to describe variation, 'far-away mean', focused on trying to get the bars of a computer display as far away as possible from the mean in order to affect the standard deviation. This is similar in approach to the descriptions given by students at the Reading and Shaughnessy (2004) D4 – *Discuss Deviation from a Central Anchor* level and indicates that refinement of the D3 and D4 levels would be warranted.

A further implication is that when designing tasks researchers need to be aware of the influence of the nature of the variable used in the task on the style of response and to try encouraging the use of graphical representation to improve the quality of descriptions of variation. Related to this is the implication that care should be taken to avoid unnecessary complication in tasks given. In this case, future use of the weather activity should only include one set of data for a particular variable, e.g., the more relevant Maximum Temperature for the temperature segment of the activity. This would remove the complication, unnecessary to this particular investigation, of having to deal with two sets of data for the one variable. A final implication is that consideration in future research should also be given to the role played by measures of central tendency, such as the mean, when describing variation.

6.2. IMPLICATIONS FOR TEACHING AND ASSESSMENT

From a teaching perspective, student responses to the weather activity demonstrate that when considering data 'in context' students may rely too much on their experience of the context itself and not enough on information provided by the data. This then influences the way students describe the variation of the data, and ultimately any predictions made. It is also evident that the nature of the data in the variable influences the way that students react to data. Students focused on varying amounts of the information depending on and treated them differently depending on whether the temperature or rainfall variable was being discussed. In addition, teachers should consider encouraging more use of graphical representation when students are engaged in activities that involve description of variation. The terminology used by students is important and there is a need to encourage students to work from their own terminology and descriptions to what is required in more statistically sophisticated discussions. Such a necessity has also been flagged by Makar and Confrey (2003). Finally, the description of responses at the various levels can be used to help teachers make sense of the unsophisticated language and reasoning of students during classroom activities.

From an assessment perspective, the hierarchy developed in this study could provide a rubric to assess the level of cognitive growth at which students are operating in terms of their description of variation, a very basic statistical concept. Such descriptions are essential if students are to be able to indicate their appreciation of existing variation and communicate such information in a statistically sophisticated manner to a wider audience. As such a hierarchy is further developed teachers should be encouraged to use it to code responses to a variety of statistical tasks, so that they will be better informed as to how students are describing the variation as part of their reasoning about variability and patterns in data.

ACKNOWLEDGEMENTS

This paper reports on research that formed the basis of the presentation *Student Perceptions of Variation in a Real World Context* at the Third International Statistical Reasoning, Thinking and Literacy Research Forum in Lincoln, Nebraska from July 23 to 28 in 2003. The author acknowledges the helpful comments made by reviewers and editors who have contributed to the revision of this paper.

REFERENCES

Bakker, A. (2003). Reasoning about shape as a pattern in variability. In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3)*. [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Ben-Zvi, D. (2003). The emergence of reasoning about variability in comparing distributions: A case study of two seventh grade students. In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3)*. [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behaviour. In H. Rowe (Ed.), *Intelligence, Reconceptualization and Measurement* (pp. 57–76). New Jersey: Laurence Erlbaum Assoc.

DelMas, R.C., & Liu, Y. (2003). Exploring students' understanding of statistical variation. In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3)*. [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, *32*, 101–125.

Jones, G. A., Mooney, E. S., Langrall, C. W., & Thornton, C. A. (2002). Students' individual and collective statistical thinking. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*, *Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. A. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning, 2*, 269–307.

Langrall, C. W. & Mooney, E. S. (2002). The development of a framework characterizing middle school students' statistical thinking. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*, *Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.

Lann, A., & Falk, R. (2003). What are the clues for intuitive assessment of variability? In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3)*. [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Makar, K., & Confrey, J. (2003). Chunks, clumps, and spread out. In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3).* [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*(1), 23–63.

Pegg, J. (2003). Assessment in mathematics: A developmental approach. In J. M. Royer (Ed.), *Advances in Mathematical Cognition* (pp. 227–259). Greenwich, CT: Information Age Publishing.

Reading, C. (1998). Reactions to data: Students' understanding of data interpretation. In L. Pereira-Mendoza, L. Kea, T. Kee & W-K Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (Vol. 3, pp. 1427–1433). Voorburg, The Netherlands: International Statistical Institute.

Reading, C., & Lawrie, L. (2004). Using SOLO to analyse group responses. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 193–200). Bergen, Norway: Bergen University College.

Reading, C., & Pegg, J. (1996). Exploring understanding of data reduction. In A. Gutierrez (Ed.), *Proceedings of the 20th International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 187–195). Valencia, Spain: University of Valencia.

Reading, C., & Shaughnessy, J. M. (2000). Student perceptions of variation in a sampling situation. In T. Nakahar & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 89–96). Hiroshima: 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, The Netherlands: Kluwer Academic Publishers.

Shaughnessy, M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*, *Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.

Shaughnessy, M., Watson, J., Moritz, J., & Reading, C. (1999). School mathematics students' acknowledgement of statistical variation. In C.Maher (Chair), *There's more to life than centers*, Procession Research Symposium, 77th Annual National Council of Teachers of Mathematics Conference, San Franciso, CA.

Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: an exploratory study. *Mathematics Education Research Journal*, *12*(2), 147–169.

Watson, J. M., & Kelly, B.A. (2003). Developing intuitions about variation: The weather. In C. Lee (Ed.) *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3)*. [CDROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*(1), 1–29.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–262.

CHRIS READING
SiMERR National Centre
Education Building
University of New England
NSW 2351
Australia

APPENDIX – SAMPLE WEATHER ACTIVITY DATA HANDOUTS

Jan-98	Jul-99	Jul-98	Jul-98	Oct-00	Oct-00
Millimetres	Millimetres	Max deg C	Min deg C	Max deg C	Min deg C
0.0	21.8	8.8	3.5	22.5	7.9
0.0	0.8	12.9	0.4	19.6	-1.9
0.0	0.0	14.1	-5.9	22.0	-2.8
4.8	0.0	12.6	3.0	23.6	-3.1
2.2	0.0	18.5	4.0	24.4	3.1
9.2	0.0	18.3	2.8	25.4	3.0
0.0	0.0	15.8	10.0	26.5	3.0
0.0	0.2	10.8	8.1	27.5	4.6
0.0	0.4	8.2	4.9	28.3	9.4
11.2	6.0	8.5	-1.2	21.7	14.3
0.0	0.8	12.9	-2.8	19.4	2.3
2.2	0.0	14.0	0.2	20.1	8.0
0.0	0.8	12.0	5.1	20.0	4.2
0.0	1.0	10.0	3.5	11.3	10.1
2.4	16.2	9.7	-1.4	17.5	2.4
0.0	0.8	12.5	-2.1	19.1	1.5
0.0	0.4	16.1	-0.2	19.2	4.6
0.0	0.0	12.6	7.1	18.4	9.4
0.0	0.0	16.7	6.4	15.9	10.7
0.0	0.0	19.0	3.1	20.6	6.9
0.6	0.0	12.4	9.9	21.9	7.7
4.8	0.0	11.4	4.1	22.3	10.8
0.0	0.0	11.6	-4.2	22.0	10.4
0.0	0.0	13.2	-0.7	21.5	11.4
0.0	0.6	12.0	5.7	20.4	11.0
8.8	0.2	13.0	6.8	21.1	4.1
0.0	1.2	16.7	9.0	20.7	4.1
0.0	0.8	11.1	10.0	19.6	1.9
0.0	0.0	6.6	0.8	21.1	5.7
0.0	0.0	5.4	0.6	15.5	10.1
0.0	0.0	4.9	0.4	15.2	9.9
Rainfall Data (January 1998) used to produce response R905	Rainfall Data (July 1999) used to produce response R706	Temperature Data (July 1998) used to produce response R1110		Temperature Data (October 2000) used to produce response R1109	