AN EMERGING HIERARCHY OF REASONING ABOUT DISTRIBUTION: FROM A VARIATION PERSPECTIVE

CHRIS READING
The National Centre of Science, Information and Communication Technology, and Mathematics Education for Rural and Regional Australia, University of New England
creading@une.edu.au

JACKIE REID
School of Mathematics, Statistics and Computer Science, University of New England
jreid@turing.une.edu.au

ABSTRACT

Recent research into students' reasoning about variation refers specifically to notions of distribution that emerge. This paper reports on research where written responses, from tertiary introductory statistics students, were coded according to the level of consideration of variation. A hierarchy of reasoning about distribution is proposed, based on the notions of distribution that were evident in these responses. The hierarchy reflects students' progression from describing key elements of distribution to linking them for comparison and inference. The proposed hierarchy provides researchers with an emerging framework of students' reasoning about distribution. The research also highlights that educators need to be aware that, without a well developed consideration of variation, students' ability to reason about distribution will be hampered.

Keywords: *Statistics education research; Reasoning about variation; Reasoning about distribution; Tertiary; Hierarchy*

1. OVERVIEW

As one of the fundamental forms of statistical thinking (Wild & Pfannkuch, 1999), reasoning about variation impacts all aspects of statistics including reasoning about distribution. Recent research into statistical reasoning (Bakker, 2004; Bakker & Gravemeijer, 2004; Ben-Zvi, 2004; Chance, delMas & Garfield, 2004; Reading & Shaughnessy, 2004) highlights the importance of both variation and distribution in the study of statistics. The Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4), held in 2005, focused on reasoning about distribution. Many questions were provided as stimulus for participants in SRTL-4, and a subset of these was relevant to the work reported in this paper: What is the nature of the connection between students' reasoning about variation and students' reasoning about distribution? How can students' explorations of variation help to unravel the mystery of distribution? How can cognitive growth in reasoning about distribution be described?

First, consider what is meant by the terms variation and distribution. Variation, in its broadest sense, will be construed as the description or measurement of the observable characteristic variability (Reading & Shaughnessy, 2004, pp. 201-202). Four components of consideration of variation were developed by Wild and Pfannkuch (1999) after

Statistics Education Research Journal, 5(2), 46-68, http://www.stat.auckland.ac.nz/serj

interviewing statisticians and students: noticing and acknowledging, measuring and modelling, explaining and dealing with, and investigative strategies. In Moore and McCabe's (2003) well-known tertiary introductory statistics textbook, the distribution of a variable is defined as "the values that it takes and how often it takes those values" (p. 5); though later the definition for probability distributions is expanded by using proportions rather than frequencies. Basic features expected in descriptions of distributions (p. 12) are the overall pattern (i.e., shape, centre and spread) and deviations from the pattern (e.g., outliers). When Bakker and Gravemeijer (2004) investigated the concept of distribution, they identified centre, spread, density and skewness, as key elements. As density and skewness provide detail about shape, the authors of the current paper propose a framework for distribution with five key elements: centre, spread, density, skewness and outliers.

Next, consider that among the latest trends in statistical reasoning, thinking and literacy research, the development of hierarchies to describe cognitive growth has become a desirable research objective. The comprehensive review of models of development in Jones, Langrall, Mooney and Thornton (2004) included a summary of the models of cognitive development that relate to specific statistical concepts. Amongst these was a model for "sampling and sampling distributions," but not for distribution. Since then, Makar and Confrey (2005b) proposed a five level hierarchy of statistical inference that referred to distribution in its upper levels. However, lack of a hierarchy describing the cognitive development of distribution as a concept provided the impetus for the goal of this study – to develop and describe a hierarchy of reasoning about distribution. Of the various theories that may be used to explain cognitive growth, one in particular, the Structure of Observed Learning Outcomes (SOLO) Taxonomy (Biggs & Collis, 1982), has been identified as a powerful tool in the assessment of mathematical reasoning (Pegg, 2003). More recently, statistics education researchers have used SOLO to develop hierarchies of cognitive development (e.g., Watson, Kelly, Callingham & Shaughnessy, 2003; Pfannkuch, 2005), leading to its selection as a suitable framework for the hierarchy to be proposed in this study.

2. LITERATURE REVIEW

While there is general agreement that both variation and distribution are fundamental concepts in statistics, debate continues over which statistical concept provides a fundamental basis for the development of the other. The following review draws together research about the close connection between students' reasoning about variation and reasoning about distribution, before expanding on the SOLO Taxonomy and its use to explain statistical reasoning.

2.1. CONNECTING REASONING ABOUT VARIATION AND DISTRIBUTION

What does reasoning about variation contribute to reasoning about distribution? Reading and Shaughnessy's (2004) overview of research into reasoning about variation helps to unfold the complexity of variation but other research refers specifically to links between variation and developing notions of distribution: the context of variability is important for shedding light on reasoning about distribution (Hammerman & Rubin, 2004); comparison of data in distributions is important motivation for students to reason about variation (Ben-Zvi, 2004); and the underlying concept of distribution is critical for understanding variation (Makar & Confrey, 2003; delMas & Liu, 2003). Importantly, Bakker (2004) considered that both variability and shape are concepts that should be developed in parallel. Bakker and Gravemeijer (2004) acknowledged the vital relationship between variation and distribution when they concluded “without variation, there is no distribution” (p. 149). Other researchers (Bakker, 2004; Ben-Zvi, 2004; Makar & Confrey, 2003) have also closely linked reasoning about variation and distribution, with Bakker (2004, p. 81) calling for more research to clarify how students can develop their informal notions of centre, clumps, spread, and shapes, into more conventional measures of distribution.

The suggestion that reasoning about variation may lead naturally into reasoning about distribution becomes apparent when research findings are considered in light of the proposed five key elements of distribution: centre, spread, density, skewness and outliers. Seven developmental stages were identified when Ben-Zvi (2004) traced the reasoning about variation of two secondary students in an activity requiring them to compare two distributions. The final three stages dealt with use of centre and spread, informal variability modelling through handling outliers, and noticing and distinguishing variability within and between distributions. Makar and Confrey (2003) argued that learning environments should be structured to help develop this link between variation and distribution by pushing students to find a need for variation in their inferential tasks and assisting them to discuss variation in such a way that develops a discussion of distribution. Reading and Shaughnessy (2004, p. 223) developed a hierarchy about describing variation that included the notions of moving from general descriptions of extreme and middle values to deviations from an anchor. Such cognitive development could help students link the key distributional elements of centre and spread.

Distributional reasoning is particularly difficult for students when dealing with sampling distributions. Chance, delMas and Garfield (2004, p. 312) noted that students were not able to reason about sampling distributions until they had a sound understanding of both variability and distribution. In the light of their proposed reasoning framework, they observed that at the *Verbal Reasoning* level a “student can select a correct definition but does not understand how the key concepts such as variability and shape are integrated” (p. 303). Lack of language, beyond the level of statistical summaries, has been identified as one of the difficulties in understanding distributions (Biehler, 1997), with students finding it difficult to take distribution concepts emphasized in probability theory and apply them in data analysis situations. An improved understanding of variation may help students to better reason about distribution by providing them with a vocabulary for describing distributions.

Learning more about the link between variation and distribution is crucial if Makar and Confrey (2005a, p. 28) are correct in claiming that distribution gives “a visual representation of the data’s variation.” The study being reported in this paper aims to develop a hierarchy of reasoning about distribution, through a re-analysis of students’ responses to various tasks. The original analysis of the responses focused on reasoning about variation. The analyses and ideas presented in this paper assist in understanding what aspects of that reasoning may be helpful and provide a foundation for reasoning about distribution.

2.2. THE SOLO TAXONOMY

The cognitive developmental SOLO Taxonomy model consists of five modes of functioning, with levels of achievement identifiable within each of these modes (Biggs & Collis, 1991). Although these modes are similar to Piagetian stages, an important difference is that with SOLO earlier modes are not replaced by subsequent modes and, in fact, often support growth in later modes. For a description of these modes see Pegg

(2003, pp. 242-243). A series of levels have been identified within each of these modes. The relevant mode for this study, the concrete-symbolic mode, is the mode which focuses on thinking through the use of a symbol system. Four levels within this mode are: *prestructural* (P) with no focus on relevant aspects; *unistructural* (U) focusing on one aspect; *multistructural* (M) focusing on several unrelated aspects; and *relational* (R) focusing on several aspects in which inter-relationships are identified. These four levels form a cycle of growth that occurs in each mode and recurs in some modes, with each cycle being identified by the nature of the aspects on which it is based. When there are recurring cycles the relational level of one cycle equates to the prestructural level of the next cycle.

The application of this model of cognitive growth has varied among researchers. Some acknowledge that SOLO has been used to inform the development of their hierarchy, but do not explain how or do not explicitly use the SOLO terminology to describe their levels. For example, Watson et al. (2003) have four levels of understanding of variation entitled: Prerequisites for Variation, Partial Recognition of Variation, Applications of Variation, and Critical Aspects of Variation. Each level is articulated in detail, including the fourth level "where consolidation of concepts occurs" (Watson et al., 2003, pp. 11-13). In both the methodology and discussion, Watson et al. stated that SOLO was the basis for the categorical coding but the actual levels described have not been linked specifically to the SOLO levels (P, U, M & R). Others use the SOLO taxonomy to inform hierarchy development and explain how it relates to the levels they describe, but do not explicitly name their levels using SOLO terminology. For example, initially Mooney (2002), and then later Jones et al. (2004), described levels for analyzing and interpreting data; Idiosyncratic, Transitional, Quantitative and Analytical, and then explained each of these levels in terms of specific SOLO levels. Finally, there are those who use SOLO as the framework to underpin their hierarchy and explicitly describe the levels of the hierarchy in terms of the SOLO level descriptors. For example, Watson and Moritz (1999) for comparing two datasets, Watson and Kelly (2003) for understanding of statistical variation, Reading (2004) for describing variation, and Pfannkuch (2005) for the nature of the various strands of the statistical process, developed levels clearly articulating the parallel with SOLO levels (especially the U, M and R levels).

The existence of more than one cycle of levels within a mode (Pegg, 2003, p. 245) is already being acknowledged by statistics education researchers. Jones et al. (2004) explained the differing coding levels of statistical reasoning at the primary and secondary level as reflecting two different cycles of SOLO levels. Watson, Collis, Callingham and Moritz (1995) described two cycles of drawing inferences from data. The first based on developing an aggregated view of data, and the second based on sorting data and hypothesizing associations. Watson and Moritz (1999) identified the use of proportional reasoning in responses as indicative of the move from the first to the second cycle, in the comparison of two datasets. Reading (2004) described a cycle based on responses of a qualitative nature followed by a cycle of responses of a quantitative nature, in the description of variation.

3. APPROACH

The continuing success of the use of SOLO as a framework for hierarchy development led to its use in developing the hierarchy proposed in this paper. However, rather than the standard use of SOLO as a framework for directly coding students' raw responses, this study used SOLO for analysing responses that had already been coded (grouped) according to the level of another variable, consideration of variation. First, the

Reading and Reid (2005a) Hierarchy of Consideration of Variation (Figure 1) was used to code students' responses to assess the level of consideration of variation. The coded responses for each level of consideration of variation were then re-analyzed to determine any reasoning about distribution. This coding and re-analysis formed the first phase of the study. The SOLO framework was then applied to the results of the re-analysis to inform the proposition of a hierarchy of reasoning about *distribution*. This development of the hierarchy was the second phase of the study.

No consideration of variation	
MP1&4:	discusses the means only as evidence of the inference, with no mention of variation
MP2:	does not mention the relevant factors to explain variation of trial outcomes
MP3:	does not mention variation in relation to the distribution
Weak consideration of variation	
MP1&4:	discusses the amount of variation but does not explain how this justifies the inference
MP2:	incorrectly applies relevant factors to explain variation of trial outcomes
MP3:	some description of variation that implies how variation influences distribution
Developing consideration of variation	
MP1&4:	discusses the amount of variation and explains how this justifies the inference made
MP2:	interprets some factors correctly to better explain variation of trial outcomes
MP3:	indicates appreciation of variation as representing distribution of values
Strong consideration of variation	
MP1&4:	indicates an appreciation of the link between variation and hypothesis testing
MP2:	interprets all factors correctly to give good explanation of variation of trial outcomes
MP3:	recognizes effect of variation on the distribution and relevant factors

Figure 1. Hierarchy of Consideration of Variation
(adapted from Reading & Reid, 2005a)

The remainder of this paper is organized according to these two main phases. Sections 4-6 describe the first phase, in which students' responses were re-analyzed for evidence of reasoning about distribution, after having been coded for consideration of variation using an existing hierarchy (Figure 1). Section 7 describes the second phase of the study, where a new hierarchy of reasoning about distribution is proposed, using SOLO as the developmental framework. The proposed hierarchy in section 7 is based on the analysis of the data in the first phase (i.e., on the work described in section 4-6), as well as interpretation of ideas by various researchers both in published papers and in discussions and intellectual debates at the two recent SRTL forums (Bakker, 2004; Ben-Zvi, 2004; Hammerman & Rubin, 2004; Ben-Zvi & Amir, 2005; delMas, Garfield & Ooms, 2005; Makar & Confrey, 2005b; Pratt & Prodromou, 2005; Rubin, Hammerman, Puttick & Campbell, 2005; Wild, 2005). For space considerations, some technical details regarding the data used in this study and coding schemes are omitted, and can be found in Reading and Reid (2005b) and Reid and Reading (2004).

4. METHODOLOGY

4.1. SUBJECTS

The study is based on responses collected from 57 students enrolled in an introductory statistics course, at a regional Australian university. The participants were those who consented to participate in the study out of 207 students in the course.

4.2. TASKS AND PROCEDURE

Students completed four minute papers, presented in the Appendix. A minute paper is an informal writing task that consists of a short question given at the beginning, or end, of a class to be completed in five or ten minutes and submitted immediately. For more detail about the use of minute papers see Reid and Reading (2004). The minute papers addressed four key themes in the course: exploratory data analysis (MP1), probability (MP2), sampling distributions (MP3), and inferential reasoning (MP4). Students responded to the minute papers during non-compulsory lectures, both before and after an instructional sequence related to each of the four themes. Later in this paper the letters a and b are used to designate before and after assessments (e.g., Minute Paper 1 had two versions, MP1a (before) and MP1b (after)). In each case, the ‘a’ paper involved as little use as possible of statistical symbols or terminology. The minute paper questions were displayed on an overhead transparency. Before each was completed, points of clarification were addressed to ensure that all students were clear about the requirements of the task, in particular understanding of graphical representations. This clarification was restricted - no explanations were given to inform the question given in the minute paper.

4.3. CODING

Initially the minute paper responses were independently coded in relation to consideration of variation by the two authors (researchers) using the Reading and Reid (2005a) hierarchy (Figure 1). This allowed a separation of the responses into groups with no, weak, developing and strong consideration of *variation* respectively. This initial grouping based on consideration of variation, as a lens through which to investigate reasoning about *distribution*, was undertaken because of the strong connection between variation and distribution in the published literature. The responses in each grouping (with the exception of ‘no’) were then re-analysed to determine any indications of reasoning about distribution. The five key elements of distribution: centre, spread, density, skewness, and outliers, were used as an organizing framework.

One of the researchers was an instructor in the course but was not involved in the data analysis until the course was completed, as required by the ethics approval. Inter-coder reliability was good (i.e., greater than 80%), for all but two minute papers. When there were disagreements about the coding level of a response, each of the two researchers explained what aspect of the response had caused her to choose the particular level. The ensuing discussion, and negotiation, about the interpretation of the response resolved itself in every case.

5. RESULTS

Results of the initial coding for level of consideration of *variation* are summarized in section 5.1. For more detailed discussion of the methodology and examples of responses at each level see Reading & Reid (2005b). Results of the re-analysis to determine any indication of reasoning about *distribution* are summarized in section 5.2.

5.1. CONSIDERATION OF VARIATION

Coding of the minute paper responses identified three levels of consideration of *variation*: no, weak and developing. None of the responses were coded as 'strong.' Table 1 is provided to inform the reader about how many responses were used to develop the evidence reported in the next section. No comparative analyses about the student before and after performance are reported in this paper, as the focus is not on measuring change but rather developing a hierarchy, treating all available responses as equally important. The number of responses available for analysis was disappointingly low, ranging from 48 for MP1a down to 12 for MP3b. This may have been partially due to the fact that the minute papers were completed in lecture timeslots and attendance varied. The last line in Table 1 reports on inter-coder reliability (ICR) for each of the minute papers. More technical details regarding coding appear in Reading and Reid (2005b).

Table 1. Consideration of Variation - Percentages of responses for minute papers (adapted from Reading & Reid, 2005b)

	Minute Paper 1 (EDA)		Minute Paper 2 (probability)		Minute Paper 3 (sampling distr)		Minute Paper 4 (inferential stat)		
Level	MP1a (*n*=48)	MP1b (*n*=26)	MP2a (*n*=40)	MP2b (*n*=31)	MP3a (*n*=26)	MP3b (*n*=12)	MP4a (*n*=18)	MP4b (*n*=22)	Total (*n*=223)
no	12	8	0	3	11	0	22	4	8
weak	71	65	70	36	81	83	45	14	59
developing	17	27	30	61	8	17	33	82	33
Total	100	100	100	100	100	100	100	100	100
ICR	82%	77%	93%	81%	85%	75%	89%	91%	

5.2. REASONING ABOUT DISTRIBUTION: EVIDENCE FROM RESPONSES

Following is the reasoning about *distribution* identified in the minute paper responses. It should be remembered that the 'weak,' and 'developing' terms referred to in this discussion are a measure of the consideration of *variation* demonstrated. So, for each minute paper, the indications of reasoning about distribution are described first for the weak (consideration of variation) responses and then for the developing (consideration of variation). For convenience, information from both the 'a' and 'b' responses are combined for each minute paper. When student responses are reproduced in full they are labeled SR1, SR2, and so forth for reference.

Minute Paper 1 (exploratory data analysis) The weak (consideration of variation) responses only focused on spread and centre, and used terms that suggested consideration of shape that were either incorrect, or not explicit enough to indicate understanding. Spread was mostly expressed as end values for the distribution, although some responses also described how the scores were positioned within the range. Some considered the dispersion (e.g., "distributed evenly throughout" - SR1), while others considered the grouping together (e.g., "major group or clump"), indicated how the spread was centred, (e.g., "condensed at a different point), or used the average to represent a modal cluster.

Attempts at describing the shape included "right skewed" and "left skewed" without elaboration, and the effect of shape on the measure of centre by claiming that the "median is pulling down causing the graph to be skewed," while skewness was indicated by a "very compact bottom 50%" (SR2). Some responses tried to link features of the distribution (e.g., linked the behaviour of the middle 50% of the distribution to that of the range).

SR1 There are more fish that weigh between 0-400 grams in the Perch Species and then the rest of the species are spread more evenly between 400-1000 grams. The bream species on the other hand primarily weigh over 200 grams and the different weights are *distributed evenly throughout*. This shows that there is a difference in weight for the two species.

SR2 There is some difference in the weight of the 2 species. However for the most part they have a similar range. Species B has a larger range than species A, the data for species B is heavily skewed, it has a *very compact bottom 50%* suggesting a concentration of weights towards the bottom.

The developing (consideration of variation) responses more clearly indicated the amount of spread and how it was centred, where "more condensed" was explained as one set "around the middle heavy weights" and the other at "the lower weights" (SR3). Some gave more information about the ends of the distribution by comparing extreme data values rather than where the majority of the data were concentrated (SR4). Many of the developing responses referred to the density of the distribution as well as the shape, sometimes explaining away inconsistent shapes as an anomaly due to outliers. Sources of variation outside the scope of the data, or sampling error, were also used to explain different ranges when the two distributions were basically the same (i.e., the middle 50% of the distributions were similar (SR5)).

SR3 Yes. The perch are more distributed in their weight than the bream making the bream heavier as it is *more condensed around the middle heavy weights* (e.g. 500-1000 grams) whereas the perch are greatly varied in weight reducing the total mass and they are more *condensed at the lower weights* than the higher masses.

SR4 Bream has an upper weight recorded at approx 1000g while perch has an approx 1100g. Therefore there is a *weight difference at the upper limit*, though not significant and due to the variability of weights can not conclude that there is an upper limit weight difference. There is a more significant weight *difference at the lower extremes of weight*, of approx 10g to 250g. However both species will have minimum weights of a gram or two. Thus not conclusive.

SR5 There is not! Though the two species have different means and different ranges, there [*sic*] *mid 50% weights in fact line up*. The differing ranges could be due to inaccurate sampling or another variable, as their lowest weights should be similar and stretch upwards from zero.

Minute Paper 2 (probability) For MP2a most weak (consideration of variation) responses opted for the always 50% red scenario, thus allowing no variability in outcomes and identifying the centre of the distribution but not shape, with such terms as “clustered” (SR6) and “around the 5 lolly mark.” MP2b gave little information on reasoning about distributions because the question was often misinterpreted, and responses did not take into account the importance of order thus considering the two different situations as the same. One weak response, in particular, showed the conflict between theory and intuition (personal experience), choosing the mixture MFFM as more likely despite producing a calculation that showed the same probability for all combinations (SR7). Others attributed similarities in probability to the small sample size.

SR6 C because they have a 50% chance of picking a red one, as there are 100 candies in total & 50 red ones. 50% of 10 = 5 & scores are *clustered* around & on five.

SR7 (a) Both are likely because the probability for having a male or a female is equal. Although it may be more likely to have a mixture (MFFM) rather than all females (FFFF) probability wise *either could happen* P(any) = 1/2 × 1/2 × 1/2 × 1/2.
(b) As in question (a) the probability for both scenarios is equal and therefore all F and FFMM are likely.

The developing (consideration of variation) responses generally indicated shape (arrangement of the numbers) as well as the centering (e.g., giving a range of numbers), “with some less than 5” and “some more than 5” (SR8), that suggest some appreciation of balancing of the distribution around the average value.

SR8 B because the majority of the lollies in the jar are red (50 out of 100). This in theory indicates that you would be most likely to pull out a larger number of red lollies. B gives a variable range of numbers-> *some less than 5 some more than 5*. B averages around 5.

Minute Paper 3 (sampling distributions) Some weak (consideration of variation) responses failed to recognize that the question was focused on the distribution of the sample means rather than of the parent population. Many thought that the distribution of the sample means would be the same as the original population, more often elaborating on the mean of that distribution than the amount of variation, or attributing the changing nature of the distribution to the possible occurrence of extremes. Some incorrectly attributed greater variance to the distribution of the sample means (SR9) rather than to the distribution of individual values, and many described clumping of data, allocating data to within one standard deviation either side of the mean. The better responses included the more detailed information required to discuss the variability of the distribution, although not explicitly acknowledging it as standard error. Some idea of the density of the data was suggested with “a lot will be close to the population mean and then fewer will extend to the edges” (SR10).

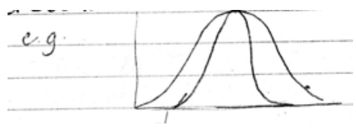
SR9 The means of the 100 sample sizes are going to have *a greater variance* than that of the whole population, the bigger the sample, the closer it is to the mean.

SR10 The values of the mean will be distributed either side of the population means [*sic*]. But *a lot will be close to the population mean and then fewer will extend to the edges.*

Developing (consideration of variation) responses gave more attention to the shape but mostly did not make it clear whether the variance was in fact less than that for the original population (e.g., "not much variance" (SR11)). Some responses indicated an appreciation of the sample size effect on the distribution (e.g., a sample of 100 limiting "the error that occurs in small samples"). One response clearly showed diagrammatically the expected shape of the sampling distribution but did not articulate this well in words (SR12), while another got closer to the notion of standard error by identifying that "σ will become smaller" with the majority of values being closer to the true mean.

SR11 The means of each of the samples would be fairly close *with not much variance*. This is because the samples are all the same size and are repeated within the same population. Also the sample size of 100 each time is a good number as this *will limit the error that occurs in small samples*.

SR12 The distribution of the sample means will mimic that of the population but over a smaller area in the center of the pop. distribution.



Minute Paper 4 (inferential reasoning) The weak (consideration of variation) responses clearly referred to measures of centre, as means or medians, or the spread, as range. One compared the number of dots on either side of the mean (SR13) suggesting that density of the data may also be important. Some mentioned overlap but did not elaborate (SR14).

SR13 Example 1 and 2 because they have a bigger range than 3. Example 1 has *more dots on one side of the mean than the other* which might change the mean.

SR14 No, all boxplots show an *overlap*.

The developing (consideration of variation) responses clearly considered the density of the distribution, usually described as clustering of some form with some responses being more specific about the location of the clusters. Overlap of data was elaborated by stating what was overlapping and connecting this to the conclusions drawn. Interpretations of overlap varied considerably: some were very specific about overlapping boxes (SR15), or whiskers, or both; but others were more vague about the overlap (SR16), not indicating what was being compared. Very few responses actually mentioned the word distribution.

SR15 Yes, there is significant difference in regards to the Wren species. It is not overlapping with the other species at all (i.e., *its central 50% doesn't overlap with the central 50% of the other species*). The Hedge Sparrow and the Meadow Pipit also are significantly different because there is no overlap between them either. The other four species overlap too much for there to be any difference in the mean.

SR16 Example 3 has a real difference in group means. None of the data plotted *overlaps* and it is *very clustered so that each mean is separate from the other.*

6. DISCUSSION OF STUDY RESULTS

What was learnt about students' reasoning about distribution from their reasoning about variation? The reasoning about distribution as evidenced in minute paper responses with weak (section 6.1) and developing (section 6.2) consideration of variation is discussed. These insights into students' reasoning about distribution contributed to conjectures (section 6.3) about how consideration of variation provides a foundation for reasoning about distribution.

6.1. REASONING ABOUT DISTRIBUTION BY STUDENTS WITH WEAK CONSIDERATION OF VARIATION

More than half the minute paper responses demonstrated weak consideration of variation. Of interest now is what these responses indicated in terms of reasoning about distribution. These responses rarely demonstrated a sound understanding of the key elements of the distribution, or ability to reason about distribution in context, and incorrectly linked increased sample size to increased variation. Most responses focused on some measure of location (mean, median, mode) and possibly the range of the data. Some responses did incorporate terminology suggesting consideration of more than the centre and spread. Terms such as "clumped" and "condensed at a different point" gave some sense of the shape and density, respectively, of the distribution, although the responses did not demonstrate a sound understanding of distribution. Those responses that made reference, using standard (e.g., "right-skewed", "outliers") or non-standard language (e.g., "a very compact bottom 50%"), to some of the other key elements that characterize a distribution rarely included sufficient detail to indicate a sound understanding of the links between these key elements. Any response attempting to link some key elements and/or use them for comparative purposes, did so incorrectly.

6.2. REASONING ABOUT DISTRIBUTION BY STUDENTS WITH DEVELOPING CONSIDERATION OF VARIATION

One-third of all minute paper responses demonstrated a developing consideration of variation. Such responses discussed and explained the amount of variation within and between distributions, explained the effect of variation on the distribution, and used that information to justify their inference. So, again, of interest is what these responses indicated in terms of reasoning about distribution. The responses moved beyond a limited focus on centre and spread, often making a link between these two key elements. Many demonstrated a sound understanding of at least some of the other key elements of distribution. Furthermore, some referred to the density of the distribution (e.g., "more condensed", "bunched"), building up a better picture of the shape of the distribution. In addition, many were able to use the information gained from linking the key elements for comparative purposes, discussing overlap of the distributions, or parts of the distributions (e.g., "clusters … are confined to different areas"). A discussion of overlap of two distributions leads towards recognition of the link between the systematic (between-group) and random (within-group) variation suggesting an intuitive analysis of variance. However, few responses were able to successfully apply their understanding of centre, spread and density to the complex notion of the sampling distribution of the mean. Chance et al. (2004, p. 314) have previously identified the difficulty of understanding the concept of the sampling distribution without an understanding of distribution and variation.

6.3. CONJECTURES ABOUT VARIATION – DISTRIBUTION LINKS

A student's ability to understand and articulate variation may be an indicator of a student's ability to reason about distribution. The minute paper analysis showed that in their efforts to discuss, explain and use the concept of variation, the responses indicated that students had developed a refinement of their understanding of many of the key elements of distribution: centre, spread, density, skewness and outliers. This suggests that consideration of variation is an important tool for unlocking the mystery of how students reason about distribution. Other influencing factors are: use of non-standard language (discussed below), interpretation of the task, interpretation of data representation, and discussion with peers (for more detail see Reading & Reid, 2005b).

Especially important to unlocking the mystery of reasoning about distribution is what was evident in developing responses but not evident in weak responses. Responses that exhibited a developing consideration of variation generally demonstrated a more advanced understanding of at least some of the key elements of the distribution compared with weaker responses. Many linked the key elements to compare distributions, thus demonstrating a more sophisticated reasoning about distribution. Although some weaker responses demonstrated intuitive understanding of key elements of distributions, only those responses that had a more developed consideration of variation were able to draw these key elements together to reason better about distribution, through the language they used and the links they made.

Responses showed a variety of non-standard terms to describe and compare distributions, such as "clustered" and "compact 50%," which reflect some appreciation of the density of a distribution. Furthermore, discussion of the overlap of distributions indicates that students are moving closer to inference based on their reasoning about distribution. Although there was an increase in the use of standard statistical terms and notation during the course, the responses continued to include ideas using non-standard terms but were able to be more precise about the meaning of both standard and non-standard terms. Makar and Confrey (2005a) found that it was important to be able to use non-standard terminology to express views, even when correct terminology is known. Students should be encouraged to use non-standard terminology to express their ideas to ensure that they understand the concepts clearly, while familiarising themselves with the corresponding statistical terminology.

This study has demonstrated that consideration of variation is important for students in developing their reasoning about distribution. Those students, who are unable to appreciate a need for variation, nor describe it, are not in a position to identify, understand, and use the key elements of a distribution. They will not have the concepts or the language to describe what they see or visualize, and consequently will be unable to reason about distribution in context.

6.4. LIMITATIONS

When interpreting the above results, the limitations of this study, in relation to the sample, task, procedure and resulting analysis, should be considered. The sample size for the minute papers, while providing sufficient responses for analysis, was not as large as had been planned. The minute papers varied in usefulness for studying reasoning about distribution. In particular, those based around the probability theme need to be redesigned to better facilitate students' expression of their reasoning about distribution. One possible limitation of the minute papers, in terms of the potential for depth of response, was the restricted time allowed for completion. As always, with qualitative research, there were issues based around the interpretation of students' responses. This research should be viewed as the researchers' interpretation of what these particular students were sharing in their responses for this particular course. While useful for guiding other educators and researchers these conclusions may not necessarily be universally applicable.

Thus far, the study has been outlined and the key indications of reasoning about distribution, which were evident in the responses previously coded on their level of consideration of variation, have been described. In the following section these indicators are combined with the findings reported by other researchers, both at SRTL-4 and elsewhere, to propose a hierarchy of reasoning about distribution (second phase) that uses SOLO as a framework.

7. EMERGING HIERARCHY OF REASONING ABOUT DISTRIBUTION

The proposed *Hierarchy of Reasoning about Distribution* (Figure 2) was informed by the observations of reasoning about distribution evident in responses coded according to their consideration of variation (in section 6) and based on SOLO levels of cognitive development. The hierarchy is arranged with increasing sophistication in dealing with the key elements of distribution: centre, spread, density, skewness and outliers. Two cycles of levels based in the concrete-symbolic mode are described.

CYCLE 1	Understanding the key elements of distribution
Prestructural (P1)	does not refer to key elements of distribution
Unistructural (U1)	focuses on one key element of distribution (centre, spread, density, skewness or outliers)
Multistructural (M1)	focuses on more than one key element of distribution
Relational (R1)	develops relational links between various key elements of distribution
CYCLE 2	Using distribution for statistical inference
Prestructural (P2)	recognizes the concept of distribution but does not use it to make inferential statements
Unistructural (U2)	makes one inferential statement described in such a way as to indicate a correct understanding of the concept of distribution
Multistructural (M2)	makes more than one inferential statement described in such a way as to indicate a correct understanding of the concept of distribution

Figure 2. Hierarchy of Reasoning about Distribution

The first, well-defined cycle of P-U-M-R levels is based on an understanding of the key elements of distribution. The responses that exhibited weak consideration of variation, as described in section 6.1, informed the P-U-M part of this cycle. The developing responses that demonstrated a linking of the key elements, as described in section 6.2, provided the background for the relational level (R2) in this first cycle. The second cycle of levels, based on using distributions for making statistical inferences, could only be partially defined based on the better developing responses described in section 6.2. The responses that were able to make some inference informed the unistructral (U2) and multistructural (M2) levels, depending on whether one or more inferential statements were made. Analysis of responses incorporating more sophisticated reasoning about distribution is needed to further develop this second cycle. It is anticipated that this may have been possible from responses that demonstrated strong reasoning about variation but such responses were not available in the study reported.

Note that the relational level (R1) of the first cycle is equivalent to the pre-structural level (P2) of the second cycle, in that the key elements have been linked to form the concept of distribution but the distribution itself is not used for statistical inference. Thus two cycles of cognitive development have been identified: the first based on understanding the key elements of distribution, and the second about using distribution for statistical inference. This is consistent with the Jones et al. (2004) and the Watson et al. (1995) descriptions of two cycles of SOLO levels of statistical reasoning: the first associated with development of understanding of concepts, and the second associated with the application of these concepts.

Before expanding on the levels of the hierarchy it is necessary to consider a terminology issue raised at SRTL-4. To allow for other non-standard ways of determining where the distribution is located on the axis, researchers suggested altering the ‘centre’ element to ‘location.’ However, the authors decided to retain the term ‘centre,’ but allow it to include references to the more general concept of location as well as standard statistical measures of centre.

7.1. CYCLE 1 – UNDERSTANDING THE KEY ELEMENTS OF DISTRIBUTION

In this first cycle the focus is on the key elements themselves (i.e., centre, spread, density, skewness and outliers), and not on the distribution as a whole. The way that data are distributed is dealt with in an informal fashion.

Prestructural (P1) Responses do not refer to any of the key elements of distribution. It is likely that such responses indicate a problem dealing with the representation, either graphical or numerical. For a discussion of levels of understanding of data representation see Reading (1999).

Unistructural (U1) Responses refer to just one key element of distribution. For example, two datasets may be compared based on the range only, rather than taking into account whether the data representation is bumpy or flat. Responses showing weak consideration of variation that described just one key element of distribution fall into this category. Generally this single key element was a measure of centre or spread (i.e., if only one key element is discussed it is less likely to be the density, skewness or outliers). Ben-Zvi and Amir (2005) found that seven year olds only see the relevance in the actual values of the data and not in how many there are of each value. This flat (one-dimensional), rather than distributional (two-dimensional), view of the data did not allow them to reason with distribution. Similarly, responses given at a unistructural level (i.e., dealing with centre or spread), indicate a one-dimensional view of the distribution of data that needs to be expanded to two dimensions.

The complexity of each key element of distribution was emphasized by delMas et al. (2005), for example, regarding what density meant in relation to the histogram representation. The complexity of any one key element would need to be resolved before it would be possible to give a multistructural response, dealing with more than one key element. Sometimes, when using the SOLO taxonomy for coding, there are responses that show features to suggest coding at a particular level but incorrect conceptualization prevents this. Such responses are described as transitional. Some responses, transitional to multistructural, tried to include another key element but not in an acceptable form.

Multistructural (M1) Responses refer to more than one key element of distribution but do not link the various key elements. Most noticeable at this level is the discussion of shape as more than one key element of the distribution has been assimilated. Some weak consideration of variation responses did incorporate terminology, using standard (e.g., "right-skewed," "outliers") or non-standard language (e.g., "a very compact bottom 50%"), suggesting consideration of more than just the centre and spread. Terms such as "clumped" and "condensed at a different point" gave some sense of the shape and density, respectively, of the distribution. There were some responses, transitional to being relational, that attempted to link key elements but this was not correctly done.

Another issue which arises at this level is "cut-points" for dividing a visually presented dataset, as discussed by Rubin et al. (2005) based on their work with teachers. Such points may indicate centre by showing where the distribution is located on the axis, but students decide where to cut based on density, thus indicating more than just a consideration of location. Rubin et al. (2005) also found that the teachers ignored outliers and chose to deal with a simpler set of data. In that instance, the software had made it easy for them to ignore the outliers and recalculate statistics for their inferences. Effectively these teachers were removing the problem of dealing with a skewed distribution. Such action may be a form of simplifying the linking process by removing some of the complicating key elements.

Relational (R1) Responses make links between the various key elements of distribution. Some of the developing consideration of variation responses explained the effect of variation on the distribution by discussing the amount of variation within and between distributions, and used that information to justify their inference. The simplest links made were between centre and spread, with links to density in some way (e.g., "more condensed," "bunched"), building up a better picture of the shape of the distribution. Bakker (2004, p. 65) emphasized the complexity of the distribution concept and the possibility of dealing with it initially in a less formal way by focusing on shape. The relational level of cognition required to deal with shape, linking the two key elements density and skewness, confirms that an appreciation of shape is important for making the difficult 'jump' to the actual concept of distribution. Relational responses consider the distribution as "an aggregate with its own characteristics," as described by Makar and Confrey (2005a, p. 28) and others (see, e.g., Bakker & Gravemeijer, 2004, p. 148). This linking of the various key elements of distribution (aggregation) allows the move to more complex reasoning using distribution.

7.2. CYCLE 2 – USING DISTRIBUTION FOR STATISTICAL INFERENCE

In this second cycle the focus is on the distribution as a whole and its use as a tool for making statistical inferences. The importance of understanding distribution to enable students to comprehend standard deviation was highlighted by delMas and Liu (2003). Hammerman and Rubin (2004) found that even when students were able to deal with data as an aggregate there were still complex processes needed to move away from just considering variation. Statistical inference, whether formal or informal, involves dealing simultaneously with signal (centre), noise (variability), sample size and shape of the distribution. This second cycle involves recognizing the distribution as an aggregate and being able to move on and use this concept of distribution for inference. The notion of a student moving from a data-centric (data spread across a range of values) to a modelling perspective (variation as a random movement away from the main effect), as outlined by Pratt and Prodromou (2005), should help to explain the important move from the first to the second cycle of reasoning about distribution. The need to recognize that there is a family of distributions that make up a model as variables change value (Wild, 2005), would also be critical to being able to use distributions for inferences, thus moving into the second cycle.

Pre-structural (P2) This is equivalent to the relational level in the previous cycle (i.e., there is no indication of statistical inferences being made using distribution which is essential for the second cycle). Responses make the necessary links to perceive the distribution as a “whole” but do not make any steps towards using distribution in detailed statistical inference.

Unistructural (U2) Responses make one inferential statement described in such a way as to indicate a correct understanding of the concept of distribution. Many of the developing consideration of variation responses were able to use the information gained from linking the key elements for comparative purposes, discussing overlap of the distributions, or parts of the distributions (e.g., “clusters ... are confined to different areas”). A discussion of overlap of two distributions indicates recognition of the link between the systematic (between-group) and random (within-group) variation and is suggestive of an intuitive analysis of variance. Also important here is the need for an understanding of the concept of distribution to be able to work with the complex notion of the sampling distribution of the mean.

Multistructural (M2) Responses make more than one inferential statement described so as to indicate a correct understanding of the concept of distribution. There were not enough responses of a sufficient quality to allow elaboration on the definition of this level.

Relational (R2) There were no responses to allow description of this hypothesized relational level of cognition for the hierarchy. It is anticipated that responses at this level would be able to link together the inference statements made thus indicating a strong understanding of the concept of distribution.

8. IMPLICATIONS

This paper has provided a suggested alternative approach for use of the SOLO Taxonomy and also contributed to the ongoing development of hierarchies in statistics education research. A conventional use of the SOLO Taxonomy would involve directly coding student responses in relation to their reasoning about distribution. The approach used in this paper was somewhat different - the researchers chose initially to code (no, weak, developing) the responses in relation to the underlying concept of variation and then apply the SOLO Taxonomy to the reasoning about distribution found in the grouped responses.

As reasoning depends heavily on an understanding of underlying concepts (Garfield, 2002), it was not unexpected that the better indications of reasoning about distribution were found in the responses with a higher level of consideration of variation. It should be remembered, however, that like the coding of all open-ended responses, the indicated levels are only what the student was able to demonstrate at that particular time to that particular question. While the described hierarchy can be used as a guide to the types of responses that may occur for other questions, there is no guarantee that students will achieve at a similar level on a different question. In fact, whether reasoning about distribution occurs at all will depend on the nature of the task. It is not sufficient to provide a situation where students are merely asked to describe a distribution. They need activities that require working with distribution in some way. For example, a comparison task can provide the motivation to reason about distribution (Ben-Zvi, 2004; Makar & Confrey, 2003).

The Hierarchy of Reasoning about Distribution proposed in this paper has added to previous research on the cognitive development of distribution as a concept. This hierarchy is consistent with, and elaborates on, the detail provided by the five level hierarchy of *use of statistical evidence* proposed by Makar and Confrey (2005b). Cycle 1 and cycle 2 of the proposed hierarchy (Figure 2) correspond to Makar and Confrey’s final two levels: Level 4 – Distribution and Level 5 – Inference. The P1-U1-M1-R1 levels described in cycle 1 (understanding the key elements of distribution) represent a deeper articulation of Makar and Confrey’s Level 4. While the P2-U2-M2 levels described in cycle 2 (using distribution for statistical inference) provide an insight into Makar and Confrey’s Level 5. More work is needed in this area to inform the description of cycle 2, especially the R2 level.

8.1. IMPLICATIONS FOR RESEARCH

Implications for researchers arise from both the existing hierarchy and the proposed hierarchy. The existing *Hierarchy of Consideration of Variation* (see Figure 1) can be used by other researchers to assess students’ consideration of variation. Such background information about students in relation to one of the key fundamental statistical thinkings, consideration of variation, can then be used to determine whether responses indicate a readiness to develop cognitively in relation to other related statistical concepts. In the case of the study reported in this paper, this background information about consideration of variation was used to arrange responses, thus separating those with the less developed indicators of reasoning about distribution from those with the more developed indicators. The *Hierarchy of Reasoning about Distribution* (see Figure 2) proposed in this paper sets two challenges for researchers. One is to elaborate on this hierarchy description, particularly in the second cycle by analysing responses that exhibit a ‘strong’ level of consideration of variation. The second is to use the hierarchy to code responses from larger and more diverse groups of students, and test the hierarchy’s validity as an instrument to allow students’ reasoning about distribution to be measured thus allowing developing reasoning to be mapped.

8.2. IMPLICATIONS FOR TEACHING AND ASSESSMENT

The conclusions drawn are useful for guiding educators in both learning activity design and assessment. When designing learning activities for students, educators need to plan for knowledge development that will assist students to move from their informal notions to more statistically sophisticated notions. Such planning, in relation to distributions, should focus heavily on nurturing students' conceptions of variation and extending these to reasoning about distribution. It is proposed that activities that use distributions, but do not expect sophisticated reasoning about distributions, be used to allow students to progress naturally through the first cycle of cognition, developing a strong understanding of the concept of distribution through its key elements, before being expected to use it for statistical inference. This is especially important for students who are identified as having weak consideration of variation and hence will need to be given the opportunity to develop a better appreciation of variation. Educators should note that this research has also demonstrated that assessment tasks designed for one purpose can be used for other purposes. Tasks designed to assess outcomes in core themes can also be used to identify indicators of reasoning about distribution. Further research is now needed to assist educators to develop more assessment tasks that are multipurpose and also to develop supportive learning strategies to nurture reasoning about variation, thus laying a firm foundation for reasoning about distribution.

ACKNOWLEDGEMENTS

This paper is based on research data presented at SRTL-4 and reported in detail in the proceedings. The work of presenters at SRTL-4 and SRTL-3, as well as other researchers who have published in this field, has contributed to the development of the hierarchy. The most important acknowledgement, though, is to the participants at SRTL-4 who suggested taking the research one step further to propose a hierarchy and provided the stimulating discussion and research reports that assisted the development of the hierarchy. Special acknowledgement goes to Iddo Gal and the anonymous referees for their constructive critiquing of this paper.

REFERENCES

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83.

Bakker, A., & Gravemeijer, K. (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, The Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42-63.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_BenZvi.pdf]

Ben-Zvi, D., & Amir, Y. (2005). How do primary school 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)*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Biehler, R. (1997). Student's difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J.

Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, The Netherlands: International Stat. Inst. [Online: www.stat.auckland.ac.nz/~iase/publications/8/14.Biehler.pdf]

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

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.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty in reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4)*, Auckland, 2-7 July 2005, [CD-ROM, with video segments],. Brisbane, Australia: University of Queensland.

delMas, R., & 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*). [CD-ROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3).
[Online: www.amstat.org/publications/jse/v10n3/garfield.html]

Hammerman, J., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Hammerman_Rubin.pdf]

Jones, G. A., Langrall, C. W., Mooney, E.S., & Thornton, C. A. (2004). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 97-117). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Makar, K., & Confrey, J. (2005a). "Variation-Talk": Articulating meaning in statistics. *Statistics Education Research Journal*, *4*(1), 27-54.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_Makar_Confrey.pdf]

Makar, K., & Confrey, J. (2005b). Using distributions as statistical evidence in well-structured and ill-structured problems. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4)*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]..Brisbane, Australia: University of Queensland.

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

Moore, D. S., & McCabe, G. P. (2003). *Introduction to the practice of statistics*. New York: W.H. Freeman and Company.

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

Pfannkuch, M. (2005). Characterizing year 11 students' evaluation of a statistical process. *Statistics Education Research Journal, 4*(2), 5-25.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4(2)_pfannkuch.pdf]

Pratt, D., & Prodromou, T. (2005). The emergence of distribution from causal roots. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4),* Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Reading, C. (1999). Understanding data tabulation and representation. In O. Zaslavsky, (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 97-104). Haifa, Israel: Technion – Israel Institute of Technology.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84-105.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Reading.pdf]

Reading, C., & Reid, J. (2005a). Consideration of variation: A model for curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable* (pp. 36-53) [CD-ROM]. Vorburg, The Netherlands: International Statistics Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/rt04/2.3_Reading&Reid.pdf]

Reading, C., & Reid, J. (2005b). Reasoning about variation: A key to unlocking the mystery of distributions. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4),* Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Reading, C., & Shaughnessy, C. (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.

Reid, J., & Reading, C. (2004). Just a minute? The use of minute papers to investigate statistical thinking in research, teaching and learning. *Adults Learning Maths Newsletter, 21*, 1-4.

Rubin, A., Hammerman, J., Puttick, G., & Campbell, C. (2005). The effect of distributional shape on group comparison strategies. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4),* Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Tippett, L. H. C. (1952). *The methods of statistics* (4th ed.). New York: John Wiley & Sons, Inc.

Utts, J. (2005). *Seeing through statistics* (3rd ed.). Belmont, CA: Thompson Brooks/Cole.

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

Watson, J., & Kelly, B. (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). [CD-ROM, with video segments] Mount Pleasant, Michigan: Central Michigan University.

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

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145-168.

Wild, C. (2005). A statistician's view on the concept of distribution. In K. Makar (Ed.), *Reasoning about Distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4),* Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

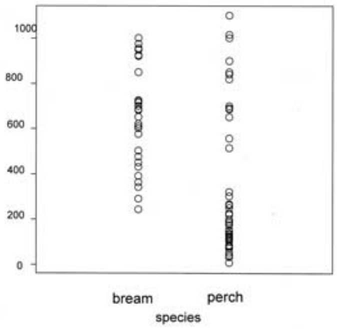
Wild, C. J., & Seber, G. A. F. (2000). *Chance encounters: A first course in data analysis and inference.* New York: John Wiley & Sons.

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

APPENDIX: MINUTE PAPER QUESTIONS

Minute Paper 1a

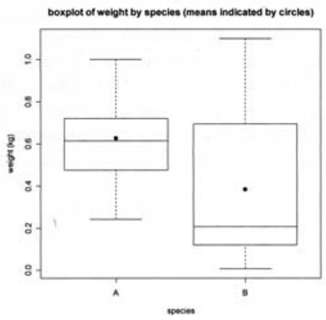
Look at the following plot. It shows the weights in grams of two species of fish (bream and perch).



Do you think that there is a difference in weight for the two species? Explain your response.

Minute Paper 1b

Look at the following boxplots. They show the weights in kg of 2 different species of animal (A & B).



Do you think that there is a difference in weights for the two species? Explain your response.

Minute Paper 1

Minute Paper 2a

A bowl has 100 wrapped hard candies in it. 20 are yellow, 50 are red, and 30 are blue. They are well mixed up in the bowl. Jenny pulls out a handful of 10 candies whilst blindfolded, counts the number of reds, and tells her teacher. The teacher writes the number of red candies on a list. Then, Jenny puts the candies back into the bowl, and mixes them all up again. Five of Jenny's classmates, Jack, Julie, Jason, Jane and Jerry do the same thing. They each pick ten candies, count the reds, and the teacher writes down the number of reds. Then they put the candies back and mix them up again each time.

From the lists choose the one that you think is most likely to represent the teacher's list for the number of reds. Explain why you chose that one.

A. 5, 9, 7, 6, 8, 7
B. 3, 7, 5, 8, 5, 4
C. 5, 5, 5, 5, 5, 4
D. 2, 4, 3, 4, 3, 4
E. 3, 0, 9, 2, 8, 5

Minute Paper 2b*

Suppose the probability of having a male child (M) is equal to the probability of having a female child (F). A couple has four children.

(a) Are they more likely to have FFFF or to have MFFM? Explain your answer.

(b) Are they more likely to have four girls or to have two children of each sex? Explain your answer.

(Assume that the decision to have four children was independent of the sex of the children.)

* Question from J. Utts (2005, p. 346)

Minute Paper 2

Minute Paper 3a

Suppose a sample of 100 women is drawn from a certain population and their heights measured. The mean of this sample is 170.1 cm. Census data indicated that the adult female population has a mean height of 168.4 cm and a standard deviation of 4.5 cm.
If repeated samples of size 100 are taken from the same population of women and the resulting means from each of the samples recorded what can you say about the distribution of these means?

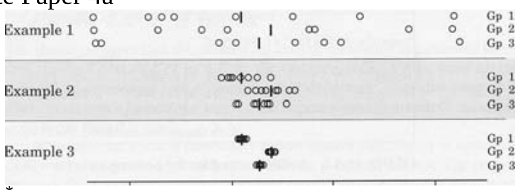
Minute Paper 3b

Suppose a sample of size n is drawn from a population. The mean of this sample is $\overline{x}$. The population has a mean E(X) = μ and a standard deviation sd(X) = σ.
If repeated samples of the same size are taken and the resulting means from each of the samples considered what can you say about the distribution of these values?

Minute Paper 3

Minute Paper 4a[*]

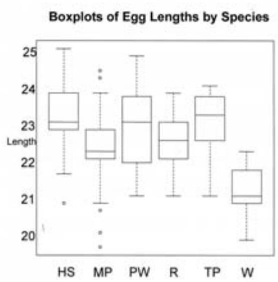
There are 3 different examples. In each example, a sample was taken from each of 3 groups and the data plotted, along with the sample means. Sample means are indicated by vertical lines. For which example(s) might you conclude that there is a real difference in group means? Explain your response.



[*]Diagram from Wild & Seber (2000, p. 439)

Minute Paper 4b[*]

Cuckoos are known to lay their eggs in the nests of other (host) birds. The eggs are then adopted and hatched by the host birds. These data give the lengths (mm) of cuckoo eggs found in the nests of other birds.
This study investigates the difference in mean egg length (mm) of cuckoos' eggs according to the species of the foster-parent. With reference to the boxplot, do you think that there are any significant differences in mean egg lengths among the six species? Justify your response.



HS: Hedge Sparrow
MP: Meadow Pipit
PW: Pied Wagtail
R: Robin
TP: Tree Pipit
W: Wren

[*]Data from Tippett (1952, p. 176)

Minute Paper 4