

## *Consideration of Variation: A Model for Curriculum Development*

Chris Reading  
University of New England  
Australia

Jackie Reid  
University of New England  
Australia

### **Abstract**

*Consideration of variation has been identified as one of the fundamental types of statistical thinking, and recent research into the understanding of variation highlights its growing importance. Curriculum needs to reflect this trend. This chapter proposes a model for an integrated approach to curriculum development where consideration of variation is used as the linking thread. The application of this process to a tertiary introductory service statistics course is presented. A hierarchy of levels of consideration of variation is developed from students' responses to "minute papers" and typical responses in the various levels of the hierarchy are discussed. The hierarchy can be used to evaluate the effectiveness of the integration of variation into the curriculum. This research has implications for teachers of statistics, in the development of curriculum, and for researchers in the growing field of students' understanding of statistics.*

### **Introduction**

A focus of statistics education research on statistical reasoning, thinking, and literacy was highlighted when delMas (2002) introduced a series of articles clarifying the definitions of reasoning (Garfield, 2002), literacy (Rumsey, 2002) and thinking (Chance 2002). Given the recognized importance of "understanding" (a recurring theme in all these discussions) and "variation" (a critical aspect in the study of statistics), the research trend in "understanding variation" is not surprising (Reading & Shaughnessy, 2004; Watson et. al, 2003; Wild & Pfannkuch, 1999, p. 235). A necessary consequence of such a trend is the need to reconsider the planning, implementation, and evaluation of curriculum. The research project, *Understanding of Variation*, funded by a University of New England Internal Research Grant, was designed to inform this trend by considering how understanding of variation is evidenced as students engage in the various learning activities and assessment tasks in a tertiary introductory service statistics course where "consideration of variation" is a core feature of the curriculum.

The project aims to refine hierarchies currently being developed to assess students' understanding of variation, to investigate how this understanding develops, and to identify teaching and learning strategies that assist this development. Such information is critical in formulating curriculum and evaluating its effectiveness. This report proposes a model for curriculum development based on an integrated approach and focuses on the analysis of responses to "minute papers" as a means of evaluating this approach. A minute paper consists of a student's response to a single question, completed on-demand, and addressing some aspect of a focus topic.

### **Background**

#### *Integrated Curriculum*

A standard curriculum development process should address: what is to be achieved; how it is to be achieved; and the expected extent of this achievement. A popular cyclic approach to such a process, prescribed by Nicholls and Nicholls (1981, p. 21) involves revisiting the steps: selecting objectives; selecting and organizing content; selecting and organizing methods; and evaluation (including assessment and feedback). The Nicholls' cyclic approach emphasizes that the approach to content, not just the

content itself, should be a key aspect of the curriculum development process. Most importantly, this approach draws attention to the need for curriculum evaluation.

In an effort to better support the learner and to avoid problems associated with students' fragmented view of the curriculum, integrated curriculum has been proposed as a more holistic approach to content. Unlike traditional curriculum, integrated curriculum aims to develop understandings through sustained interaction, conversation, and discussion (Pidgon & Woolley, 1992). According to Murdoch and Hornsby the goal of an integrated curriculum approach is to extend and refine students' developing knowledge (Murdoch & Hornsby, 1997). The approach is "understanding-driven" and, in addition to providing a means for giving a more holistic view of the content, an integrated approach to curriculum will also strengthen students' understanding of concepts as they are explored through different topics (Murdoch and Hornsby, 1997).

The trend from traditional content-based to approach-based curriculum development necessitates development of a model for planning the approach. One model used to plan integrated curricula is "threading". Potentially useful threads for helping students make connections between various content areas relate to four main "ways of working". These ways include cooperating and interacting, reasoning and reflecting, imaging and inquiring, and assessing and evaluating (Murdoch and Hornsby, 1997, pp. 14-15). Pidgon and Woolley (1992) tie curriculum to iterative assessment in their description of the threads as "planned understandings," which become the driving force for assessing what students already know and what they may have learned as a result of engaging in the learning activities.

### Statistics Curriculum Development

The conventional approach to curriculum, which involves teaching statistics as a series of techniques, is now being challenged. For example, Ballman (1997) attempted to promote greater emphasis on variation by proposing a new approach to probability using activities that challenge students' conceptions of randomness. The need to support such changes was recognized within the American Statistical Association/Mathematics Association of America (ASA/MAA) Joint Committee. As a result, several Focus Group Guidelines were based on the following three fundamental goals: emphasize "statistical thinking"; use more data and concepts, less theory, and fewer recipes; and foster active learning (Cobb 1992, in Scheaffer, 2000, p.159). The importance of these fundamental goals is acknowledged by many who propose changes. Moore (1997) uses the guidelines to structure his discussion of changing content for learners in introductory statistics courses; however, these guidelines are not evidenced in his proposed content-pedagogy-technology synergy. Hoerl, Hahn, and Doganaksoy (1997) remind us of the importance of one specific guideline, statistical thinking, when they respond to Moore about the importance of process. Scheaffer (2000) demonstrates how these guidelines can be used to broaden and deepen the teaching and learning of statistics by basing curriculum on a process of "thinking through a problem from inception, to clarification, to data, to analysis, to conclusion" (p.173), but the suggested overview of a sequence of disjointed topics does not reflect this discussion.

What, then, could be used as a thread for integrating curriculum in a statistics course? Students' understanding, identified as fundamental to curriculum integration by Pidgeon and Woolley (1992), is an obvious suggestion. The link between students' understanding and curriculum development was highlighted by Petocz and Reid (2003). They developed categories of statistics understanding based on how students reported their understanding. They then used these categories to inform curriculum development, concluding that there is a need to accommodate variation in student conceptions.

With this new approach to curriculum comes the need for new approaches to evaluating the curriculum and assessing student learning. These are tasks that would be intertwined under the notion that the model for curriculum development depends on students' understanding. These expanded goals for student learning led Garfield and Gal (1999, p.7) to investigate the development of alternative assessment procedures, such as group or individual projects, portfolios of student work, minute papers, and concept maps. They pose a number of assessment challenges, including assessing students' understanding of "big ideas" and developing models for evaluating and comparing curricula (p. 11).

They conclude that it will be difficult to evaluate the effectiveness of statistics courses that approach content in a new way without first shifting the assessment focus from computations to reasoning.

### Model for Curriculum Development: Fundamental Types of Thinking

When choosing a thread, or “big idea” for understanding, the importance of fundamental concepts should not be underestimated. The understanding of such concepts is critical to the development of better statistical reasoning (delMas, Garfield & Chance, 1999). Wild and Pfannkuch’s (1999, p.226) five fundamental types of statistical thinking could be considered as suitable threads from which to weave the content of a statistics curriculum. The five types of thinking, which emerged from their interviews with statisticians and students, are: recognition of the need for data; transnumeration (changing representations to engender understanding); consideration of variation; reasoning with statistical models; and integrating the statistical and contextual.

Pfannkuch and Horring (2005) focused on the development of statistical thinking in a research project with secondary students and used the Wild and Pfannkuch (1999) framework for statistical thinking to communicate to teachers the types of thinking that need to be fostered in students. Their proposal for a model for curriculum development emphasizes the need to use one or more of these fundamental types of thinking as a thread and to map the teaching and learning activities across content areas.

### Consideration of Variation

As a trial for this basis for curriculum development we decided to focus on only one of Wild and Pfannkuch’s (1999) five types of thinking, *consideration of variation*, as a thread. We chose *consideration of variation* from amongst these fundamental types of thinking because most recent discussions of new approaches to teaching statistics highlight the importance of understanding variation (see, e.g., Garfield & Gal, 1999; Ballman, 1997). We also chose this focus because the ASA/MAA Joint Committee on Undergraduate Statistics (Cobb 1992) included the need for students to recognize the omnipresence of variability and the ability to measure and model variability in their definition of statistical thinking.

When discussing the interaction of teachers in the development and implementation of syllabi, MacGillivray (2005) reinforces the importance of variation by introducing the notion of statistics as the “science of variation”. Furthermore, Pfannkuch, Rubick and Yoon (2002) focus on the *consideration of variation* thinking type in their exploratory study of the thinking involved when secondary students are engaged in data investigations. Meletioui and Lee (2002, p.23) highlighted the negative impact of a deterministic approach in the mathematics curriculum on statistics instruction. Such an approach may leave students with a reliance on the “signal” component of the statistical model without recognition of the “noise” component (i.e., unexplained variation). Furthermore, deterministic approaches may reinforce the fragmented view of statistics, rather than Reid and Petocz’s (2002) more desirable holistic view, which should be central to a complete statistical model. A sound understanding of variation could help curriculum developers, teachers, and students capitalize on the benefits of a holistic view.

Wild and Pfannkuch’s (1999, p. 226) *consideration of variation* includes four components:

1. noticing and acknowledging variation: recognizing the omnipresence of variation and the need to record this variation in discussions;
2. measuring and modeling variation for the purposes of prediction, explanation, or control: creating summaries (numerical or graphical) to represent the variation in the data and using these summaries to represent the impact of variation;
3. explaining and dealing with variation: looking for the causes of variation and considering the impact on design and sampling; and
4. using investigative strategies in relation to variation: formal procedures for looking at the properties of the variation itself.

To best investigate students' understanding of variation it is necessary to delve into as many of these components of *consideration of variation* as possible. This approach was taken by Torok and Watson (2000) when developing their categories of the appreciation of variation.

## **The Study**

### Research Focus

The development of a curriculum to nurture the understanding of variation leads to two key questions. How can *consideration of variation* be used in the modelling process for a curriculum that aims to develop statistical understanding? What evidence of *consideration of variation* emerges in students' *statistical thinking* during assessment activities? The *Understanding of Variation* project aims to identify whether the various learning activities and assessment strategies used in a course where variation is a core concept are effective in developing students' understanding of variation. Pfannkuch, Rubick and Yoon (2002) took a similar approach with their analysis of students' statistical investigations, identifying evidence of the components of *consideration of variation*.

Although traditional teaching and learning activities (including assessment) are still useful, with a new approach to the curriculum comes a need for new tools to assess deeper understandings not just the students' attainment of the course objectives. Ridgway, McCusker and Nicholson (2005) identified a need to embed rich assessment tasks to produce more valuable curriculum materials. Obtaining information on deeper understandings is crucial for the development and refinement of an approach-based curriculum "threaded" on "statistical thinking". This trend to use less traditional strategies for both teaching and assessment (Garfield & Gal, 1999, p. 4) is evident in Stromberg and Ramanathan's (1996, p. 159) suggestion to use "entrance" and "exit" slips (i.e., minute papers) to encourage students' writing skills. Minute papers are an innovative model for assessment (Garfield & Gal, 1999, p. 7). They are brief, anonymous written comments on a question in which students describe their understanding of a particular concept or procedure.

This report focuses on the analysis of responses to minute papers and extends the Pfannkuch, Rubick and Yoon (2002) approach by arranging the *considerations of variation* into a hierarchy and focusing on two additional research questions: What hierarchical levels can be used to describe tertiary students' consideration of variation? Can minute papers be used effectively to illustrate students' consideration of variation?

### Methodology

The two-phase research offered a model for developing curriculum with variation as the core and then evaluated its effectiveness by assessing the evidence of consideration of variation in students' responses. The research targeted a one-semester introductory service statistics course (with an enrollment of 46 students) that is studied in science-related fields at a regional Australian university. The course included a variety of topics with four organizing themes: exploratory data analysis, probability, sampling distributions, and inferential statistics. The text for the course was authored by Wild and Seber (2000). The presentation of the content in this text was considered to support the proposed approach, although throughout each topic the lecturer explicitly referred to the core concept of variation more frequently than the text. The topics forming the content of the course in each of the four themes were mapped across the four components of *consideration of variation*, thus providing a basis for the lecturer to more closely examine how these aspects of variation were treated. Exploratory investigation and categorization of the level of *consideration of variation* evidenced in students' responses were used to evaluate the effectiveness of modeling variation as the core for the curriculum.

Although all enrolled students were expected to complete the various learning activities and assessment tasks as an integral part of the course, responses were only analyzed for the students who

agreed to participate in the research. Data collection and analysis were performed by two researchers, one of whom was the lecturer in the course. The complete project included analysis of student responses to pre-study and post-study questionnaires, follow-up interviews with selected students, four separate minute papers, a question from a class test, and a question from an assignment. Among these activities only minute papers were new to the course. The minute paper analysis is the focus of this report.

The minute paper questions (see Figures 1 – 4) reflect the curriculum themes. Minute paper (MP) 1 focused on exploratory data analysis. MP2 focused on probability. MP3 focused on sampling distributions, and MP4 focused on inferential statistics. The minute papers were displayed on overheads, and each student was expected to write a concise response on a 15 cm x 10 cm (6" x 4") card in five to ten minutes and submit the response immediately. The minute papers were worth an optional 5 percent of course credit for participation.

Although the feedback, given in the form of a teaching sequence, was based on assessment of the minute papers in relation to the course objectives, the focus of the analysis here is the consideration of variation that was evidenced in the responses. The Torok and Watson (2000) proposed *Developing Concepts of Variation* hierarchy was useful as a guide when the researchers identified, and arranged hierarchically, key concepts of Wild and Pfannkuch's (1999) *consideration of variation* that emerged in the responses. Detailed analysis of the responses follows the discussion of the implementation of the model for curriculum development.

### **Implementing the Model for Curriculum Development**

In the following course outline the four components of *consideration of variation*, as defined above, are mapped across the four themes of the curriculum and the impact of the variation-oriented approach on the topics of the four themes is explained. The mapping is summarized in Table 1.

Table 1  
Mapping Curriculum Themes against Components of *Consideration of Variation*

	<i>Noticing and Acknowledging</i>	<i>Measuring and Modeling</i>	<i>Explaining and Dealing with</i>	<i>Investigative Procedures</i>
<i>Exploratory Data Analysis</i>	Introductory examples	Graphical and numerical summaries (MP1)		
<i>Probability</i>	Randomness (MP2), Law of Large numbers, small and large sample variation	Probability models	Calculate probabilities that observed differences are real or random	
<i>Sampling Distributions</i>	Representative sampling, repeated sampling	Simulations of sampling distributions for means and proportions	Relationship between sample size and standard error (MP3)	
<i>Inferential Statistics</i>		Variation within and between groups (MP4), margin of error, confidence intervals	Statistical models; t-tests, simple linear regression, ANOVA, two-way tables	Residuals plots, normal probability plots

The course begins with *exploratory data analysis*, emphasizing the motivating factor behind statistical analysis. Most students are comfortable with measures of central tendency from the secondary curriculum, but the concept of variability is at best a vague notion for many. In the current course students are exposed to a broad range of examples that act as a starting point for discussions about design, data production, sampling, analysis, and the unifying concept of the statistical model. Examples of data collected from the students, themselves, are used to initiate discussions about data collection and

representative sampling. This leads to discussions about repeated sampling and variability in the results. Students are exposed to the concept of variation and its importance without resorting to statistical terminology, such as Central Limit Theorem and sampling distributions, until much later in the course.

The importance of variation is reinforced when comparing distributions. Examples with side-by-side box plots, histograms, dot plots, and scatter plots ensure that students are exposed to a variety of graphical representations. In addition, calculation of numerical summaries reinforces the need for measurements of variation and central tendency. Students begin to recognize, for example, that the consideration of the difference between means or medians is not sufficient when comparing distributions. MP1 requires students to use the graphical summary to draw an informal conclusion about the difference between groups.

Comparative data sets are used to illustrate the need for *probability* in statistics. Students are expected to determine if differences between two distributions are “real” or can be explained by chance variation. Probability is the required tool for making decisions in the face of uncertainty. In MP2 students have the opportunity to explore the concept of randomness.

Simple classroom experiments allow students to look at the “Law of Large Numbers” and its misleading counterpart the “Law of Small Numbers”. Discussions of small sample variation compared with large sample variation are crucial to the course. Computer simulations of repeated sampling from binomial and normal distributions allow students to explore the concept of *sampling distribution* in an interactive setting. Their ability to describe the concept of sampling distribution is assessed in MP3.

In the study of *inferential statistics* the course emphasizes concepts such as reliability and margin of error. Unlike the more traditional focus on hypothesis testing, emphasis is placed on confidence interval estimation. Students begin to realize that reporting a p-value for a hypothesis is not sufficient and that they must also provide an interval estimate of the parameter, thus acknowledging the influence of variation in the data.

Variation provides a common link for topics in this introductory statistics course. The statistical model can be stated as  $DATA = FIT + RESIDUAL$ , or,  $DATA = SIGNAL + NOISE$ . Students are made aware of the impact of unexplained variation (noise). When comparing distributions, they are asked whether the variability between groups (signal) is large compared with the variability within groups (noise). This concept is emphasized graphically with a number of examples, where group means have the same relative value, but the variability differs. Following exploration of these issues, students complete MP4, which requires an inference about the differences between the groups. When examining one and two-way tables of categorical data, students are asked how much variation exists between the observed data and the results suggested by the proposed model, and whether the difference is more likely the result of chance variation or an error in the proposed model. At every opportunity throughout the course, students are reminded of the importance of considering variation in the data and the resulting impact on their interpretation of the underlying model. This common thread links important themes across the curriculum, encouraging students to take a logical and holistic approach to statistical analysis rather than viewing statistics as a fragmented list of techniques to be memorized.

## Results

As the minute papers were implemented during lecture times, not all students completed every minute paper. The response rate was 88 percent for MP1, 85 percent for MP2, 48 percent for MP3, and 52 percent for MP4. The reduced response rates for MP3 and MP4 were due to some students withdrawing from the course and, of those remaining, less students attending lectures towards the end of the course. For each minute paper key concepts in the consideration of variation were evidenced in student responses and provided the basis for describing four hierarchical levels of consideration: *No*, *Weak*, *Developing*, and *Strong Consideration of Variation* (see Table 2).

Table 2

Rubric for Levels of *Consideration of Variation* in Each Minute Paper

<i>Level 0 - No consideration of variation</i>	
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
<i>Level 1 - Weak consideration of variation</i>	
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
<i>Level 2 - Developing consideration of variation</i>	
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
<i>Level 3 - Strong consideration of variation</i>	
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

When statistically immature vocabulary and the developing nature of student understanding made it difficult to determine whether certain key criteria had been met to warrant coding at a particular level, responses were coded as transitional between one level and the next. Following is a discussion of typical responses at the various levels for each minute paper. All responses have a reference code indicating the minute paper being answered, and the student responding, including gender (1: male, 2: female). For example, R215M1 is the response to minute paper 1 from the fifteenth female student.

Minute Paper 1: Exploratory Data Analysis

Look at the following box plots. They show the weights in grams of 2 species of fish (Bream & Perch). Do you think that there is a difference in weights for the two species? Explain your response.

**Figure 1. Minute Paper 1 Question**

MP1 (see Figure 1) was given at the end of a lecture discussing box plot representations and the importance of measures of variation but before any detailed discussion about comparing box plots. The assessment focus here is how well a response demonstrates the use of variation in the justification of the conclusion drawn. The key to providing a better response is the progression from just comparing the

spread of data to linking the spread to the conclusion drawn. Table 3 provides typical responses for each of the levels described in Table 2.

Table 3  
Typical Responses for Minute Paper 1

<i>Level</i>	<i>Response</i>
0 No consideration of variation	(R215M1) There isn't a difference in weight between the two species as seen on the scale both box plot measure roughly 750 grams. (R115M1 <sup>T</sup> ) The Bream have a higher average weight compared to the Perch. This difference could be clue to different environment (Both salt, fresh, one of each). The mean of the two species isn't that different but the medians are very different. In conclusion the Bream have a higher weigh to the Perch but more information is required to get a clearer picture of the two species.
1 Weak consideration of variation	(R219M1) No, I don't think that there is a difference in the weights of the two fish. Perch seems to look different but there is a chance that the weight of the Perch has some outliers. Therefore making it look as though there is a difference in weights. (R114M1 <sup>T</sup> ) The Range of middle 50% of bream is much smaller than the range of the middle 50% for perch showing that a larger percentage of the bream are of similar size whereas the average size of perch vary much more. The mean of bream is in the middle of the range of the weights showing that there are approx. equal numbers on either side whereas the mean for the perch is lower showing that a larger majority of the perch population has less weight. Therefore: feel that the Bream has a heavier population than that of Perch.
2 Developing consideration of variation	(R202M1) I don't think there is that great a difference b/w the two fish weights accept that Perch has a greater weight distribution. Since the boxes overlay each other and the averages are fairly close with no outliers. The median shows the middle of the numbers but still no great diff. only around 100gms. (R213M1 <sup>T</sup> ) The weights for the perch are positively skewed, whereas bream is symmetrical. This could indicate a larger variation in values for perch. The plot shows that Bream weight sizes are all relatively around the same. Obviously Perch species grow to a variety of different weights. The means for Perch are higher. The median indicating a larger variety in weights.
3 Strong consideration of variation	(R118M1) There is a difference between the weights of the species. Although, at a glance, the fish seem to have an approximate equal weight, the mean and median suggest otherwise. In the bream, the median and mean are almost the same, denoting that average weight is around 600 grams. The variance is not as large as in the perch and weight is more predictable. In the perch, weight varies greatly with the large outlying values raising the mean. The median shows us that although perch can grow large, most are around the 200 gram mark. In conclusion, bream and perch grow to be the same weights, but there are more bream of a larger weight than perch.

“T” indicates a response transitional to the next level

Response R215M1 is typical of Level 0, with no mention of variation at all, while response R115M1 is transitional to Level 1 because it is able to identify that just mentioning the measures of centre



are not enough to make a conclusion. R219M1 (Level 1) mentions spread in the form of outliers but does not link this to the conclusion about the means, while R114M1 is transitional because the description of the inter-quartile range suggests, but does not clarify, that variation is being considered to explain the mean differences. R202M1 (Level 2) refers specifically to the *overlay* of the boxes, hence using the variation to justify the conclusion about the weights. R213M1 is transitional because the amount of skew, although mentioned, is not directly related to the decision about the means. The only Level 3 response, R118M1, discusses predictability of the mean based on the amount of variance, indicating recognition of the significance of the role of variation in hypothesis testing.

Minute Paper 2: Probability

In an experiment we toss a coin six times and record Heads (H) or Tails (T) on each toss.

Which of the outcomes given below is the most likely? Explain your reasoning.

HTHTTH      TTTHHH      THTTTH

**Figure 2. Minute Paper 2 Question**

MP2 (See Figure 2) was given early in a series of lectures on probability after a discussion on the need for probability in statistics and the concept of randomness, including whether observed differences are “real” or “random”, but before discussion on the “Law of Large Numbers”, the likelihood of “runs” of heads or tails, and independence. The assessment focus here related to how well the response explains the effect of the factors (relative likelihoods of outcomes, the independence of trials, the randomness of outcomes, and the sample size) on the variation considered possible in trial outcomes. The key to providing a better response was to correctly interpret the various notions that affect the variation. Typical responses are presented in Table 4.

Continued on Next Page

Table 4  
Typical Responses for Minute Paper 2

<i>Level</i>	<i>Response</i>
0 No consideration of variation	Note: All responses at this level were coded as transitional. (R217M2 <sup>T</sup> ) I feel that the last outcome THTTTH would be the more likely outcome because people sometimes intrude some bias into small experiments by trying to toss the coin to land on what they want, or not toss it much, which could possibly be what we are seeing with the 3 tails all coming in one go. Just because there are only 2 subjects doesn't always mean that there is a 50/50 chance of an outcome, therefore having an extra head than tail or vice versa could possibly be more likely.
1 Weak consideration of variation	(R203M2) HTHTTH is the most likely outcome. The coin is tossed 6 times and each time Heads has a 1/2 chance of landing upright as does Tails. Tails - 3/6 chance all up Heads - 3/6 chance. In the outcome each comes up 3 times. The order of the above outcome also appears to be completely random. The second outcome is not likely as it is not random. (R212M2) I think the outcome is more likely to be HTHTTH. Because the likely chance of outcome 50/50 and it seem like a randomly result. Most of toss is difference from previous toss. But also we are considering the tossing skill of the tosser, the way of toss & catch. (R210M2 <sup>T</sup> ) Each toss has the same probability of getting heads or tails, a one out of 2 possible outcomes. All 3 of them have three head & three tail landings but the first choice is more common than the other two. Probability wise, they are all 3 even just because each toss is independent of each other and have a 50/50 chance of hitting either heads or tails. The first one then third, the second options are probably the ones seen.
2 Developing consideration of variation	(R213M2) (Student has drawn diagram example of coin toss.) HTHTTH - theoretically because there are only 2 probable outcomes, there is a 50/50 chance. It is possible though to flip heads 6 times or tails 6 times, although it is highly unlikely. The probability of getting TTTHHH is the least probable outcome, THTTTH or HTHTTH are more likely. TTTHHH is less probable because the results are less random. Hard to predict because it pure chance. (R114M2 <sup>T</sup> ) I believe that all 3 outcomes are just as likely as each other as with each toss, there is 50% chance of either heads or tails, the next toss has nothing compared with what occurred on any pervious tosses. So just because a pervious toss may have been a head, there is still just as much chance of a head or tail with the next toss. So out of 6 tosses the chance of one outcome is just the same as the chance of all other outcomes.
3 Strong consideration of variation	No responses were coded at this level.

"T" indicates a response transitional to the next level

All Level 0 responses were transitional, and one such response, R217M2, assumes the presented outcomes are possible and then questions the appropriateness of the theoretical proportions. At Level 1, R203M2 misapplies the 50:50 proportion notion to restrict the possibilities to exactly three heads and three tails, effectively reducing the amount of variation considered possible, and R212M2 misapplies the independence notion to the point where one toss must be different from the previous toss. The transitional response, R210M2, recognizes that the tosses are independent and allows more variation, but then makes a choice based on restricted variation and also incorrectly observes *all 3 of them have three head & three tails* so making it unclear whether the notion of the proportion being 50:50 has been

correctly applied. R213M2 (Level 2) correctly recognizes that the 50:50 proportion should not restrict the outcomes to exactly three heads and three tails in the short term but then misapplies the notion of randomness by distinguishing *a less random* selection, while R114M2 is transitional because the independence and randomization notions are applied correctly but the implications of the influence of the 50% proportion are not elaborated. No responses were coded as Level 3.

### Minute Paper 3: Sampling Distributions

Suppose a sample of  $n$  observations is taken from a population with mean  $E(X) = \mu$  and standard deviation  $sd(X) = \sigma$ . Explain what is meant by the sampling distribution of the sample mean,  $\bar{x}$ .

#### Figure 3. Minute Paper 3 Question

MP3 (see Figure 3) was given following a detailed examples-based discussion of sampling distributions including repeated sampling, sampling distribution of the mean, and the effect of sample size. The assessment focus here was on how well the response describes the effect of the variability on the sampling distribution and the factors affecting such variation. The key to a better response was recognizing that the variation produced by the action of taking the samples will be reduced if sample size is increased. Table 5 provides typical responses.

Table 5  
Typical Responses for Minute Paper 3

<i>Level</i>	<i>Response</i>
0 No consideration of variation	(R215M3) Mean $E(x) = \mu$ Standard deviation $sd(x) = \sigma$ The sampling distribution of the sample mean, $\bar{x}$ , means that the sample distribution of $\bar{x}$ is equal to the population mean. (R117M3 <sup>T</sup> ) Mean $E(x) = \mu$ $Sd(x) = \sigma$ The sampling distribution of the sampling mean, $\bar{x}$ , the expected number ( $\mu$ ) in the distribution is not biased.
1 Weak consideration of variation	(R113M3) You are taking a sample from a population finding the mean, $\bar{x}$ , for the sample and comparing it to the mean $\mu$ of the population, taking lots of samples of $\bar{x}$ gives us a closer estimate to $\mu$ . The mean of the sampling distribution of $\bar{x}$ is equal to $\mu$ . (R225M3 <sup>T</sup> ) Sampling distribution is the variability of a population which can be represented by $\mu - \sigma / \sqrt{n}$ and with the use of a (simple graph of a bell-shaped curve is drawn here).
2 Developing consideration of variation	(R103M3) I really struggled with what exactly the question was asking. The sampling distribution of the sample mean? The sampling distribution is the variability of the sample mean, which is inherent purely because the sample doesn't have every measurement in the population. There is always a chance that of a sample of 20 individuals in a population of 200 all the individuals sampled may be smaller than the true mean. (R207MP3 <sup>T</sup> ) The sampling distribution of the sample mean, $\bar{x}$ . If a number of samples of $n$ observations are taken from a population, each of these will have its own mean. If large enough number of samples are taken the mean of each should be seen to be centering around the true mean. Its distribution and variance can be given by $\sigma / \sqrt{n}$ .
3 Strong consideration of variation	(R210M3) The sampling mean, $\bar{x}$ , is the mean of multiple samples taken of $n$ observations. It is an overall $\bar{x}$ that better reflects the 'true' $\bar{x}$ of the population or entire samples. The $\sigma$ , $sd$ , can then be view on a population $n$ basis versus an individual sample basis by $\sigma / \sqrt{n}$ for $sd(\bar{x})$ . The $\bar{x}$ of multiple samples gives a better population mean with less variability and higher precision.

“T” indicates a response transitional to the next level

Typical of Level 0 responses is R215M3 which quotes  $sd(x) = \sigma$  but does not describe the effect of variability on the distribution. R117M3 is transitional as the phrase *not biased* suggests, but does not clarify, consideration of how the values of the sample means may have varied as sampling progressed. R113M3 (Level 1) indicates recognition of less variation in the resulting distribution by acknowledging that the estimate of the population mean will be better while R225M3 is transitional in that it tries to introduce the notion of the variation being reduced by a factor dependent on the sample size but gets the representation incorrect. R103M3 (Level 2) appreciates that there can be variability as the sample only contains some of the population and elaborates by giving an extreme example of what could happen, while R207M3 is transitional because although it acknowledges that the samples' means are spread around the true mean and that the *variance is  $\sigma/\sqrt{n}$* , it only implies, rather than explains, that the variation will be reduced by a factor dependent on the sample size. The Level 3 response, R210M3, not only states that the standard deviation of the means is  $\sigma/\sqrt{n}$ , like R207M3, but also relates *less variability and higher precision* in giving a population mean.

Minute Paper 4: Inferential Statistics

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.

Note: Diagram is from the course text (Wild & Seber, 2000, p. 439)

**Figure 4. Minute Paper 4 Question**

MP4 (see Figure 4) was given after a series of lectures on analysis of variance, including discussion of variation within and between samples, the concept of the F-ratio, and exposure to the diagram included in MP4. The assessment focus here is on how well a response considered the implications of variation in the justification of the inference made about the means. The key to providing a better response was the progression from simply comparing the spread of the data to realizing how the positioning of that variation, e.g. *either side of the mean* or *overlapping*, can influence the inference. Table 6 provides typical responses.

Table 6  
Typical Responses for Minute Paper 4

<i>Level</i>	<i>Response</i>
0 No consideration of variation	(R115M4) There seems to be no real differences in the group means. But example 3 shows more of a difference in the groups than the other examples. Group 1 has more of a difference in the means, then the other two but only a little bit, the means for 2 & 3 are around the same.
1 Weak consideration of variation	(R117M4) We can conclude that example 1 has a real difference than example 2 and 3. Since the data is spread more. Example 2 also shows a real difference in mean because the data is considerably spread compare to example 3. (R224M4 <sup>T</sup> ) In example 1, there may be a large difference in the group means, the data has a lot of spread, indicating high variability. Example 2 has less spread, and in 3 most of the data lie very close together. Even though the variability varies in each example, they are all unbiased, with approx. equal amounts of data lying on either side of the mean. Therefore none of the examples have a large diff. in group means.
2 Developing consideration of variation	(R207M4) In example 3 there is a real difference in means between groups 1 and 2 and 1 & 3. This can be seen by the lack of overlap of the results of the different groups. (R116M4 <sup>T</sup> ) Example 3 appears to be the only one with a real difference in group means as this is the only example in which the samples do not all overlap. Because of this it is likely that the spread of means is significant whereas for those that do overlap (1,2) the difference is more likely to be a byproduct of sampling difference.
3 Strong consideration of variation	(R210M4) A real difference would most likely occur in example 3. Because Ex.1 has the most variability (and standard deviation) the difference in means is not significant. The same with Ex. 2 but there is less variation, still difficult to tell extent of real difference. With Ex.3 though, such small standard deviation (variance) gives a clue that a real difference may exist between group 1 & 2, 1 & 3, but not a likely between 2 & 2 within Ex. 3 smaller the standard deviations, the more precision and smaller confidence intervals.

“T” indicates a response transitional to the next level

The Level 0 response R115M4 only refers to actual differences between means. R117M4 (Level 1) considers the spread of each of the examples as if the separate groups do not exist and there is one set of data, while R224M4 is transitional because variation has been used in the justification of the inference (considering the amount of data lying either side of the mean to conclude that the example is unbiased), but the degree of bias is not the focus of the inference and each group does not appear to have been considered independently. R207M4 (Level 2) identifies the *lack of overlap* in example three and concludes there are differences between the means without exploring why such a conclusion can be reached, while R116M4 is transitional in that it identifies that differences may be due to *sampling difference* thus allowing for the possibility that even though the sample means might be different the population means may not be. The only Level 3 response, R210M4, clearly indicates that the smaller standard deviation gives more precision in estimating the mean, hence acknowledging the influence of the amount of variation on precision.

### Overview

From the coding level comparison (see Table 7), it is evident that the proportions of responses across the four levels are similar for MP1, MP2 and MP4, with the majority of responses demonstrating weak or developing *consideration of variation*. The exception, MP3, has 50% of responses showing no *consideration of variation*. Interestingly, students' performance on MP4 is slightly stronger than for MP1, with a greater proportion coded as developing rather than weak. Both MP1 and MP4 are minute papers that focus on the measuring and modeling component.

Table 7  
Coding Level Comparison

<i>Level</i>	<i>MP1</i>	<i>MP2</i>	<i>MP3</i>	<i>MP4</i>
0	14% (4)	14% (4)	50% (8)	6% (1)
1	59% (17)	61% (17)	19% (3)	35% (6)
2	24% (7)	25% (7)	19% (3)	53% (9)
3	3% (1)	0% (0)	12% (2)	6% (1)
Total	100% (29)	100% (28)	100% (16)	100% (17)

Evaluations completed at the end of the course provided insights into student perceptions of the benefits of minute papers. While 93 percent of students agreed that their understanding of the concept of variation had improved since the beginning of semester, only 71 percent agreed that the minute papers had helped to clarify this understanding. Feedback from these students included that the minute papers helped to put to the test what was learnt in lectures; that they were an indication of whether the lectures had been understood; and that they should occur more frequently. Some problems identified were that some of the minute papers were hard to understand; that more feedback would have been appreciated; and that students were embarrassed by not knowing the answer to minute questions.

### **Discussion**

The process of threading the four components of *consideration of variation* through the various themes in the course and analyzing the minute paper responses for evidence of this *consideration of variation* has proved beneficial.

### Curriculum Mapping Process

During the mapping it became apparent that not all cells in [Table 1](#) would necessarily have an entry. Not unexpectedly, the bottom-left and top-right corners of the table were empty. The nature of the four components of *consideration of variation* suggest a hierarchical structure from left to right with investigative procedures only being possible after the other three components have been addressed. Similarly, the curriculum is based on a hierarchy of themes from exploratory data analysis to inferential statistics, with some themes being revisited during the presentation of the course. Consequently, for example, what is involved with inferential statistics is more complex than mere exploratory data analysis. However, the presence of an empty cell on or near the leading diagonal, from top left to bottom right of the table, may indicate a potential gap in the curriculum. Addressing such an occurrence is part of the cyclic approach to developing an integrated curriculum.

### Minute Papers as Evidence of Consideration of Variation

Some minute papers proved more useful than others in terms of eliciting information about students' consideration of variation. A number of issues relating to the creation and interpretation of the minute papers became apparent when developing the hierarchy and coding the responses. The wording of the minute paper questions has repercussions for the quality of the response. Responses to MP1 suggest that the question was too restrictive due to data being given in a summarized form. A question presenting data in a less summarized form, for example as dot plots, may provide responses with more consideration of variation. This appears to have occurred in MP4. Responses to MP3 suggest that the closed nature of the question was not very useful because students could simply reproduce a definition without demonstrating their consideration of variation in relation to the sampling distribution. Complexity of the task also proved to be an issue. Responses to MP4 suggested that comparing variation both between and within samples was more challenging than was anticipated. The increase from two samples, as in MP1, to three and then the further complication of being given three different examples to consider rather than just one, caused some students to oversimplify the situation and consider all three groups as one large sample. These issues are discussed in more detail in Reid and Reading (2004).

Interpretation of the responses is often difficult due to literacy issues and lack of sophistication in the explanation and application of notions. Some MP4 responses, in particular, flagged the literacy issue. For example, one response talks about the means falling *in around one and other*, which may be taken to suggest that they are close together. However, more may have been meant by this expression, as further on in the same response another example is discussed where there is a little less spread and the means fall *within each other*. This may be considering the variation as well as the position of the mean. The lack of clear English expression prevents such a determination and the coding level can only be based on what can be best interpreted from the response. In a similar vein, in MP2, immaturity in students' understanding of probability, independence, and randomness makes it difficult to distinguish between notions when coding for evidence of correct application. For example, insistence that one trial be different from the previous could be misapplication of the notions of independence or randomness, or even probability (i.e.,  $P(H) = 0.5$ , so H cannot occur on the next toss). The fact that MP2 was given at the beginning of the probability teaching sequence may have contributed to this finding.

Not only did students' responses suggest that they had performed better in the final minute paper, MP4, but students also indicated in their evaluations that they felt that their understanding of variation had improved by the end of the course. However, it should be noted that the improved performance in MP4 could be due to one or more of the following:

- attendance decreased later in the course and it is likely that those attending were the more attentive students;
- students had by then completed most of the course where variation was the core concept;
- students had already been exposed to the diagram a week prior to completing MP4; and
- the data in MP4 were presented in a less summarized form, unlike in MP1.

Interestingly, responses indicated a tendency for students to focus on the extraordinary, such as the large gap between the median and third quartile for Perch in MP1. With the research focus on variation it was pleasing to note that many responses did not just automatically use the average as the measure for comparison when asked to compare the differences in weights in MP1. An important consideration that arose with the MP2 responses is the meaning of variation. There is variation in what is possible between the various sequences of outcomes, with students incorrectly reducing variation by misapplying notions. In addition, there is the variation that students view as the difference between what is shown as actually happening and what they expected would happen, rather than what theory dictates. For example, some responses clearly indicate what has been given in the question is different from what they would expect, being: *a similar percentage*, in one case; and *3 heads and 3 tails* in another. It should

be noted that the concept of “expected values” was not discussed in the course until after MP2 was completed.

### *Evaluating the Integrated Curriculum*

The ways in which students considered variation in their responses has provided useful information for evaluating how successfully the consideration of variation thread has been used to structure the integrated curriculum. Specifically, the use of minute papers focuses on students’ understanding and reasoning rather than merely their ability to perform calculations. What students chose to discuss in the responses differed, in some instances, from what the lecturer anticipated. For example, in MP3 a large proportion of students failed to mention variation at all, despite the fact that variation is critical in any discussion of distributions. It may be that the closed nature of the question restricted such discussions. However, it may also be possible that the activities incorporated into the sampling distributions theme have not adequately conveyed the concept of variation to the students. Consequently, to better thread consideration of variation through that theme, the lecturer needs to rethink the activities and the way they are presented.

### ***Implications for Teaching and Research***

As consideration of variation is a key concept in statistical thinking, students’ success in introductory statistics courses will depend on how well they incorporate this into the various statistical procedures and inferences that are studied. The model for an integrated approach with consideration of variation as a thread provides a means for course presenters to review their curriculum. In addition, the hierarchy could assist those evaluating such an approach by providing a means for assessing the consideration of variation in student responses.

Further iterations of the cyclic approach to the mapping, as begun in [Table 1](#), could indicate components of variation that are not being covered adequately in the curriculum and assessment. When evaluating this approach, a benefit of this hierarchy is that special assessment items do not have to be developed, as the coding levels can be applied to responses to tasks that students are already undertaking as part of the course. However, the information gained from the assessment of student responses helps to identify specific assessment tasks that provide useful information about consideration of variation. Those tasks, and associated learning strategies and course objectives, should be encouraged in future curriculum development. Carefully worded minute papers have proven to be useful, and their use should be encouraged. By applying minute papers before and after the coverage of each theme in the curriculum it would be possible to examine any progress in students’ understanding of variation within themes as well as throughout the course, thus identifying specific themes or topics where more attention to variation is required.

This integrated approach to curriculum development and the hierarchy for assessing consideration of variation would be cost-effective in all socio-economic settings. Resources required are the time and expertise to review the curriculum, plan suitable assessment tasks and learning activities, and analyze the student responses. This exploratory study also provides an important basis for further research into assessing student performance in statistical thinking.

Early indications are that researchers can use evidence from already established assessment tasks, such as minute papers, to assess key statistical thinking. Reports on the analysis of responses to other tasks (pre-study and post-study questionnaires, follow-up interviews, class tests and assignments) completed by these students as part of the larger *Understanding of Variation* project, will help to further inform the proposed hierarchy for consideration of variation. Similar work on *consideration of variation* at the secondary level could help to inform the “appreciation of variation” component in the Statistical Literacy Construct described by Watson and Callingham (2005). Based on the findings of this and other studies, future research plans include using a larger cohort of students studying within the sciences,



following one cohort of students longitudinally through their tertiary study of statistics, and considering cohorts of students in other disciplines.

The *Understanding of Variation* project, and in particular the *Consideration of Variation* hierarchy, aim to help students make links so that they move from a fragmented view of statistics as an unrelated list of techniques, towards a more holistic view where they can appreciate statistics as a way of understanding real-life situations using a variety of statistical models as advocated by Reid and Petocz (2002). Developing a holistic view promotes the development of, not only statistical literacy and the ability to reason statistically, but the higher-order ability to actually think statistically.

## **Acknowledgement**

This research was funded by a University of New England, Faculty of The Sciences, Internal Research Grant.

## **References**

- Ballman, K. (1997). Greater emphasis on variation in an introductory Statistics course. *Journal of Statistics Education*, 5(2).
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3).
- delMas, B. (2002). Statistical literacy, reasoning and learning: A commentary. *Journal of Statistics Education*, 10(3).
- delMas, R.; Garfield, J. & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3).
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3).
- Garfield, J. & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67(1): 1-12.
- Hoerl, R.; Hahn, G. & Doganaksoy, N. (1997). Discussion: Let's stop squandering our most strategic weapon. *International Statistical Review*, 65(2): 147-153.
- MacGillivray, M. (2005). Coherent and purposeful development in statistics across the education spectrum. In G. Burrill and M. Camden (Eds.), *Curriculum Development in Statistics Education: International Association for Statistics Education 2004 Roundtable*. Voorberg, the Netherlands: International Statistics Institute.
- Meletiou-Mavrotheris, M. & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal*, 1(2): 22-37.
- Moore, D. S. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review*, 65(2): 123-137.
- Murdoch, K. & Hornsby, D. (1997). Planning curriculum connections: Whole-school planning for integrated curriculum. Armadale, Australia: Eleanor Curtain Publishing.
- Nicholls, A. & Nicholls, H. (1981). *Developing a curriculum: A practical approach*. George Allen & Unwin Ltd.
- Petocz, P. & Reid, P. (2003). Relationships between students' experiences of learning statistics and teaching statistics. *Statistics Education Research Journal*, 2(1): 39-53.
- Pfannkuch, M. and Horring, J. (2005). Developing statistical thinking in a secondary girls school: A collaborative curriculum development. In G. Burrill and M. Camden (Eds.), *Curriculum Development in Statistics Education: International Association for Statistics Education 2004 Roundtable*. Voorberg, the Netherlands: International Statistics Institute.

- Pfannkuch, M.; Rubick, A. & Yoon, C. (2002). Statistical thinking: An exploration into students' variation-type thinking. *New England Mathematics Journal*, 34(2): 82-98.
- Pidgon, K. & Woolley, W. (Eds.) (1992). *The BIG picture*. Armadale, Australia: Eleanor Curtin Publishing.
- 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). Dordrech, The Netherlands: Kluwer Academic Publishers.
- Reid, A. & Petocz, P. (2002). Students' conceptions of statistics: A phenomenographic study. *Journal of Statistics Education*, 10(2).
- 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.
- Ridgway, J.; McCusker, S. & Nicholson, J. (2005). Uncovering and developing student statistical competencies via new interfaces. In G. Burrill and M. Camden (Eds.), *Curriculum Development in Statistics Education: International Association for Statistics Education 2004 Roundtable*. Voorberg, the Netherlands: International Statistics Institute.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3).
- Scheaffer, R. L. (2000). Statistics for a new century. In M. J. Burke & F. R. Curcio (Eds.), *Learning Mathematics for a New Century* (pp.158-173). Reston, VA: National Council of Teachers of Mathematics.
- Stromberg, A. J. & Ramanthaman, S. (1996). Easy implementation of writing in introductory statistics courses. *The American Statistician*, 50(2): 159-163.
- 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. & Callingham, R. (2005). Statistical literacy: From idiosyncratic to critical thinking. In G. Burrill and M. Camden (Eds.), *Curriculum Development in Statistics Education: International Association for Statistics Education 2004 Roundtable*. Voorberg, the Netherlands: International Statistics Institute.
- 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): 223-265.
- Wild, C. J. & Seber, C. A. (2000). *Chance encounters: A first course in data analysis and inference*. New York: John Wiley & Sons.