Statistics Education Research Journal

Volume 5 Number 2 November 2006

International Association for Statistical Education
http://www.stat.auckland.ac.nz/~iase

International Statistical Institute
http://isi.cbs.nl/

Statistics Education Research Journal

The Statistics Education Research Journal (SERJ) is a peer-reviewed electronic journal of the International Association for Statistical Education (IASE) and the International Statistical Institute (ISI). SERJ is published twice a year and is free.

SERJ aims to advance research-based knowledge that can help to improve the teaching, learning, and understanding of statistics or probability at all educational levels and in both formal (classroom-based) and informal (out-of-classroom) contexts. Such research may examine, for example, cognitive, motivational, attitudinal, curricular, teaching-related, technology-related, organizational, or societal factors and processes that are related to the development and understanding of stochastic knowledge. In addition, research may focus on how people use or apply statistical and probabilistic information and ideas, broadly viewed.

The Journal encourages the submission of quality papers related to the above goals, such as reports of original research (both quantitative and qualitative), integrative and critical reviews of research literature, analyses of research-based theoretical and methodological models, and other types of papers described in full in the Guidelines for Authors. All papers are reviewed internally by an Associate Editor or Editor, and are blind-reviewed by at least two external referees. Contributions in English are recommended. Contributions in French and Spanish will also be considered. A submitted paper must not have been published before or be under consideration for publication elsewhere.

Further information and guidelines for authors are available at: http://www.stat.auckland.ac.nz/serj

Submissions

Manuscripts must be submitted by email, as an attached Word document, to co-editor Tom Short <tshort@iup.edu>. Submitted manuscripts should be produced using the Template file and in accordance with details in the Guidelines for Authors on the Journal's Web page: http://www.stat.auckland.ac.nz/serj

Publication: IASE/ISI, Voorburg, The Netherlands
Technical Production: California Polytechnic State University, San Luis Obispo, California, United States of America
Web hosting and technical support: Department of Statistics, University of Auckland, New Zealand

ISSN: 1570-1824

SERJ EDITORIAL BOARD

TABLE OF CONTENTS

EDITORIAL

This issue is special in two ways, both because it is a special issue focused on research on reasoning about distribution, and because it marks the end of SERJ's fifth year of operation. This editorial elaborates both on the topic chosen for this special issue, as well as on SERJ's status at this point in time and thinking ahead.

We are pleased to offer this special issue, which is based on SERJ's collaboration with SRTL – the biannual international research forum on Statistical Reasoning, Thinking, and Literacy. Two years ago, based on papers discussed at SRTL-3, we published a Special Issue on reasoning about variation and variability, with guest editors Joan Garfield and Dani Ben-Zvi, who have been organizing SRTL since 1998. The current issue includes several papers related to reasoning about distribution, the topic chosen for SRTL-4 which took place in 2005. We thank Maxine Pfannkuch and Chris Reading for serving as Guest Editors for this issue.

The papers in this issue highlight the centrality of distribution as a core construct and a set of interactive tools which learners of statistics, or adults having to make sense of statistical information in the world, have to cope with or relate to at various stages and levels of learning or working with statistics. Think of just three situations where notions of distribution come up: dealing with visual representations of counts of data in basic charts and graphs, making predictions about the results of using random generating devices such as dice based on results of prior events, and having to grasp second-order abstractions embodied in sampling distributions. In these and many other contexts, students at all levels of learning, and thus their teachers, find themselves continuously engaged in generating, interpreting, communicating about, and using distributions and thinking about variability in data. While seemingly simple, distributions present many challenges to teachers and researchers alike, because of the multiplicity of areas within the domains of statistics and probability where they appear, and their changing level of complexity in different contexts of learning or using statistics.

The research papers in this issue explore in depth some of the many issues associated with making sense of and reasoning about distribution, and in so doing also illustrate that a range of methodological approaches are needed to study learning and reasoning processes in this area, both qualitative and quantitative. We hope to see additional papers on reasoning both about distribution and variation in upcoming issues of SERJ, so as to add to the cumulative knowledge base on these foundational areas of statistical knowledge.

Over the last five years, the journal has developed substantially and is attracting a growing number of both readers and authors. The editorial board includes experts from 11 geographically dispersed countries. Papers being submitted represent work being carried out in many places around the world, and the flow of manuscripts is growing. In the 12-months period October, 2004 through October, 2005, we received 22 manuscripts. Of these, eight were found not suitable for SERJ and were not refereed; five had potential but needed revision even before refereeing; three were rejected after an external review, and six were rewritten and resubmitted for further review. Only five of these 22 manuscripts have been published by now. In contrast, during the most recent calendar year the number of submitted papers has almost doubled. This can be attributed both to the growing centrality and recognition of the journal, as well as to the fact that more research is being carried out on statistics education.

Statistics Education Research Journal, 5(2), 2-3, http://www.stat.auckland.ac.nz/serj

As the journal continues to grow, it faces many challenges. SERJ operates in an emerging area which is of interest to diverse scholars and practitioners, and to both new and established researchers. Our submitting authors and our external referees, come from diverse disciplines with somewhat different traditions, such as statistics, education, psychology, natural sciences, medicine, business, engineering, and others. Given the international nature of SERJ, we have to be aware of diversity and accommodate variations in aspects of scientific reporting and academic writing. As a research and practice community we will need to seek ways to maintain high standards and convey high expectations for quality both to authors, referees, and researchers alike.

Towards helping the maturation of the field of statistics education research, SERJ has initiated several activities in the last two years. At the 55th meeting of the International Statistics Institute in Sydney (2005), SERJ arranged a workshop for prospective authors designed to educate about writing high-quality research papers. At the 7th International Conference on Teaching Statistics held in summer 2006 in Salvador, Brazil, SERJ arranged two workshop, one for prospective authors similar to that held at ISI, and a new innovative one for referees on writing good reviews. The feedback from participants who attended these workshops suggests that they can help both new and continuing researchers interested in the emerging field of statistics education research, and we thus will aim to continue with such workshops in coming years.

Looking forward into the next few years, we hope to see a growth in the range of topics addressed by research, and in the range of the methodologies employed. Many areas in statistics education require more research, such as learning and reasoning about inferential statistics, correlations and associations, probabilistic reasoning, or the understanding of statistics encountered in everyday contexts or in official statistical publications, to name just a few. Exploratory research on these and other core areas will surely require the continued use in years to come of diverse types of qualitative designs and descriptive quantitative designs. Yet, as time goes by we do hope to see a growing number of studies employing experimental and comparative designs which can provide evidence in line with the growing expectations that educational research provides solid information about the relative efficacy of different interventions or teaching methods.

In closing, we want to thank our many referees, whose challenging role is to help the journal maintain high scholarly standards. SERJ serves a diverse and expanding community of practitioners and researchers interested in statistics education and learning in diverse fields and contexts. We encourage SERJ readers to send us reactions and ideas regarding the journal, its scope, papers it publishes, and possible future plans.

IDDO GAL AND TOM SHORT

REASONING ABOUT DISTRIBUTION: A COMPLEX PROCESS

GUEST EDITORS:

MAXINE PFANNKUCH
The University of Auckland, New Zealand
m.pfannkuch@auckland.ac.nz

CHRIS READING
SiMERR National Centre, University of New England, Australia
creading@une.edu.au

1. THE NATURE OF DISTRIBUTIONS

We are very pleased to introduce this special issue of the *Statistics Education Research Journal* (*SERJ*) on *Reasoning about Distribution,* which presents research at the forefront of building conceptual foundations for statistics education. According to Moore (1990, p. 136) statistical thinking is an "independent and fundamental intellectual method that deserves attention in the school curriculum." Equally he could have stated that statistical thinking deserves attention by research. He also hoped that "in the future pupils will bring away from their schooling a structure of thought that whispers, 'Variation matters … Why not draw a graph?'" (Moore, 1991, p. 426). With considerable foresight Moore not only encapsulated the building blocks for statistical thinking but also two deep research questions with which statistics education researchers are currently grappling: How do students actually reason about variability and distribution? How do these two types of reasoning develop?

Variation is at the heart of statistical thinking but the reasoning about variation is enabled through diagrams or displays that "represent intuitively the original reality via an intervening conceptual structure" (Fischbein, 1987, p. 165), such as graphs or frequency distributions of data. The conceptualization of variation "through a lens, which is 'distribution'" (Wild, 2005) was originally fostered by Quetelet in the 1840s (Porter, 1986). Connecting variation in nature to distribution structures was a major conceptual obstacle in the history of statistics. It was not until the end of the 19th Century that the astronomers' error curve was re-conceptualized as a distribution governing variation in social data. According to Bakker and Gravemeijer (2004) distribution is the conceptual entity for thinking about variability in data. Therefore a discussion about the nature of distributions involves both conceptual and operational aspects to be considered. A conceptual perspective focuses on clarifying what notions underpin distributions and why these notions are important whereas an operational perspective focuses on how a specific set of data is captured, displayed and manipulated by distributions. Reasoning about distributions involves interpreting a complex structure that not only includes reasoning about features such as centre, spread, density, skewness, and outliers but also involves other ideas such as sampling, population, causality and chance. These other ideas lead towards connecting empirical data with probabilistic notions, which in turn develop cognizance of empirical and theoretical distributions. In fact Bakker and Gravemeijer (2004), in the context of data analysis, believe that focusing on distribution might bring

Statistics Education Research Journal, 5(2), 4-9, http://www.stat.auckland.ac.nz/serj

more coherence to the statistics curriculum. Similarly, Scheaffer, Watkins, and Landwehr (1998, p. 17) considered that “the unifying thread throughout the probability curriculum should be the idea of distribution.”

It would seem that distribution provides a strong connection between statistics and probability, a connection that is currently lacking in curricula and teaching. Distribution is a key concept in statistics yet statisticians and educators may not be aware of how difficult it is for students to develop a deep conceptual and operational understanding of distributional structures. When students are given tasks involving comparing distributions or making inferences, they often fail to utilize relevant information contained in the underlying distributions. Curricular materials often focus on construction and identification of distributions, but not on the meaning and interpretation of these distributions or on how to manipulate them to derive further information from the data. Different distributions of the same data require students not only to understand how their structures are connected but also how these different distributions may unlock different parts of the story of the same dataset.

Thus, distributions are conceptual organizing structures or mental devices that allow for a statistical intellectual method to develop. These structures are complex and subtle and require a long enculturation into understanding them. Many questions arise about conceptual, pedagogical, and research-related aspects of reasoning about distributions.

Some questions that need to be addressed by research are:

- What does distribution mean to students?
- What are the simplest forms and representations of distributions that children can understand?
- When and how do children begin to develop the idea of distribution?
- How does reasoning about distribution develop from the simplest aspects or forms of distribution to the more complex ones?
- What type of understanding of distribution is sufficient for a statistically literate person?
- What instructional tasks and technological tools can promote the understanding of distribution?
- What are the common misconceptions involved in reasoning about distribution?
- What are the difficulties that students encounter when working with, analyzing and interpreting distributions?
- How does an understanding of distribution connect and affect understanding of other statistical concepts and how does it relate to other kinds of statistical reasoning (e.g., reasoning about variation, covariation, inference)?
- What methods can be used to assess understanding of distribution?
- What are useful methodologies for studying (researching) the understanding of distribution?

2. ABOUT THIS SPECIAL ISSUE

Since reasoning about distribution is a complex and challenging research topic, this special issue presents a series of papers which address some of the questions posed above. This special issue arose from the fourth international research forum on Statistical Reasoning, Thinking and Literacy (SRTL-4) and from a subsequent call from SERJ for other researchers to submit papers on this topic. After considering “reasoning about variability” in the third forum (SRTL-3), the fourth forum (SRTL-4) held in July 2005 at The University of Auckland, New Zealand built on the core idea of variation by focusing on "reasoning about distribution." These SRTL forums bring together a small number of researchers whose work is focused on a particular area and presented in extended sessions that permit lengthy discussions among the participants. In addition, many researchers present primary data in the form of video clips and transcripts of students or teachers in the process of reasoning as well as discussing and explaining their actions; this allows for intensive review and discussion of findings and research methods by all participants. At SRTL-4 twenty researchers in statistics education, from six countries, discussed eight studies that examined different aspects of reasoning about distribution (Makar, 2005). After five days of presentations and discussion the participants believed that they were only in the initial stages of understanding reasoning about distribution but felt that as a community they were getting closer to important breakthroughs. The papers in this special issue represent some of the many efforts now underway to deepen our knowledge and respond to some of the challenging research questions listed earlier.

The first paper in this Special Issue, by *Wild*, a well-known statistician, is based on his opening address at SRTL-4 and is designed to delve deeply into issues of distributional reasoning and its purpose from a statistician's perspective. He presents distribution as a lens through which variation is viewed and then discusses the conundrums of connecting empirical distributions to theoretical distributions, the position of the sampling distribution, and why all distributions are conditional. The paper by *Pfannkuch* proposes a model for reasoning from the comparison of box plots based on one secondary teacher's articulation of these comparisons whilst teaching. This model is intended as a guide for teacher reasoning and to inform the design of teaching sequences. The paper by *Reading and Reid* describes levels of reasoning about distribution based on the SOLO taxonomy that could be used to assess students' development of such reasoning ability and to structure learning sequences. This hierarchy emerged from the re-analysis of tertiary students' responses that had shown various levels of reasoning about variation. The paper by *Prodromou and Pratt* reports on a virtual simulation designed to allow students to use causality to articulate features of distribution. This latest iteration of software under development acts as a "window on thinking-in-change" by allowing the students to explore the relationship between causality and variation. The paper by *Leavy* reports on the developing understanding of distribution as elementary pre-service teachers compared distributions of data that were created during practical investigations. The use of the experimental context in this study was found to support the construction of a distributional perspective.

3. EMERGING KEY THEMES

There are four themes common across these papers that have important implications for future statistics education research. These relate to research purpose, educational context, methodology and the importance of variation. First, education research is evolving to have a more cognitive focus. The purposes of the various research studies undertaken were to either describe the reasoning about distributions (Pfannkuch; Reading & Reid) or investigate ways to assist students to develop such reasoning (Prodromou & Pratt; Leavy). In unpacking the concept of distribution, Pfannkuch's key elements of reasoning help to position "distribution" within the wider "inference" context, while Reading and Reid's "understanding" and "using" cycles provide a cognitive developmental framework for assessing the concept. In assisting development of the concept, Prodromou and Pratt are improving a microworld to assist students in co-ordinating different perspectives of distribution, while ways of building on existing notions, as identified by Leavy, provide a foundation for creating richer learning environments. Future research must continue to address both the assessment of reasoning about distribution, and ways of supporting the development of this reasoning.

Second, the educational context in which research is positioned is becoming increasingly important for generating meaningful qualitative data. In all four studies the concept of distribution was investigated by the researchers during learning activities and involved comparison of datasets. Increasingly research data are being collected during actual teaching/learning episodes, rather than with participants who have been withdrawn from their classes or given artificial tasks as part of a research project. This is reflected in all but one of the studies, and even then Prodromou and Pratt worked with students who were involved in a learning situation, although it was outside class time. The comparison of datasets was either explicit, as the actual task given, or implicit, as a necessary action to achieve a more general task. These studies have demonstrated that rich environments are available for collecting qualitative research data on reasoning about distribution, when learners are allowed to explore their own meanings for distribution and the necessary related reasoning process.

Third, the analysis of qualitative data is proving to be a rich source of information for investigating reasoning about concepts. The methodologies employed in all four studies reflect this recent research trend. In each case the researcher(s) analysed qualitative data based on episodes or responses that were produced during learning activities, from either the teacher's (Pfannkuch; Leavy) or the student's (Reading & Reid; Prodromou & Pratt) perspective. This often necessitates smaller sample sizes to achieve the depth of analysis desired, with the implication that findings are more in-depth but sometimes more exploratory in nature. Frameworks provided by Pfannkuch and by Reading and Reid are valuable stepping-stones to more detailed assessment of students' reasoning. The "thinking-in-change" investigated by Prodromou and Pratt provides a particularly interesting approach to the analysis of student thinking in action and could profitably be pursued by future researchers.

Finally, variation is a recurring concept in each paper. The underlying importance of variation is demonstrated in its role in the various descriptions of models, frameworks and understandings of distribution, and in supporting key decisions when adjusting distributions. Variation was acknowledged as one of the key elements in being able to reason about distributions (Pfannkuch) and, as such, it was used as an initial variable for identifying better quality student responses before searching for what constituted weaker and stronger reasoning about distribution (Reading and Reid). Increase in awareness of variation (Leary) and co-ordination of two different perspectives of variation (Prodromou & Pratt) were both found to be important in supporting the development of the concept of distribution. Thus all four studies reinforce the now generally accepted linking of the concepts of variation and distribution. Future studies of either concept should not preclude the other.

The juxtaposition of these four studies also raises a question about the connection between teaching methods and students' reasoning about distribution. Starting with students' tendency to think deterministically, Prodromou and Pratt develop a novel way of building up students' concepts of distribution. Their teaching strategy raises questions about the other research. For example, if new teaching methods different from the current practice are used by teachers and researchers: Will Reading and Reid's hierarchical model change? Will Leavy's students' reasoning show the same misconceptions? Will Pfannkuch's teacher have the same problems with her reasoning? Conversely: Will Prodromou and Pratt's method give rise to new student misconceptions? To improve teaching, future research needs to approach the problematic issue of reasoning about distribution from many different perspectives.

4. CLOSING THOUGHTS

Together these studies have provided an insight into research methodology for investigating the reasoning process, as well as detailed knowledge and frameworks on which to base investigations into the concept of distribution. The focus thus far on qualitative studies to inform exploratory research in the area of reasoning about distribution has been enlightening, but given the limitations of qualitative research, researchers now need to develop quantitative studies to substantiate the wide range of findings being espoused. At the same time it is worth recognizing that there is a noticeable trend to investigate the cognition of teachers and those training to become teachers, as well as their students. These papers suggest that such research would be profitable for the development of statistics education. Wild's paper on the concept of distribution also points to many research avenues that need to be explored and thought about by researchers.

Importantly, researchers need to expand on this useful research work on reasoning about distribution. We hope to see further papers on reasoning about distribution and related issues such as variation appearing in future issues of *SERJ*. We challenge researchers to determine when the first notions of distribution begin to develop for students and how they extend into an understanding of more complex forms. Integral to this is the need to determine how the understanding of distribution connects to and affects understanding of other statistical concepts and related statistical reasoning. In particular, statistics educators are interested in knowing about common misconceptions held, and difficulties encountered, by students when reasoning about distribution and which instructional tasks and technological tools promote a better understanding of distribution. In particular, there is a lack of research with post-secondary and college level learners, who encounter distribution and variation in a more formal context of learning about statistics, that needs to be addressed. Underlying all this work, researchers should continually strive to identify useful methodologies for studying student understanding of distribution. By responding to these challenges the statistics education research community will enrich the available knowledge relating to reasoning about distribution and thus assist statistics educators to improve the quality of learning about fundamental statistical concepts.

We appreciate the opportunity to collate and devote a set of research papers to reasoning about distribution. Especially, we value the contribution of the coeditor, Iddo Gal (University of Haifa, Israel), who co-ordinated this special issue and offered many suggestions to improve the quality of the papers. Special thanks also go to all SRTL-4 participants who contributed to the research forum discussions of earlier versions of some of the papers and to those researchers who contributed as reviewers of all papers. Readers are now invited to comment or make suggestions by contacting the authors. Finally, all researchers are invited to consider contributing to the forthcoming SRTL-5 (see "Forthcoming Conferences" in this issue), to be held in 2007 in England, which will be devoted to *Reasoning about Statistical Inference: Innovative Ways of Connecting Chance and Data.*

REFERENCES

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.

Fischbein, E. (1987). *Intuition in Science and Mathematics*. Dordrecht, The Netherlands: Reidel.

Makar, K. (Ed.). (2005). *Reasoning about distribution: A collection of research studies. The Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

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

Moore, D. (1991). Statistics for all: Why? What and how? In D. Vere-Jones (Ed.) *Proceedings of the Third International Conference on Teaching Statistics (ICOTS-3)*, Dunedin, August 1990, Vol. 1 (pp. 423-428). Voorburg, The Netherlands: International Statistical Institute.

Porter, T. M. (1986). *The rise of statistical thinking 1820-1900.* Princeton, NJ: Princeton University Press.

Scheaffer, R., Watkins, A., & Landwehr, J. (1998). What every high-school graduate should know about statistics. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K-12* (pp. 3-31). Mahwah, NJ: Lawrence Erlbaum Associates.

Wild, C. J. (2005). The concept of distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. The Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy,* Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

MAXINE PFANNKUCH
Department of Statistics
The University of Auckland
Private Bag 92019
Auckland, New Zealand

CHRIS READING
National Centre of Science, Information and Communication Technology and
Mathematics for Rural and Regional Australia
Education Building
University of New England
Armidale, NSW 2351, Australia

THE CONCEPT OF DISTRIBUTION

CHRIS WILD
The University of Auckland, New Zealand
c.wild@auckland.ac.nz

ABSTRACT

This paper is a personal exploration of where the ideas of "distribution" that we are trying to develop in students come from and are leading to, how they fit together, and where they are important and why. We need to have such considerations in the back of our minds when designing learning experiences. The notion of "distribution" as a lens through which statisticians look at the variation in data is developed. I explore the sources of variation in data, empirical versus theoretical distributions, the nature of statistical models, sampling distributions, the conditional nature of distributions used for modelling, and the underpinnings of inference.

Keywords: *Frequency distributions, Statistical models; Sampling distributions; Statistical inference; Types of distribution; Variation*

1. INTRODUCTION

There are aspects of statistics that are so basic to the way we think in the subject that no one abstracts, enunciates and examines them. We encountered this phenomenon frequently in conducting the research for Wild and Pfannkuch (1999). It is not a problem for the statistical practice of professionals since they have long since been successfully encultured into these ways of thinking. It may well, however, be a root cause of some of the problems we face in statistics education. "Variation" was one of these unenunciated givens until quite recently and still is for many communities of statisticians. "Distribution" is another fundamental given of statistical reasoning. I can find a great deal written about specialized usages and definitions of "distribution" but almost nothing about "distribution" itself as an underlying conceptual structure. For example, Wiley's massive 16 volume *Encyclopedia of Statistical Sciences* does not contain an entry for "distribution" as an entity although it contains over 300 different sections in which "distribution" appears in the title.

The main aim of this paper is to try to explore for teachers and statistics education researchers where the ideas of "distribution" that we are trying to develop in students are leading to, and where they are important and why. We need to have such considerations in the back of our minds when designing learning experiences. They are a logical precursor for a planned educational development; a platform upon which the educational "when?," "in what order?," "by what means?" and so on, can be built. Our journey towards an understanding of "distribution," and the need for concepts of distribution, begins with the pervasive nature of variation.

Section 3 of Wild and Pfannkuch (1999) was entitled "Variation, randomness and statistical models." The genesis of that story was, "In the beginning was variation." Variation is an observed reality detectable in all systems and entities. It is, in a word, omnipresent. A statistical response is generated when the variation we have to deal with

Statistics Education Research Journal, 5(2), 10-26, http://www.stat.auckland.ac.nz/serj

in pursuing a real-world goal is not completely predictable at levels of precision that are of practical importance and we have given up, at least temporarily, on the ability to understand differences between individuals at a level that might make them predictable. The statistical response is to investigate, disentangle and model patterns of variation in order to learn from them. We will see that the notion of "distribution" is, at its most basic, intuitive level, "the pattern of variation in a variable," or set of variables in the multivariate case. Thus the notion of "distribution" underlies virtually all statistical ways of reasoning about variation. So it is particularly fitting that the first special section of the *Statistics Education Research Journal* (Garfield and Ben-Zvi, 2005) had the theme of "variation" and this, the second special issue, has the theme of "distribution."

Statisticians look at variation through a lens which is "distribution" (Figure 1). Provided "variation" is in the background of our thinking, we are looking through the "distribution" lens as soon as we look at our data in any way that sets aside case labels. Setting aside case labels is no small matter, however. There has been a good deal of work about how children at elementary and middle school levels relate to data. Bakker and Gravemeijer (2004, p. 147) write that such students "tend to conceive a dataset as a collection of individual values instead of an aggregate that has certain properties." What is important and interesting to children is the particular. Case labels (e.g., the names of people) inform us that a particular data record describes a specific entity, often a person. It is a very big step indeed from this to thinking about data in aggregate terms.

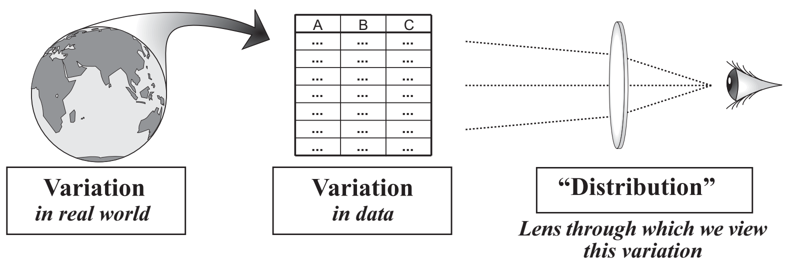


Figure 1. "Distribution" as a lens

In statistics we are seldom interested in a dataset as a collection of separate snapshots of particular individuals taken in a particular way at a particular instance in time. Rather we look at data to learn more widely applicable lessons. These lessons are not, we believe, to be found in the individual data points themselves but in patterns discernible in the dataset as a whole. So we put aside (temporarily ignore) the links between data points and individuals as distracting detail in order to better focus on patterns. When case labels are set aside individuals with identical values for the variables of interest become indistinguishable so that, without any loss of information, we can reduce the data to a set of distinct values and their corresponding frequencies, that is, to a frequency distribution. All of the information about patterns of variation is in the (typically multivariate) frequency distributions. All summary statistics and almost all the graphs we look at are summaries and graphs of frequency distributions. We use them to discover and describe aspects of the patterns in the variation contained in the frequency distributions. We convert frequency distributions into relative-frequency distributions to facilitate the comparison of batches of data (e.g., to compare data from different subgroups) containing different numbers of observations.

Where does the variation we see in data come from? There is typically real variation in the systems we are investigating and this is inevitably overlaid with additional variation induced by the observational process as in Figure 2. Why do we summarise and model patterns of variation? Primarily we do it for the purposes of prediction, explanation or control; that is, in order to be able to make better predictions, better understand the mechanisms generating the data, or to enable us to change the pattern of variation in the system in the future, at least to some partial but useful extent such as by reducing mortality rates.

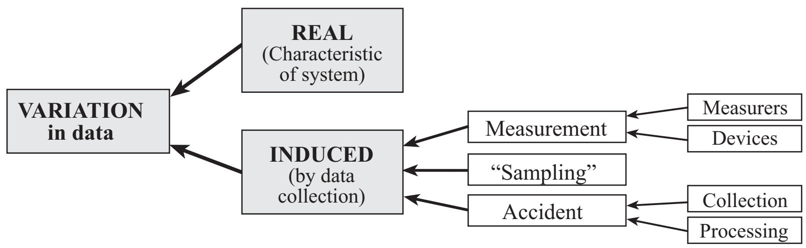


Figure 2. Sources of variation in data

Section 3 of Wild and Pfannkuch (1999) went on to discuss how statisticians look for sources of variability by looking for patterns and relationships between variables and in particular for those patterns that are likely to persist. It talked about explained and unexplained or residual variation. The majority of the section discussed "the quest for causes" and I don't want to touch on that here (except to promulgate "variation causes statistics"!). It concluded with the following: (1) variation is an observable reality; (2) some variation can be explained; (3) other variation cannot be explained on current knowledge; (4) random variation is the way in which statisticians model unexplained variation; (5) this unexplained variation may in part or in whole be produced by the process of observation through random sampling; (6) randomness is a convenient human construct which is used to deal with variation in which patterns cannot be detected.

We look for regularities or patterns in the observed variation and those that we believe, considering what we see in the data and what we understand about the mechanisms generating the data, are likely to be real and not ephemeral correspond to "explained variation." Unexplained variation, or "noise," is what is left over once we have "removed" all such patterns. It is thus, by definition, variation in which we can find no patterns. We model unexplained variation as being generated by a random process, implicitly if not explicitly. The simplest such models are regression models. We are papering over, at this point, a rather large crevasse which is the difficulty in deciding whether an apparent pattern in our data is likely to be a persistent characteristic of the process generating the data, and thus form a structural element in our model, or ephemeral and should be swept up in random elements of a model.

There is an old saying that goes, "If it looks like a duck, walks like a duck and quacks like a duck, then it is a duck." If it looks/walks/quacks like a duck, the statistician will use the inferential reasoning appropriate for ducks, despite having no real assurance that this bird actually has duck DNA. When modelling unexplained variation, because it looks random when viewed in any of the ways we have devised for inspecting it, we will draw the inferences that we know would be appropriate if it was in fact randomly generated. We do this because we do not know any better ways of proceeding (and don't believe anyone else does either). For further discussion, see Section 3.4 of Wild and Pfannnkuch (1999).

Having established “distribution” as a lens through which we view variation in data and explored the nature of explained and unexplained variation we will now start looking at distinctions between types of distributions that draw on these ideas.

2. EMPIRICAL VERSUS THEORETICAL DISTRIBTIONS

2.1. INTRODUCTION

In an effort to understand better how statisticians use “distribution,” I pointed Google at a number of sites including the American Statistical Association (ASA) where it searched the pages of the *Journal of the American Statistical Association*, the *Journal of Statistics Education*, other ASA journals, and many other resources. The adjectives and other qualifiers that I found used with “distribution” are collected in Appendix 1. By far the most common usages fell into two classes, “named theoretical distributions” (e.g., normal, binomial, …) and “the distribution of …” referring to the empirical or frequency distribution of some particular measured quantity, so that will be our starting point.

The distinction that underlies discussions of *empirical* versus *theoretical* distributions is between the variation we see in our data and a potential model for the process that gives rise to that variation (Figure 3). The empirical, *frequency* or observed distribution of our variable(s) contains the variation that we can see directly in our data. There is no inferential component, just a description of what exists in the data. When we move on to try to learn wider lessons from features seen in the current dataset, we conceive of unexplained variation present as having been generated by some unknown distribution. We often refer to this as the “true” or “underlying” distribution even though it is almost always a conceptual entity. When we use a full parametric model in our analysis we choose some named parametric distribution, such as the normal distribution, which we then assume to be what generates the data. This is the *theoretical* distribution, which describes or defines a probability model.

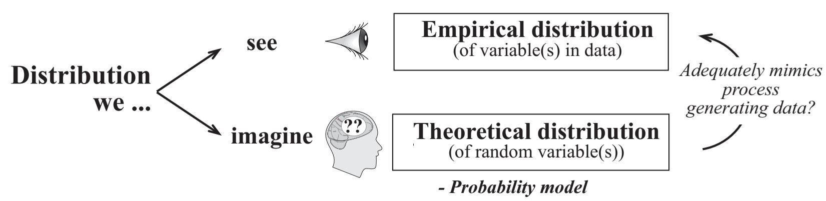


Figure 3. Empirical versus theoretical distributions

We have hundreds of humanly invented distributions for such purposes. In certain application areas, experience has shown that certain distributions are useful, but there is no way of ever knowing that our data are being generated from some particular distribution. So we never really believe our assumed theoretical distributions. The best we can hope for is that the act of sampling from the assumed theoretical distribution adequately mimics the most important features of the process which generated our data. Our lack of trust in the theoretical distribution leads to considerations of “robustness” and “goodness of fit.” That is, we would like to use inferential procedures that are comparatively insensitive to departures from distributional assumptions (robust) and we want to avoid using a theoretical distribution for inference that demonstrably does not “fit” the data – by which we mean that the distribution would be unlikely, in some sense, to produce the dataset we have in hand. To have any hope of making sense of this modelling process, students need to experience the behaviour of data which are generated from truly random sources (or as close as we can get to that) and lay that alongside real data. This is the crux of “connecting chance and data.” I will expand on this point in the next subsection.

Where do *outliers* fit? Outliers are observations we suspect are not being generated by the process which is generating the bulk of the data, but by a different (e.g., gross error) process. When we detect outliers we go back to the case labels hoping that there is some additional information we can uncover about that case that might help us understand why it appears so different. For example, we may be able somehow to determine whether the outlier is an error that can be corrected or removed.

2.2. “HEIGHTS ARE NORMALLY DISTRIBUTED”

At risk of belabouring points, I will now approach the ideas in Section 2.1 from another direction. How can we understand a statement like “heights are normally distributed”? We should not understand it in an absolute, literal sense because the statement is far more precise than anything we could ever actually know. Usually we are using “heights are normally distributed” in a loose descriptive way. The shape of the empirical distribution of the heights that we have seen (in whatever context) looks as though it is reasonably well approximated by the probability density curve of some particular normal distribution. We may make a leap of faith by believing that the approximation would still be good if we could look at the empirical distribution of heights from everyone in the parent population from which the heights we have seen have been drawn. We could make an even greater leap by thinking that this is probably also the way it would turn out if we looked at heights of people drawn from some other population that we have not yet investigated.

If we add the idea or reality of sampling at random from a population where the height distribution is well approximated by a normal distribution, then it follows that the behaviour of the data we get from sampling people and measuring their heights should be almost indistinguishable from the type of data we would get from taking random draws from a normal distribution. That latter behaviour can be investigated directly mathematically or via simulation.

If we make the assumption that our data on heights have been sampled from a Normal probability model then inferential statements (e.g., a confidence interval for the mean of the heights population distribution) follow from statistical theory as a consequence of that assumption. This is analogous to mathematics where, if one takes a set of conditions as holding true (axioms), then many other statements deduced as a logical consequence of these initial axioms (the theorems), must also hold true.

Some of the distributional leaps of faith in the first paragraph may be informed by a nonsignificant test of normality for our height data. But how much does this tell us? It tells us only that we cannot rule out the possibility that sampling variation alone may have produced the degree of “departure from normality” that we see with these data. Experience shows that, in virtually every situation, any theoretical distributional assumption we care to make will be shown to be implausible given enough data. What we are doing is never about the assumed theoretical distribution being right. It is only ever about the assumed theoretical distribution being a close enough approximation so that the methods of drawing inferences that follow from the assumptions we make are not misleading in any important way. This brings us back to robustness and goodness of fit as discussed in Section 2.1. *We make distributional assumptions in order to come up with* *methods of drawing inferences from data that still work when those distributional assumptions are not quite right.*

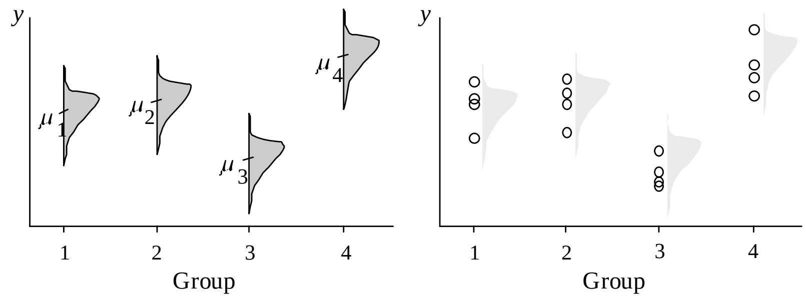
There is a very understandable desire to drive the teaching of probability models and distributions using real data because dice, coins, and so on are boring, even irrelevant. This runs into the problem discussed above. We can never know that any specific set of real data has been generated from any specific probability model. At best we can believe that the model is a good approximation. Probability models are abstract constructs that are used to model real-world behaviour. Their successful operation stands on two legs. The first leg consists of understanding the abstract construct that is the model, the sort of "data" the model generates, and how we reason inferentially in that idealised environment. The second leg consists of seeing the parallels that suggest to us that the model may provide a reasonable approximation to a given reality, of applying the model-based reasoning, and then of interpreting the results in terms of the original context. For most of the mental connections that have to be built in order to understand the model and the nature of its random behaviour, real-world context is simply a distracting irrelevance. That is not the place of real-world context. Interaction with context occurs in the recognition of model applicability, the interpretation of model parameters and the interpretation of any inferential statements that follow from applying the model-based reasoning.

As a non-traditional illustration, what students are experiencing in the fascinating basketball environment described by Prodroumou and Pratt (2006) is the stochastic behaviour of simulated "data" generated by a statistical model. While students do not directly learn anything new about basketball, by adjusting model parameters they can make the behaviour exhibited by the simulated environment (i.e., the statistical model) feel a lot like that of basketball. They can play with strategies that affect their performance in the simulated game and if they believe that the simulation gives an approximation that is close enough in key features to basketball they should then be willing to transfer some of the lessons learned in the simulation environment to the actual game.

2.3. MORE ABOUT DISTRIBUTIONS AND MODELS

We now explore some complications neglected in the discussion in Section 2.1. When we choose "a" theoretical distribution as a model for some variable, typically we are actually referring to an assumption that the true distribution is an unknown member of a parametric family of distributions such as the Normal(μ,σ^2) family. Here assigning different values to the parameters μ and σ^2 gives rise to different distributions within the family and we make statistical inferences about the unknown "true" values of the parameters that "produced the data."

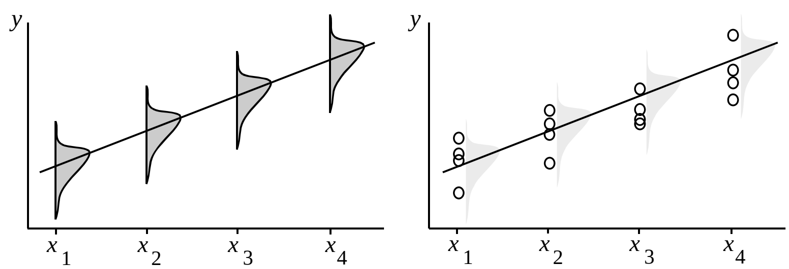
Beyond the simplest models, we do not just specify "a distribution." We actually build a construct using structural and random elements where each random element has a distribution. The simplest models of this form are the one-way analysis of variance model depicted in Figure 4, which underlies traditional inferential methods for comparing groups, and the simple linear regression model depicted in Figure 5. In Figure 4 the shapes are little normal curves coming up out of the page. The normal distribution for the y-values in group i is centred at μ_i. Under this model, "observations" belonging to the ith group are generated by sampling from a normal distribution with mean μ_i and some variance σ^2, which is the same for all of the groups. This is represented on the right hand plot, which also retains a "ghost" of the generating distribution. The model generates a type of pattern that we often observe in real data when we are trying to compare groups and thus forms a model for the mechanism generating such data that we can often apply in practice. In reality, the values of μ_i and σ^2 are unknown. Standard statistical inferences include testing for, or finding confidence intervals for, differences between the true group means.



(a) One-way Anova model (b) Data sampled from the model

Figure 4. The one-way analysis of variance model

The simple linear regression model in Figure 5 is essentially the same except that the true means μ_x when plotted against the *x*-value at which an observation is taken are constrained to lie on a line. The structural part of this model is the linear relationship between *x* and the mean value of *y*. The random part is the distribution of *y*-values taken at a given *x* around that the mean and that is what generates the observed scatter about the linear pattern.



(a) The simple linear model (b) Data sampled from the model

Figure 5. The simple linear model

Variation seen in even very simple data structures stems from a variety of sources (e.g., person-to-person, measurement, or occasion-to-occasion). There is a need to be able to think in quite sophisticated distributional ways to tease these things out. Hierarchies of random components (multilevel modelling) can be very helpful here. Luckily, in many commonly encountered problems it is not necessary to do so. Naïve approaches that sweep the subtleties under the carpet are actually valid. Suppose, for example, we want to compare the blood pressures of a drug-treated group and a control group on placebo. People do not have “a blood pressure.” At the very least there is person-to-person variation in the levels of their average blood pressure, occasion-to-occasion variation in actual blood pressure of the same individual, and measurement error is a third source of variation adding to the other two. The variability of blood-pressure readings seen within each of the two treatment groups is the result of all of these sources. Nonetheless, with only a single observation per individual (and admittedly under certain idealised assumptions) a 2-sample *t*-test, or confidence interval, for a difference in mean levels is still a valid analysis regardless of whether we have all of these sources of variation operating or if only person-to-person variation was operating. What differs is how we interpret the within-group variances. We would not raise the complications of multiple sources of variation for inferential beginners to avoid cognitive overload but suspicions about them might well cause unease in some.

Many complex models for processes involving time and space are built up in terms of chains or hierarchies of conditional distributions. For building models for processes evolving in time, for example, we often build up probability models conditionally by thinking in terms of what might happen next given the history of the process up until that point. This is a way of conceptualizing that permits prediction and also enables us to cope with data features like censoring.

With Bayesian inference we push the conceptual envelope out still further with the idea of describing the state of knowledge (prior to collecting the data) about parameters in a distributional model in terms of distributions called *prior distributions*. A Bayesian treatment of one-way analysis of variance, for example, would include prior distributions for all of the μ_i's and the variance σ^2. Inference proceeds by updating these prior distributions using information in the data to form corresponding *posterior distributions* intended to encapsulate the new data-informed state of knowledge.

2.4. ALL DISTRIBUTIONS ARE CONDITIONAL

All distributions we work with are really “conditional distributions.” This is not to say that we need complicated conditional probability ideas to think about them, just that they apply to particular subpopulations or systems operating under particular conditions or “settings” or to a particular time. We want to plant the idea that as conditions (or the groups we look at) change, the pattern of variation in an outcome variable often changes too and that we can learn useful things when we can quantify or otherwise describe the nature of those changes. If we can do this there is useful predictive information in such things as group membership and an impetus is given to trying to understand why the patterns might change. The regression problem can be conceived of as an investigation into how the distribution (pattern of variation) of a response variable y changes as the setting (x) changes. Group comparisons (two-sample, analysis of variance, etc.) can be conceived as an investigation into how the distribution (pattern of variation) of a response variable changes as we move from subpopulation to subpopulation (group to group) as shown, in an idealised way, in Figures 4 and 5. In the models depicted in Figures 4 and 5, all that changes about the distribution of y–values as group membership changes (Figure 4) or x changes (Figure 5) is confined to the mean level of the response. Spread, shape and everything else remains identical. Of course, even if this was true of the mechanism generating the data, in any observed dataset all of the features of the empirical distributions will still differ from group-to-group at least to some extent.

Regression and analysis of variance problems are not usually presented at this level of generality. The emphasis in most textbooks is not on how the distribution changes but on how the mean changes. Why this emphasis just on means? There are many reasons. One is a desire to look at the simplest feature of the distribution first. Then there is the historical influence of having well-worked out theory for simple models in which the mean is the only thing that changes as x changes (or as we move from group to group). Additionally, the parsimony principle (or Keep-It-Simple-Stupid principle) leads us to model only changes in mean unless the data forces us to do something more complicated. Other characteristics are much harder to make inferences about. For example, normal theory-based inferences for means are quite robust but those for spreads are extremely sensitive to departures from normality assumptions. As an indicative convention, the more "detailed" the feature being compared, the more data we require to usefully characterise or compare it.

2.5. SAMPLE DISTRIBUTION VERSUS POPULATION DISTRIBUTION

For beginning students we usually introduce the distinction between empirical and theoretical distributions gently via the distinction between the *sample distribution* and the *population distribution*. More precisely this is the distinction between the distribution of values for a variable for individuals represented in our dataset versus what that distribution would be if we had data on everyone in the population. Beginning this way is consistent with our desire in teaching to move sufficiently slowly from the concrete to the conceptual so that students do not drown in subtlety. Distributional models for data from processes are necessarily conceptual and immediately raise all sorts of difficult questions, for example, about the stability of the process through time and space and about dependencies. With data from a population, however, we can think in much simpler terms, namely of sampling from a large set of individuals at one point in time, and measuring one or more characteristics on each individual selected.

In practice, however, nothing is ever quite that simple, quite that "concrete." The really concrete, (real finite population measured once with one device at one time by one person in one way) is not really of interest to anyone because the quantities of interest are confounded, at the very least, by measurement-process variation. What we see is not exactly what is there. As soon as we allow for a contribution of the measurement process to the variation present in the data we are immediately transported from a manageable easily understood world to a world where data are generated by sampling from a conceptual population (an imagined construct) or are generated by some sort of random process (see Konold & Pollatsek, 2004). Urban myth has it that mediaeval mapmakers alluded to dangers lurking beyond the borders of the known world with the phrase "Here There be Dragons." Our maps of the statistical/inferential world made for beginning students need to be inscribed very carefully for teachers with "Here There be Dragons" underlined with "These Dragons be Real."

3. SAMPLING DISTRIBUTIONS

Next in importance, after the empirical and theoretical distributions of observations, are *sampling distributions* (see Figure 6). The former two relate to the *unit-to-unit variation* that we can see within a study or dataset and to a model for the generation of that unit-to-unit variation. (I prefer to personalize this and speak in terms of individual-to-individual variation.) Sampling distributions relate to *study-to-study variation* in estimates or statistics (e.g., sample means, proportions, regression slope estimates and *t*-statistics) which cannot be demonstrated from any particular study because each research study provides only one study-level data point. It is most accessibly introduced to students, I believe, in terms of the sampling variation in a parameter estimate, for example, of a population mean or proportion. Statistics educators now have a very good array of complementary ways of enabling students to experience the sampling variation generated by the process of "conduct a study and calculate an estimate" (see Chance, delMas & Garfield, 2004, pp. 294-297). This sampling variation can be modelled using either a (theoretical) probability distribution deduced from the distribution used to model the unit-level data or an asymptotic (large-sample) approximation, or it may be simulated using a resampling technique be it bootstrap, jackknife or permutation depending upon the situation and the analyst's taste.

The main priority with sampling distributions is to get across the idea that estimates and other statistics change every time we do a new study even if we perform each study according to exactly the same protocols. Properly appreciated, this becomes the prime motivator for the need for inferential methods which incorporate uncertainty, be they significance tests and confidence intervals or Bayesian. A second priority is the Central Limit Theorem for means which lays the groundwork for commonly used inferential techniques for a range of simple, but common, situations.

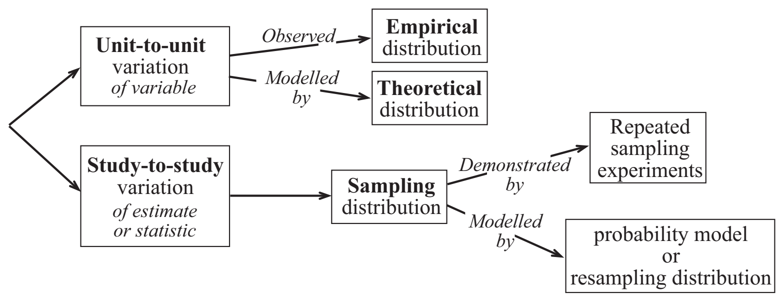


Figure 6. Incorporating the sampling distribution

Although the expression "sampling variation" appears often in the statistics education literature it appears very rarely in the statistics literature. There is no article on sampling distributions in the *Encyclopedia of Statistical Sciences* and of the over 300 section headings that contain the word "distribution" the phrase "sampling distribution" appears in only three. It is a background concept that underpins much of what we do in statistical inference but once the idea has been established it is seldom explicitly referred to. The adjective "sampling" is either dropped, to be inferred from the context, or it may appear in other guises such as the *null distribution* (of a test statistic), which is the sampling distribution that the test statistic would have if the null hypothesis was true.

4. COMPARISONS CURE UNIVARIATITIS

Teaching about the features of distributions for beginners tends to be in the context of a single variable, that is, in a univariate setting. All too often this has led to students being fed, year after year, a constant diet of univariate data and contrived univariate situations. I plead with teachers to move on to multivariate notions such as comparisons between groups and relationships between variables as soon as the most basic foundations have been laid. This is necessary to avoid infecting students with the dread disease univariatitis which is notorious for causing its victims to experience sensations of drowning in irrelevance and, ultimately, death by boredom. We may have to keep revisiting the univariate world but should take extreme care not to end up living there.

One important reason for using multivariate data early is that it gives a time-efficient environment in which students can themselves generate interesting questions to investigate using data, for example, by making interesting comparisons or investigating possible relationships. As pointed out in Wild (1994, p. 164), "not only is question generation arguably the most important part of the investigative process, the bubbling up of questions from an awakened curiosity provides much of the excitement of investigation."

Single distributions and the features of single distributions are seldom of interest in and of themselves. Interest generally lies in changes in these features, between places or groups, or over time. So why do we spend so much time working with single univariate distributions? Our main purpose is to lay the conceptual ground work that facilitates thinking about comparisons and relationships but traditional statistics teaching spends far too much time on it. We start with data and talk about centre, spread, modes, gaps, clusters, skewness, quantiles (particularly medians and quartiles), outliers, and other words are also starting to be used out of concerns about language being child-friendly and descriptive (e.g., "spreadoutness," "clumps," and "bumps"). This all too easily turns into what my colleague Matt Regan pejoratively terms "name calling." Let us put the simple data features to work in comparisons before we start naming and worrying about more detailed data features. Useful inferences about the latter are much less common in practice and much less reliable as well. So as soon as we introduce ideas like centre and spread we should put them straight to work in making some real and interesting comparisons – having visited the dull, grey, univariate world we need to bring the learning straight back into the vibrant real world. The same applies for notions like skewness. The fact that data for some variables are severely skewed is interesting mainly because data on other variables are not (another type of comparison) and because of practical implications of distributional shapes.

One of the many things we want students to be able to do when looking at plots of their data is to react to and wonder about causes for "the unexpected," particularly outliers – things that fall beyond "the expected pattern of variation." In order to do this students need some ideas about what to expect. A good place to start is the patterns of variation produced by sampling from a normal distribution or a finite population in which the characteristic of interest is approximately normally distributed. Particularly with small to moderate samples, the extent of what we might think of as "non-normal behaviour" present in data generated from a normal distribution can be astounding. Meaning should only be sought in those features of the data that correspond to features of the parent population or other mechanisms generating the data. Because exploratory data analysis is seldom coupled with exploration of models and random behaviour, many of the features beginning students point to, name and ponder causes for (gaps, clumps, outliers, skewness, bimodal behaviour, etc) are within the threshold of random error.

A recent innovation for the beginnings of inference introduced explicitly by Bakker and Gravemeijer (2004, pp. 158-165), but also used by others, for example, Konold and Pollatsek (2004, pp. 172, 180, 193), is the mind game for children of "growing the sample" which is basically concerned with conjecturing about what we might expect to happen to a display if "we added more people." In our terms, a data feature is meaningful only if it would still be present if we grew the sample substantially. For example, a gap would not be filled in, or apparent clusters would not coalesce. We move beyond name calling to statistical thinking when we can relate the features that we can see and name in our dataset, and believe will persist, to what we know about the world in order to arrive at some level of real-world insight, however small. We may, as a simple example, identify two clusters in a distribution and through further detective work determine that they are composed of identifiably different classes of individuals.

With categorical data, the most important reasons for working with relative frequencies (equivalently proportions or percentages), such as in relative frequency tables and resulting bar graphs, is to facilitate the comparison of datasets of different sizes, and to form a bridge to probability. With continuous measurement data, the real reason for teaching standardized histograms in which proportions are represented by areas is to lead in to the idea of probability density and density curves. This form of standardisation also permits comparison of datasets which have been summarised using different class intervals and to display a single set of data that has, for some other reason, been summarised using class intervals of different width. In practice, however, the need to do either of these things is so rare that I would never give it class time.

Some in the statistics education research community have found proportions of a sample below/above cut points (e.g., proportions of girls and boys with a height above 120 cm) provide a child-friendly introduction to the making of comparisons between groups when the response variable is continuous. It appears that this is something that many children do almost spontaneously. The reason we seldom see it in more advanced treatments is because the choice of cut point tends to be arbitrary and because this method of making comparisons is statistically inefficient. For example, more data is required to demonstrate a significant difference between groups this way than by comparing means. Statistical inefficiency does not provide a convincing argument against beginning students engaging with data in a way that is natural to them, however. It is much more important that they are enculturated to engage. Moreover, there are important areas in which the cut-point method is used, at least for communication purposes. Medical reporting often employs 5-year survival rates, for example.

5. DISCUSSION

The ultimate goal of statistical investigation is learning about some external reality and this involves forming and updating models of this context reality. In applied statistics there are three main elements that are brought together: current understandings of the context reality, data, and the use of statistical models and knowledge to guide how we collect data and learn from our data (understandings). Figure 7 attempts to represent the interrelationships.

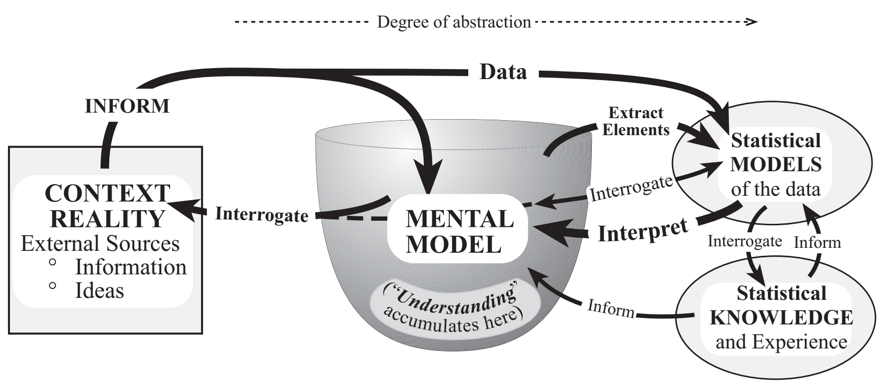


Figure 7. Learning via statistics

The need for statistics flows from variation, particularly the presence of "unexplained variation," in data. The statistical response to variation is to investigate, disentangle and model patterns of variation in order to learn from them. Virtually all of the ways statisticians do this involve looking at data through a lens which is distribution. While labels are interesting to children they have to learn to set aside labels and move beyond "who is this?" to start seeing and focussing on the patterns of variation and then to thinking about what aspects of these patterns might be expected to persist more generally. As Rubin et al. (2005) state, "aggregate views are preferable, as they are required to look beyond the data towards making inferences about the underlying populations or processes represented by data samples." When case labels are removed data records with identical values for the variables become entirely exchangeable and we are left with frequency distributions. The graphs and summaries we use are ways of looking at, summarising and conveying aspects of the information present in these distributions. Fundamentally, the notion of "distribution" is the pattern of variation in a variable (or set of variables).

The operation of the thinking processes represented in Figure 7 rests heavily on the interplay between the behaviour of our data and understanding the stochastic behaviour of potentially useful statistical models. To do this well requires bringing together the two elements of experience with exploratory data analysis and experience with the stochastic behaviour generated by models. Empirical (frequency) distributions tell us about data behaviour, whereas theoretical distributions are critical conceptual building blocks for statistical models. Put another way, the distinction underlying empirical versus theoretical distributions is between the variation we see in our data and a model for the process that generates that variation. We conceive of unexplained variation present as having been generated by some "true" or "underlying" distribution. In a full parametric analysis we assume these distributions are unknown members of a known, named parametric family of distributions. The idea of "population distributions" may paper over some complications for beginners but the paper is usually very thin.

Whereas empirical and theoretical distributions of observations relate to within study (or within dataset) variation that we can imperfectly see, sampling distributions relate to study-to-study variation in estimates or statistics which cannot be seen from any particular study because each study provides only one study-level data point. Sampling distributions motivate the need for and are a component of the development of statistical inference.

All distributions are conditional in the sense that they apply to particular subpopulations or systems operating under particular conditions or "settings" or to a particular time. The regression problem can be conceived as an investigation of how the distribution of a response variable changes as the setting (x) changes and group comparisons can be conceived as an investigation of how the distribution of a response variable changes as we move from group to group.

Because distributions are such a fundamental component of statistical reasoning our main goal should not be, "How do we reason about distributions?" but "How do we reason with distributions?," moving from a world where individual atoms are what is interesting to reasoning using aggregates. As Watson (2005) writes, children are beginning to learn about distributions from an early age starting when they first create pictograms of favourite fruits or modes of transport. They are not told, and do not need to be told, that they are learning about "distribution." Students typically first encounter summary features of distributions such as means, medians and even interquartile ranges long before they have any but the vaguest idea of "distribution." We look at graphs of distributions long before we develop the notion of distribution. Indeed our more complex notions of distribution and the nature of various features of distributions draw heavily upon the behaviours that have already been seen exhibited in graphs of data.

So do students need to be able to form and articulate a concept of distribution to be able to operate in a statistical way? Or, to steal from Nike, can students "Just do it" using graphs, summaries and an intuitive appreciation of variation? My feeling is that an explicit notion of distribution is not needed until we want to motivate, understand and then use probability models. Although distribution is the second foundation stone on which statistics is built ("variation" is the first), what is critical for early learners is much less, *"What is 'distribution'?"* than, *"How are my data distributed?"* and beginning to answer that question using appropriate graphs and summaries. One of the usual English-language meanings of the word “distributed,” taken from the *Oxford English Dictionary*, “is spread or disperse(d) abroad through a whole space or over a whole surface.” The first step is that measured characteristics of individuals (e.g., heights) are not identical – their values are distributed across a range – and that we can learn useful things by looking at how they are distributed. “How are my data distributed?” points someone like me in the right direction but does it speak to the variety of students that might experience it? We can be very grateful we have committed researchers in statistics education who are prepared to pick up questions like the above, and the issues raised below, and move the discussion beyond conjecture and anecdote.

It is my belief that we should be forming mental habits in which raw data (numbers) prompt students immediately to reach for pictures of that data, or as Moore (1991, p. 426) says, “a structure of thought that whispers, ‘Variation matters … Why not draw a graph?’” Summaries should be related conceptually to these pictures. Pictures of data, of distributions, do not need to be conventional pictures though they should converge to them over time as statistical knowledge develops; there are good reasons why conventional pictures have taken hold. Someone’s student somewhere at some time may very well come up with something startling and new which should inform the way everyone else does graphics but we must expect this to be rare. A trap for teachers is the presumption that conventional pictures are easy to read. Tools which are so transparent to the initiated (nothing to teach, it’s completely obvious, how could anyone *not* get it?!) can be quite opaque to beginners. Students need also to learn that there is not just one correct picture, that we can form a better overall view of reality by using an array of different pictures that better highlight different features of the data. Section 4 of Gould (2004) provides good examples of this and also how the insights so obtained feed into the development of statistical models for the data.

The key drivers for successful statistical practice, and thus the most critical elements to be instilled by statistics education are three propensities: the propensity to collect data that usefully addresses the question of interest, the propensity to question the applicability of data to the problem in hand, and the propensity to seek meaning in data. Everything else is about how to act on these propensities.

Most of the papers at STRL-4 and in this special issue deal with students’ engagement with empirical distributions, their features, and with comparisons of these features between groups. When features like location shifts between groups show up in a set of boxplots, for example, the following questions are never far off. “But does it actually mean anything?” “If we did it again would it come out much the same? Or could the order of the groups even be reversed?” Instantly we are transported to the realm of inference.

The inferences beginning students are able to make are necessarily informal, but therein lies the rub. There are great difficulties with informal inferences as Pfannkuch (2006) discusses. Assessment of “significance” balances the three factors: effect size, variability and sample size, in a very complicated way. Sets of standard boxplots that look identical, except for being based on different sample sizes, must be interpreted differently (notched box plots, Garret & Nash, 2001, provide a workaround). It may well be that there are no easy answers. There were some very sound imperatives that drove the development of our formal schools of statistical inference! As statistics educators we need to encourage our students into the mental habit of continually seeking meaning in data, which includes trying to make inferences, even using inadequate tools. The focus for the next SRTL Research Forum (SRTL-5 in 2007) and, I hope, a future special issue of *SERJ,* is informal ideas of inference. I look forward to the results with extreme interest.

ACKNOWLEDGEMENTS

The author is very grateful to Maxine Pfannkuch, Chris Reading and the participants at SRTL-4 for helpful suggestions and discussions.

REFERENCES

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

Chance, B., delMas, R. and 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.

Garfield, J., and Ben-Zvi, D. (2005). Reasoning about variation [Special Section]. *Statistics Education Research Journal*, *4*(1), 27-99.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4(1).pdf]

Garrett, L. and Nash, J. C. (2001). Issues in teaching the comparison of variability to non-statistics students. *Journal of Statistics Education*, *9*(2).
[Online: www.amstat.org/publications/jse/v9n2/garrett.html]

Gould, R. (2004). Variability: One statistician's view. *Statistics Education Research Journal*, *3*(2), 7-16.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Gould.pdf]

Konold, C. and Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 169-199). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Moore, D. (1991). Statistics for all: Why? What and how? In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics (ICOTS-3)*, Dunedin, August 1990, Vol. 1 (pp. 423-428). Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M. (2006). Comparing boxplot distributions: A teacher's reasoning**.** *Statistics Education Research Journal*, *5*(2), 27-45.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ5(2)_Pfannkuch.pdf]

Prodroumou, T. and Pratt, D. (2006). The role of causality in the Co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, *5*(2), 69-88.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ5(2)_Prod_Pratt.pdf]

Rubin, A., Hammerman, J. K., Puttick, G., and Campbell, C. (2005). Developing models of distributions using tinkerplots. 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*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Watson, J. (2005). Developing an awareness 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*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Wild, C.J. (1994). On embracing the 'wider view' of statistics. *The American Statistician*, *48*(2), 163-171.

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

CHRIS WILD
Department of Statistics
University of Auckland
Private Bag 92019
Auckland, New Zealand

APPENDIX: USAGE OF DISTRIBUTION

(Below "~" represents "distribution" to focus attention on accompanying adjectives and other modifiers)

NAMED theoretical ~ (e.g., normal, binomial, …)
~ OF SOMETHING OBSERVED
ACTING UPON ~s (e.g., comparing ~s, …)
Some sort of characteristic of ~ (mean, sd, quantiles, …)
Descriptor of (~ is skewed , symmetric, bimodal, long-tailed, …)

vertical ~, horizontal ~, length ~, Circular ~s, spherical ~s, Geographical ~
Of some sort of extreme (e.g., maximum flow ~)
Efficacy ~
Spatial ~, temporal ~
Stationary ~
Spectral ~
Survival ~
Shape ~s
Shot-noise ~s
Latent root ~s
Frequency ~
Empirical ~
Sample ~
Observed ~

Probability ~
Parametric ~
Dependence of ~al shape on parameters
Multiparameter ~s
Continuous versus discrete ~
Derived ~
Probability mass ~
Cumulative ~
Inverse cumulative ~
(cumulative) ~ function
Univariate ~, bivariate ~, multivariate ~
Conditional ~, Marginal ~, Joint ~
Truncated ~
Tolerance ~
Inflated ~s
Run length ~

Target ~
Theoretical ~
Unknown ~
Underlying ~, Population ~, True ~,
Sampling ~
Confidence ~
Null and Alternative ~s (testing)
Nonnull ~ theory
~ of a test statistic
Independence ~

Expected ~
Reference ~
~ free

Permutation ~
bootstrap ~, bootstrap resampling ~
jackknife ~
Simulated ~
Imputation ~
Predictive ~s

Error ~
Residual ~
Studentized ~

~ theory
~al properties
Spaces of ~s
Families of ~s
Asymptotic ~
Limiting ~
Convergence in ~
infinitely divisible ~s

Mixture ~
Mixing ~
Contaminated ~

Initial ~
Equilibrium ~, steady-state ~

Random effect ~
Frailty ~
Latent ~
~ of one or more latent variables

BAYESIAN
Prior ~, hyperprior ~
Posterior ~
Reference prior ~
Improper prior ~
(unmodified, or Bayesian or posterior) predictive ~
(Metropolis-Hasting) candidate ~
Target ~ (in MCMC sampling)
weighted

COMPARING BOX PLOT DISTRIBUTIONS: A TEACHER'S REASONING

MAXINE PFANNKUCH
The University of Auckland, New Zealand
m.pfannkuch@auckland.ac.nz

ABSTRACT

Drawing conclusions from the comparison of datasets using informal statistical inference is a challenging task since the nature and type of reasoning expected is not fully understood. In this paper a secondary teacher's reasoning from the comparison of box plot distributions during the teaching of a Year 11 (15-year-old) class is analyzed. From the analysis a model incorporating ten distinguishable elements is established to describe her reasoning. The model highlights that reasoning in the sampling and referent elements is ill formed. The methods of instruction, and the difficulties and richness of verbalizing from the comparison of box plot distributions are discussed. Implications for research and educational practice are drawn.

Keywords: *Statistics education research; Box plots; Distributional reasoning; Secondary statistics teaching; Informal statistical inference*

1. OVERVIEW

Traditionally, statistics instruction focuses on the construction of graphs, which results in students not knowing why graphs are constructed in the first place (Friel, Curcio, & Bright, 2001). Graphs are frequently used as illustrations of data rather than as reasoning tools to learn something new in the context sphere, gain new information, or learn from the data (Wild & Pfannkuch, 1999; Konold & Pollatsek, 2002). A shifting of the instructional focus to reasoning from distributions for the purposes of making sense of data, for detecting and discovering patterns, and for unlocking the stories in the data, presents many challenges. In particular, a challenge is to understand the nature and type of reasoning involved when making informal inferences from sample distributions about population distributions. Without research that attends to the complexity of informal inference and its role in the building of concepts towards formal statistical inference, statistical inferential reasoning may continue to elude many teachers and students. There is a need to understand inferential reasoning about many different types of distributions but this paper will focus on the comparison of box plot distributions. Box plots condense, summarize, and obscure information, incorporate statistical notions such as median and quartiles, and are conceptually demanding for students (Bakker, 2004). Therefore the aim of this paper is to achieve a greater understanding of the informal inferential reasoning necessary for comparing box plot distributions through analyzing one teacher's reasoning.

Statistics Education Research Journal, 5(2), 27-45, http://www.stat.auckland.ac.nz/serj

1.1. REVIEW OF RELEVANT LITERATURE

Exploratory data analysis (EDA) gave rise to a number of new graphical techniques. Tukey (1977) invented box plots as a powerful way of summarizing distributions of data to allow visual comparisons of centers and spread through the five-number summary (minimum, lower quartile, median, upper quartile, maximum), which divides the data into four equally sized sections. Further refinements can be made to basic box plots by visually representing extreme values or outliers, means, and significant differences. Basic box plots are introduced to students from as young as 12 in the USA to as old as 17 in France, whereas some countries, such as China and Israel, do not have them in the curriculum (Bakker, Biehler, & Konold, 2005). In New Zealand box plots have been in the curriculum for 14-year-olds for the last 20 years.

When comparing two box plot distributions traditional instruction assumes that inferences will *not* be drawn and hence focuses on describing features of box plots. Recent changes, however, to Year 11 (15-year-old) assessment in New Zealand assume that conclusions will be drawn from visual comparisons (Pfannkuch & Horring, 2005). At this year level students have not been exposed to confidence intervals and significance testing to draw conclusions; rather their reasoning must involve informal inferential reasoning. That is, in the case of box plots, being able to infer that one group is generally greater than a second group, or that no distinction can be drawn, based mainly on looking at, comparing, and reasoning from box plot distributions. The question arises as to what elements of reasoning are necessary to draw informal inferences.

Because formal inferential reasoning focuses on the centers of distributions the question arises as to how to scaffold students' understanding towards viewing centers as being representative of a set of data. Konold and Pollatsek (2002) note that research has demonstrated that students know how to compute averages but few use averages to characterize a dataset or to make comparisons between datasets. Such a situation does not provide conceptual foundations for the development of students' inferential reasoning. Furthermore, Konold and Pollatsek (2002) identify four views of average – signal amongst the noise, data reduction, fair share, and typical value – which are dependent upon the goal the person has in mind when using an average. They argue that all goals are valid but if students do not have the "signal amongst the noise" view of average then this can result in a reluctance to use averages to compare two groups, a fact noted by Biehler (2004) in his research on box plots. Therefore, adopting the position that the middle part of the data usefully characterizes the group and that the middle parts of the distributions should be compared is a necessary element of the reasoning process.

Box plots illustrate the signal (the center) and noise (the spread of data from the center) in their representation yet according to Biehler (2004) the interpretation of spread can result in five different views, namely: location information, regional spreads and densities, global spread as a deviation from the median, median upward and downward spread, and classification information. Whatever view is taken, a spread element of reasoning must include notions of comparing variability within and between box plots. Biehler (2004) and Friel (1998) identified that the cut-off points represented in the box plot result in students using these for comparing distributions. That is, the nature of the representation leads students to argue intuitively with the data by comparing equivalent and non-equivalent five-number summary points. Thus another element of reasoning associated with comparing box plot distributions is of the summary type. Other elements of reasoning identified by Biehler (2004) as lacking in students are the "shift" interpretation and intuitions about sampling variability. He describes the "shift" element, where all the five-number summary values are higher for one box plot compared to the other, as being an essential notion in comparison whereby students can determine the amount of the shift and the type of shift, uniform or non-uniform. If this shift type of reasoning does not work with the box plots under consideration then the comparison becomes complex (Bakker et al., 2005). Biehler's (2004) reference to sampling variability accords with Bakker (2004), who states that a key concept in developing a notion of distribution is sampling, and with Pfannkuch (2005), who believes that sampling reasoning is essential for building concepts towards formal inference. Furthermore, Bakker and Gravemeijer (2004) argue that in instruction students should experience summarizing dot plot distributions by intuitively dividing the data into groups. Such instruction can gradually develop a student habit to overlay box plots on dot plots.

Currently, studies are focused on how to introduce students to box plots and how students interpret them. There appears, however, to be no research on how teachers reason when comparing box plot distributions, nor any definitive account of how teachers or students should draw informal inferences. According to Bakker and Gravemeijer (2004) reasoning with shapes forms the basis of reasoning about distributions whereas Friel et al. (2001) refer to visual decoding, judgment, and context as three critical factors in students' abilities to derive meaning from graphs. Furthermore, Friel et al. consider that research is needed on understanding what it is about the nature of the reasoning that makes comparing datasets such a challenging task. Whatever the nature of the reasoning is, it is complex and may depend on the ability to decode representations, to attend to a multiplicity of elements represented within and between the box plots, and to make judgments.

1.2. RELATED RESEARCH

The research described in this paper is part of a larger project that is concerned with developing students' statistical thinking based on the Wild and Pfannkuch (1999) framework. In 2003, the first year of the project, informal inferential reasoning was identified as a problematic area. Focusing on the comparison of box plots, the videotape data of the classroom teaching revealed that the teacher in only one instance out of a possible eight opportunities communicated and wrote down how she would draw a conclusion from such plots (Pfannkuch & Horring, 2005). Over half the students, in an open-ended questionnaire, identified that they did not know how to draw evidence-based conclusions. An analysis of student responses to an assessment task requiring the drawing and justifying of inferences from the comparison of box plots concluded that 90% compared equivalent and 50% non-equivalent five-number summary statistics, a "summary" element; 50% mentioned the difference in the ranges, a basic "spread" element; and 30% had a very basic "shift" element of reasoning.

Realizing that drawing conclusions from the comparison of box plot distributions is not an easy task the researcher and five statisticians met in 2003 to discuss the type of reasoning that could be expected for informal inference. In teaching situations where there is no access to technology and no student experience of sampling variability, such informal inference was considered problematic not only for students but also for the statisticians (Pfannkuch, Budgett, Parsonage, & Horring, 2004). From the perspective of formal inference for the comparison of data plots the statisticians determined that there were four basic aspects to attend to in order to understand the concepts behind significance tests, confidence intervals, p-values and so forth before drawing a conclusion. These were comparisons of centers, comparing the differences in centers relative to the variability, checking the distribution of the data (normality assumptions, outliers, clusters), and the sample size effect. The discussions raised further questions as to what types of learning experiences would develop students' inferential reasoning towards a more formal level.

Since articulating the messages contained in box plots and justifying inferences either verbally or in writing was considered difficult for both teachers and students, the idea of providing a framework to support the reasoning process was conceived. The framework would support learning in terms of what should be noticed and attended to when looking at the plots. Since Year 11 students had not been exposed to ideas of sample and population or of sampling variability and the effect of sample size, the group conjectured that perhaps students should work with clear-cut comparisons that had similar spread, no unusual patterns, and samples sizes of 30. For writing a conclusion they proposed that it should begin with the words: "These data suggest …" and that the justification should be focused on comparing the centers and on comparing the differences in centers relative to the variability. After that the students could comment on features such as variability within and between the box plots, the shapes of the distributions and compare the median of distribution *X* with the percentage of the distribution *Y* that was below it. Finally the students should check whether their conclusions made sense with what they knew from their own knowledge and consider possible alternative explanations for the findings – an explanatory element of reasoning. The statisticians and researcher also suggested dot plots should be kept with box plots (Bakker & Gravemeijer, 2004; Carr & Begg, 1994) and gave ideas on how students could experience variation (Pfannkuch, 2005).

When these ideas were presented to the teacher who was being researched, she was adamant that she wanted to deal with the inherent messiness of data where clear-cut decisions are not obvious. She also felt that some suggestions for justifying inferences and comparing features were too hard for students. At this stage, the teacher was not ready to deal with sampling variation ideas or putting dot plots and box plots together, but she was ready to try and reason from box plots. Since there seemed to be no account in textbooks and in research of how to draw informal inferences from box plots in a school teaching situation and no consensus on the statisticians' suggestions, the teacher and researcher were placed in the situation of learning in and from practice.

1.3. RESEARCH QUESTION

As part of a larger project on developing Year 11 students' statistical thinking, the following research question is addressed:

> *What reasoning does a teacher articulate when learning to communicate statistical ideas and make informal inferences from the comparison of box plots?*

2. RESEARCH METHOD

The research method is developmental in that an action-research cycle is set up whereby problematic areas are identified by a teacher and researcher through observations and critical reflections on the implementation of a teaching unit and by the researcher through analysis of student assessment responses (see Pfannkuch & Horring, 2005, for a more complete account). The teacher and researcher then discuss how the current situation might be changed for the following year when the unit is taught again.

According to Ball and Cohen (1999) actual teacher learning requires some disequilibrium since learning will only occur when existing practices are challenged. From the teacher's perspective, her practice had been challenged and hence in the 2004 implementation of the statistics-teaching unit, the teacher decided to make a conscious effort to communicate and articulate how she was looking at and what she was thinking about when comparing two box plots. She was also aware of the need to write down the justifications for her conclusion. In this research the teacher is being put in the position of a learner in and from her practice, that is, actively learning while she is teaching. Therefore the action-research method is appropriate in such a situation.

The teacher and researcher decided before the teaching of the unit that when reasoning with box plot distributions she would refrain from using the summary element of reasoning and instead focus on the following five elements: comparison of centers, spread, the degree of overlap of the two box plots, sampling, and explanatory. The teacher decided when to introduce each element, what language she would use and how she would reason within those broad elements. After each teaching episode, the researcher and teacher had brief conversations about the type of reasoning used and what possibly to emphasize in the next lesson. Each lesson was videotaped by the researcher.

2.1. PARTICIPANT

The school in which the project is based is a multicultural, secondary girls' school. The teacher is in her mid-thirties, and has taught secondary mathematics for twelve years. In Year 10, students are introduced to the graphing of box plots. The class is taught mathematics by the teacher for four hours per week. The teacher is in charge of Year 11 mathematics and therefore, in consultation with the other Year 11 teachers, writes an outline of the content to be covered, together with suggested resources and ideas for teaching the unit. The researcher previously knew the teacher on a professional basis.

2.2. THE TEACHING EPISODES

This paper focuses on two of the three teaching episodes in which box plots were introduced and discussed. The teacher chose the tasks she gave to the students. Before the first teaching episode on comparing box plot distributions the students compared data using back-to-back stem-and-leaf plots and calculated the five-number summary for data. For homework the students were given an example from a textbook (Figure 1(a)) for which they were required to draw a back-to-back stem-and-leaf plot (Figure 1(b)) and calculate the five-number summary. The first teaching episode on constructing a box plot started at this stage. The teacher discussed and interpreted the stem-and-leaf plot, then used the five-number summary to remind the students how to draw box plots. She drew the box plot of males' pay with the class (Figure 1(c)). After the students had drawn the box plot of the females' pay she discussed the plots with them and built up a written conclusion on the board (Figure 1(d)).

In the second teaching episode the focus was on interpreting box plot distributions. The students were given a brief account of where the data had come from (Figure 2 (a)) and the dataset. They were asked to reflect on the background information and the given data and to think of questions they might pose. After the students suggested a number of questions the teacher said she had already drawn the plots for one of their questions (Figure 2(b)). The teacher articulated her reasoning from the comparison of the box plots with the class responding to and asking her questions about the data. Figure 2(c) is the conclusion she wrote on the board.

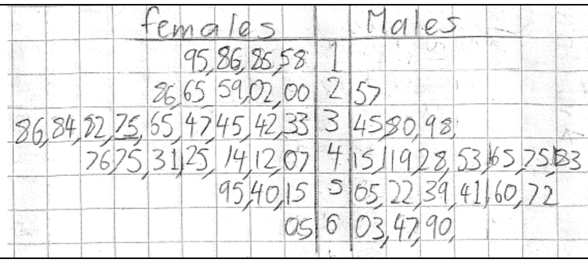
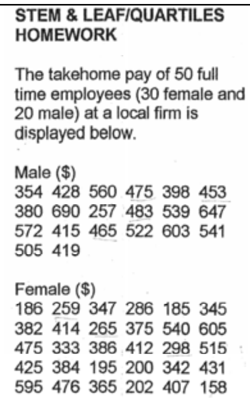


Figure 1(b)



STEM & LEAF/QUARTILES HOMEWORK

The takehome pay of 50 full time employees (30 female and 20 male) at a local firm is displayed below.

Male ($)
354 428 560 475 398 453
380 690 257 483 539 647
572 415 465 522 603 541
505 419

Female ($)
186 259 347 286 185 345
382 414 265 375 540 605
475 333 386 412 298 515
425 384 195 200 342 431
595 476 365 202 407 158

Figure 1(a) (Source unknown)

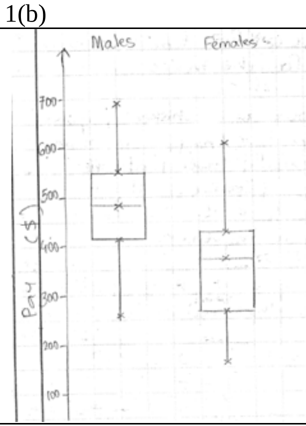
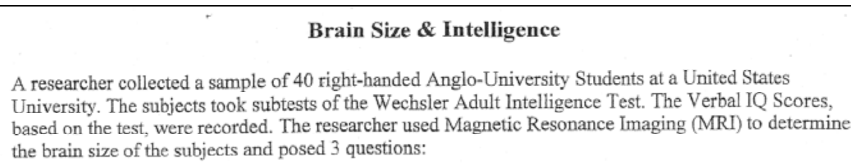


Figure 1(c)

Teacher's written conclusion was:
Each of the statistics (median, UQ, LQ) for females is lower than for the males. The female graph is clustered between the median and the UQ. The box & whisker graphs overlap but the female graph is generally lower than the males' graph. Overall, it appears that the females earn less than the males.

Figure 1(d)

Figure 1. Teaching episode one – graphs are from a student's book



Brain Size & Intelligence

A researcher collected a sample of 40 right-handed Anglo-University Students at a United States University. The subjects took subtests of the Wechsler Adult Intelligence Test. The Verbal IQ Scores, based on the test, were recorded. The researcher used Magnetic Resonance Imaging (MRI) to determine the brain size of the subjects and posed 3 questions:

Figure 2(a)

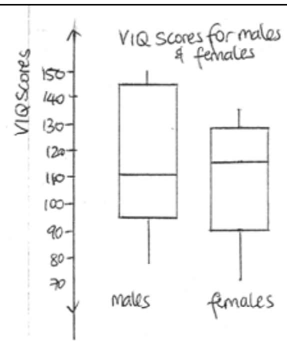


Figure 2(b): Teaching episode two

Teacher's written conclusion was:
Males' graph – slightly higher. Males' IQ is more spread than females. Interquartile range = UQ-LQ. Males IQR=39, females IQR=50. Graphs are overlapped – especially between LQ and UQ. Measure of central tendency – Female median (116) is slightly higher than males. Based on these data values, we're not certain that males have higher IQ. There is some evidence to suggest that males have higher IQ for these Uni. students.

Figure 2(c)

Figure 2. Teaching episode two – graph given to students by teacher

3. RESULTS

A qualitative analysis by the researcher of the teacher's discussion on the comparison of box plot distributions extracted ten elements of reasoning (Figure 3). Elements 4 to 7 were based on the a priori agreement, as described earlier, between the teacher and researcher on the type of reasoning elements that should be emphasized in instruction, whereas elements 1, 2, and 3 arose during teaching and the analysis. The eighth element was considered after discussion with another researcher (Tim Burgess, personal communication, 7 July 2005). It was not determined before teaching how the teacher would reason within these broad classifications. The analysis of the data suggested that the eight elements of reasoning are non-hierarchical, interdependent but distinguishable. The two moderating elements of reasoning, 9 and 10, arose from the analysis and are contained within each of the other eight elements.

ELEMENTS OF REASONING	
1. Hypothesis generation	Compares and reasons about the group trend.
2. Summary	Compares equivalent five-number summary points. Compares non-equivalent five-number summary points.
3. Shift	Compares one box plot in relation to the other box plot and refers to comparative shift.
4. Signal	Compares the overlap of the central 50% of the data.
5. Spread	Compares and refers to type of spread/densities locally and globally within and between box plots.
6. Sampling	Considers sample size, the comparison if another sample was taken, the population on which to make an inference.
7. Explanatory	Understands context of data, considers whether findings make sense, considers alternative explanations for the findings.
8. Individual case	Considers possible outliers, compares individual cases.
MODERATING ELEMENTS OF REASONING	
9. Evaluative	Evidence described, assessed on its strength, weighed up.
10. Referent	Group label, data measure, statistical measure, data attribution, data plot distribution, contextual and statistical knowledge.

Figure 3. Teacher's model of reasoning from the comparison of box plots

3.1. THE ELEMENTS OF REASONING

The goal of the teacher was to make an informal inference about populations when comparing sample distributions and to justify that inference. Since informal inferences were being drawn, visuo-analytic thinking was used by the teacher. She gradually built up, in her communication, the multifaceted ways in which she looked at and interpreted the comparison of box plots. Within some elements there are sub-elements, not all of which are illustrated below. The focus of the analysis is on her reasoning within each element as she learns more about the data under consideration. Her reasoning, however, is linked to how she teaches and therefore consideration is given to instructional methods in the analysis. It should be noted that the teacher uses the term graph when talking about stem-and-leaf plots or box plots but to be consistent the analysis of her reasoning will use the term plot.

Element 1: Hypothesis Generation Hypotheses may be formed at the beginning of an investigation before data are collected or on inspection of a given dataset, or during an investigation when analyzing a graph, or at the end of an investigation. In teaching episode one (see Figure 1) the teacher discussed the back-to-back stem-and-leaf plots before showing the students how to construct the box-and-whisker plots. In the hypothesis generation element she communicates that the inference "males earn more than females" does not capture the actual story in the data.

Teacher: All right, what's the first thing that strikes you when you looked at this graph [Figure 1(b)]?
Student: Males earn more.
Teacher: Males earn more – what gave you that impression from the graph?
Teacher: Yes, the higher amounts go down, so the mass or the bulk of the graph is lower down than the females. So that gives us the impression, that the males earn more than the females. Is that true for every single person? Did every single male earn more than every single female? No. Okay. This person here, this woman here, earns $605, she earns way more than this male here $257 …[later on] … because there's a bit of an overlap, I have to be a bit more subtle about my language and say it appears from the graph that the bulk of males, appear to earn more than the bulk of the females.

To generate a data-based hypothesis, consideration is given to variability through acknowledging that the reasoning is about the group trend and not about individual cases.

Element 2: Summary In this element the five-number summary is located on the plots and equivalent summary points are compared. For example in teaching episode one:

Teacher: Right if we were to compare each measure like the median, the females are lower than the males. Lower quartile, females are lower than the males. Upper quartile, females are lower than the males, top – right, so each of the statistics is lower for the females than it is for the males.

The box plot representation also encourages the comparison of non-equivalent summary points such as "75% of males earn more than 75% of females," which is a comparison of a lower quartile with an upper quartile. The teacher briefly mentioned that "a quarter of the ladies, women, are earning more than these guys" in teaching episode one. Sometimes her focus is specifically on the comparison of the medians as was the case in teaching episode two:

Teacher: The next thing that I think is a really important factor for helping me make up my mind is the measure of central tendency or average. What have I got to help me figure out what the central tendency is?
Student: The median.
Teacher: The median, so next I'm going to look at my median because that's the middle of the data, that gives me where the bulk of the data is and I've got females slightly higher than males. So I'm going to say the feature that I've noticed for the female's median I'm going to say what it is (116), is slightly higher than males.

The notion of the median being representative of the distribution or the signal amongst the noise is unclear in the communication but the teacher is drawing attention to its importance as a factor for making a decision under uncertainty. Mentioning the "bulk of the data" in terms of the median may be a misleading interpretation for these datasets.

Element 3: Shift In the shift element the plots are looked at as a whole and compared in terms of whether one is higher or further along than the other. This shift element could be incorporated into the summary element when equivalent summary values are compared but the way of reasoning is different in that for the shift element the box plots are looked at as a whole visually not as a straight comparison between values. From teaching episode two:

> Teacher: All right, one of the biggest factors that helps me make up my mind, … is whereabouts is the whole of the graph in relation to each other. And when I look at this, straight away I notice that the males' graph is a little bit higher up than the girls' graph. Pretty much?
> Teacher: You reckon it's a long way?
> Teacher: No, not a long way up, they're quite overlapped, they seem to be quite next to each other but one's a little bit higher than the other. So, I'll write that down. Male's graph is slightly higher.

At a more detailed level she compared the plots in terms of the shift of the majority with statements such as "the mass or the bulk of the graph is lower down." At no stage did she quantify the shift, preferring instead to use qualitative statements such as "slightly higher."

Element 4: Signal The signal element could be incorporated into the shift element but is given a separate category since reasoning about measures of center is important for formal inference. The signal element referring to the middle 50% of data, the box, may represent the starting point for informal reasoning about center. In her communication in teaching episode two the box appears to be used as a crude measure of the center and is represented as the "typical value" or the "signal" of each distribution. She is interested in how much overlap there is between the middle 50% of data so a sense of comparison of the "middle" is conveyed to students. In response to a student's query on the meaning of the term overlap she drew double-arrowed lines in the "central boxes" to demonstrate the term.

> Teacher: All right, I also might notice that the graphs are overlapped okay.
> Student: What does that mean?
> Teacher: Well that means that there's not one graph separate from the other graph, they're overlapping. So see this central box here and this central box here, remember that gives me the middle 50% of the data, they are quite overlapped, okay. Especially between, and I'll write this down, especially between the lower quartile and the upper quartile. They're very overlapped in this central part. Now in terms of which part of the graph gives me the most information or the most significant information okay, this middle box is most important, okay 'cause that's where the middle bulk is.

Possibly the teacher is laying down intuitive foundations for formal inferential reasoning where the difference in centers are compared relative to the variability. The drawn lines may also be conceived as visual foundations for confidence intervals for population medians.

Element 5: Spread The teacher drew attention to the spread element by focusing on comparing ranges and interquartile ranges visually and quantitatively. At a more detailed level she drew attention to the location of the data, that is, where and how the data are distributed. From teaching episode one:

Teacher: What does each of these sections represent? Because that's the highest, the upper quartile, median – because each part, there's four parts. See how I've broken the graph essentially into four parts.
Student: Is it the spread of money?
Teacher: It's the spread, yes, it's the spread of how much they get paid. So there are, of all the people in this study, 25 of them – percent sorry, 25 percent – a quarter right, are sitting in here. … Okay, in here is another quarter of the people, in here, another quarter, and in here another quarter. So if you imagine these women, and they were standing on this number line, and they would be –
Student: Squashed up.
Teacher: Squashed up, yes. Right because there's the same number here, same number here, same number here. But because this is a smaller area, they were close together. What about up here?
Student: Sparse.
Teacher: Yes, spread out – sparse. We've found a new word. What's the new word for today? So they're more spread out.

When comparing densities of the data there is a dual comparison: for each quarter within one group and between the two groups. This dual comparison was not communicated in this episode, rather she focused only on the female box plot. In the second teaching episode the spreads at a global level were compared and therefore the dual comparison tended to be communicated to students in a vague manner. In her reasoning the distinction between comparing variability within and between box plots was not well articulated. Although she did not consider the shape of the distributions, such as symmetry or skewness, in these teaching episodes she did in another teaching episode involving a matching exercise between histograms and box plots. At no time did she consider whether the shape was expected or unusual.

Element 6: Sampling The sampling element is underpinned by the belief that the data have been sampled from a population and that an inference will be made about the underlying population distributions from the sample distributions. David Pratt (personal communication, 7 July 2005) observed that confusion existed between the teacher and students about the game being played in this element. The students believe the teacher is making inferences about the samples, which Pratt refers to as *game one*, whereas the teacher is attempting to make inferences about the populations from the samples, which Pratt calls *game two*.

In the sampling element consideration is given to the sample size of each group and its effect on any inferences, whether a repetition of the experiment would give rise to the same difference, and determining the population for which the inference is applicable. In the 2003 analysis of the student assessment data about half the students, on the basis of fairness, mentioned that the sample size of the two datasets being compared should be the same (Pfannkuch, 2005). In cognizance of this finding the teacher brought students' attention to the datasets from teaching episode one:

Teacher: Yesterday's graph …did you remember that one set had 20 in it and the other set had 30 in it – were we still able to make comparisons between those sets?
Student: But that's not very fair.
Teacher: Not very fair?
Student: Like, with males and females.
Teacher: Yes, it's not but it doesn't necessarily affect your conclusion.
Student: Oh cause with box and whisker it doesn't matter cause it's just percentages aye?
Teacher: That's right, good call. Okay, so with the box and whisker it doesn't matter so much, although if one set could be smaller like say 5 people and you had say

another set which had 30 and you were comparing them – then you probably would make some mention on that okay. But if they're roughly the same that's fine, doesn't have to be exactly the same.

The student seems to understand the sample size effect from a proportional basis, but the teacher's point appears to be that small samples (n=5) are more variable than larger samples (n=30). If she was playing game two she would point out that it would be unwise to draw conclusions from the comparison of such samples, but instead she states that it would be noteworthy. There is considerable conflict in this interchange since the teacher does not seem to have resolved which game she is playing.

When discussing the plots in teaching episode two the teacher attempted to hypothesize what would happen if she did the study again. Two ideas appear to be present in her discussion. The first idea is "if I took another sample from the population would I get the same results?" and the second idea may not have been the teacher's intention but it is worth considering, "if I repeated this experiment again would I get the same results?"

Teacher: Now I'm going to throw in one more idea that I hope will convince you. This is 20 people, 20 people here. Okay, do you reckon, if we went back to the same place and did the same study and got another different 20 people, 20 boys and 20 girls, do you reckon we'll get exactly that same graph?
Student: No.
Teacher: Do you think, but don't you think that the median will be just a little bit higher for the girls?
Student: Yes.
Teacher: Do you think? Is it possible that maybe the results will be the other way around if it was another 20 people?
Student: Yes.
Teacher: Okay, can you see that really, they're so close, that if you were to get another 20 people, that it might just come out the other way. And then maybe in Mrs. L's classroom, maybe they've got that dataset and the girls graph might be a little bit higher than the boys and they'll be saying oh yes, and here we are looking at these, we read them the other way. So we have a bit of a problem here, I've only done this study once, we only did it with 20 girls and 20 boys, and probably *if we repeated the experiment*, we would find that we would have slightly different results.

The two ideas, taking another sample and repeating the experiment, are distinct. The first idea centers on the resultant outcome if a different sample was taken from the population and hence game two is being played. The second idea is to consider the consequences if the experiment was repeated on the same people. The implication of this idea is that another source of variation, namely measurement errors, should be considered but the game being played with this idea is game one. Again the sampling element becomes muddied.

The conflict between making inferences about samples, game one, and about populations, game two, is further illustrated by a student who wanted a definite conclusion and the teacher's subsequent conclusion.

Student: But couldn't you say, from the graph, that males do have a little bit higher IQ than females?
Teacher: … We're writing our conclusion: *"Based on these data values we are not certain that males have higher IQ."* It's not certain okay. *"There is some evidence to suggest that males have higher IQ for these students."*

The teacher’s first statement draws a conclusion about populations whereas her second statement draws a conclusion about the samples. When writing her second statement (“There is some evidence to suggest that males have higher IQ for these students.”), she draws students’ attention to who was studied:

> Teacher: I’m even going to write, Uni students, all right, because these are all University students. I mean if we were trying to find out some information about all men and for all women and their IQ then this study wouldn’t be enough. We would want to do surveys of people who are older, who are younger, who have different types of jobs, males, females, who are from New Zealand, Australia, India, China, Japan, Scandinavia, right we want to do that with people from all over the place. So we want to be really careful.

Unfortunately this reference to “Uni. students” was a game one statement, which further confuses the situation. Enculturating students into looking at who was studied in the sample and then being careful about determining the population on which the results can be generalized is part of learning about inference space judgment. Such a judgment is only possible when sufficient information is known about the data, which was not the case in teaching episode one.

Furthermore, during her discussion in teaching episode one she mentioned that if the distributions overlapped she would be careful about making a claim that men earned more than women, whereas if there was no overlap she would make the claim. The lack of overlap in sample distributions could be an artifact of sampling variation and hence the indeterminacy of her sampling reasoning is continued.

Element 7: Explanatory In the explanatory element, the background to and findings from the investigation are considered by referring to one’s own real-world knowledge. This contextual knowledge is used to check whether the findings make sense or whether other variables should be considered before venturing a conclusion or hypothesis about the situation under consideration. Before students can compare box plots they need to understand the origin of the dataset, where and how the data were collected, and how the measures were defined. In teaching episode two the teacher first of all engaged the students’ interest by telling them about an interesting talk on brain development that she had recently heard. Secondly, she discussed the measures used and on whom the study had been conducted. Part of her conversation was:

> Teacher: What, where does this data come from?
> Student: United States.
> Teacher: It comes from the United States, okay. What else do we know about this data?
> Student: They’re all right handed.
> Teacher: They’re all right handed, good, so all the people in this survey were right handed. What else do we know about these people?
> Student: They’re university students.
> Teacher: They’re university students, okay. What age group are university students usually?
> Student: Twenties?

Such information sets the data in context and lays the foundations for drawing reasonable inferences from data. In comparison, in teaching episode one, the data came from a textbook and all that was known about the data was that they were collected from a local firm. Such a paucity of background information led the teacher to consider female and male salaries in general, in an attempt to discuss whether the findings made sense with what they knew about the world.

> Teacher: Did anyone see the recent results on the average salaries for men and women? I remember seeing something on the news about that. I think it was to do with people who work for the government and public service and that includes people like teachers, nurses, policemen, officials who work in government departments – everyone who gets paid by the taxpayer if you like, right, they did a survey to have a look at who earned more – men or women and they found that it appeared that men earned a little bit more than the females. …

Later, the discussion considered whether there was an alternative explanation for the findings rather than gender being the discriminating factor for salaries:

> Teacher: Like, teaching for example, whether we're female or male maybe doesn't affect how much we earn, but maybe it affects things like –
> Student: What position you're in.

Such a discussion enables more variables to be considered before making an inference based on given data. Thinking of confounding variables and alternative explanations for findings are part of the argumentation with data, and more information on this dataset, together with other relevant data, could have provided a richer exploration.

Element 8: Individual Case When reasoning from distributions, observations which appear to be outliers are inspected as individual cases to determine whether they are part of the dataset or are errors and can be corrected or removed. Because box plots are not drawn with outliers at this Year level and dot plots were not kept under the box plots this element was not articulated. However, the teacher did reason with individual cases when she was arguing from a hypothesis generation element:

> Teacher: This person here, this woman here, earns \$605, she earns way more than this male here \$257.

The comparison of individual's earnings between the datasets is a method of argumentation based on particular instances that is used to illustrate that definitive statements cannot be made for all cases.

3.2. THE MODERATING ELEMENTS OF REASONING

The moderating elements of reasoning, evaluative and referent, serve two distinct supporting functions in the reasoning process. The evaluative element's function is to support the reasoning process by qualitatively judging the strength of the evidence provided by an element and then weighing up that evidence towards making a decision about whether there is a real difference between the two groups under consideration. The referent element's function is to ground and maintain the reasoning process within contextually-based data, since the box plot is a representation that compresses and obscures information.

Element 9: Evaluative As each of the eight elements is considered, the evidence provided by that element is described, assessed on its strength. and weighed up in the process of making a judgment on the data. For example, for teaching episode one, a description would be "the male graph is *higher* than the female graph," whereas the strength of the evidence is conveyed by "the male graph is a *lot higher* than the female graph." Weighing the evidence is conveyed by statements such as "*even though the graphs overlap* these data suggest males on average earn more than females." In teaching episode two, the teacher describes, assesses and then weighs all the evidence she has accumulated. The language of this evidence is italicized.

> Teacher: Now I know the numbers are different, the males are *bigger* than the females, but *it's not that different*, it's not like one's 100 and the other's – you know? So, it's another contributing thing, men's stuff is *more spread out*. But *it's not massively different*, especially when you see it on the graph, you know, *it's not that different,* can you accept that? Okay, so at the moment, I've *got some conflicting kind of information*, right median – females are *more* clever, but when I look at the whole graph, the whole graph's a *bit more higher* for males. They're a little bit different in their spreads but you know, so *I'm still not ready to say yes* males have got a higher IQ than females.

The weighing of evidence involves qualitative judgments and a subtle use of language to convey how a decision is being reached. Since informal inferences are being made it may be hard for students to determine in inconclusive situations what evidence is taken notice of by the teacher when making a decision (see Figure 2(c)).

Element 10: Referent When the teacher is comparing two distributions represented as box plots in a symbolic system, then reasoning from this symbolic system necessitates a constant reference to other systems. The box plots are constantly being decoded in a back-and-forth switching between the visual symbol system and the concepts and ideas to which it refers. For example, the teacher in the spread element of reasoning decodes the visual system, a rectangular box divided by a line with a whisker at each end, when she imagines a quarter of the females standing in each section. Such an imagining, with some females standing closer together than others, is a switch to another reference system or another representation of the box plot. Her main referents were the context or the statistical measures the symbol system was portraying. For example, she said "*female* graph is higher," "*male earnings* are higher, or "*female median* is higher." Sometimes her referent was the imagined underlying distribution of the data, "this central box here gives me the *middle 50% of the data*." Her language did not refer very often to the underlying plots, which had been summarized by the box plots. Furthermore, her referents to the data plot for her justifications in the written conclusions (Figures. 1(d), 2(c)) seem to be insufficient.

3.3. LIMITATIONS

There are two main limitations to this research. First, the study has only captured one teacher learning to communicate her reasoning from box plot distributions. Second, one researcher categorized the elements and hence there is no triangulation from independent sources, although the teacher did take the opportunity to assess the interpretation. Because there seemed to be no account of how to draw informal inferences from the comparison of box plots at an introductory level, an action-research method, where learning to reason occurred in and from practice, was deemed appropriate. Hence, the research can only offer some insight into possible ways teachers and students could be expected to reason informally and into possible pitfalls in the reasoning process. The research also makes the case for developing sampling reasoning concepts and keeping data with the box plots, but again this is based on one teacher's reasoning. Therefore the discussion that follows draws on the literature from students' reasoning to support some findings but remains speculative in terms of teachers' reasoning.

4. DISCUSSION

Informal inference should be stimulating intuitive foundations for formal inference. Making informal inferences based on distributions alone is not the usual statistical practice and hence the teacher in this study should be viewed as a learner in a new situation struggling to convey the messages in data. Indeed Biehler (1997, p. 176) stated that "there are profound problems to overcome in interpreting and verbally describing statistical graphs that are related to the limited expressability of complex quantitative relations by means of a common language" and that researchers need to become more aware of the difficulties.

The key finding from this research is the proposal of a descriptive model (Figure 3) of reasoning from box plots. The model is complex and is the beginning of an exploration into the elements of reasoning that could be considered when structuring teaching towards formal inference. The elements, hypothesis generation, summary, shift, signal, spread, and individual case, have been described by other researchers. The shift element could be incorporated into another element and may not be as important as the others in the reasoning process. The summary element, especially the comparison of equivalent five-number summaries, may be considered unimportant but nevertheless such reasoning does exist and the purpose is to document all the types of reasoning invoked. The main findings from this research, however, are the description of ways in which the elements of sampling, explanatory, evaluative and referent are also part of a reasoning process that leads towards formal inference. Each element will now be discussed.

The hypothesis generation element is an aggregate-based reasoning approach that Konold, Pollatsek, Well, and Gagnon (1997) and Ben-Zvi (2004) believe is essential if students are to reason about trends and patterns in distributions. The teacher's discourse incorporates such notions. Her reasoning also highlights the link between the nature of the representation and the nature of the reasoning, particularly in the summary and shift elements. The teacher's reasoning is led intuitively towards comparing the five-number summary boundaries, a facet of reasoning Biehler (2004) and Friel (1998) noticed in their students. The shift element documented by Biehler (2004) appeared to be intuitively inherent in the visual nature of the representation and hence in the teacher's reasoning.

Reasoning with measures of center is expressed by the teacher in the signal element. Both Bakker (2004) and Biehler (2004) report that the median as a representative value of a distribution is difficult to develop. When Bakker (2004) and Konold et al. (2002) searched for an alternative notion of center, they hypothesized that students' intuition of middle group or modal clump could support the development of center as being a characteristic of the distribution. The teacher does use the middle 50% of data as an intuitive device for the signal. Also the nature of the representation leads to this type of argumentation.

Within the spread element of reasoning by the teacher, two comparisons are evident: comparing the densities within one box plot and comparing the densities between the two box plots. Such a discussion was not clear to students nor was the purpose of the discussion of how comparing spreads helped in making an inference. Biehler (2004) noted that his students did not comment on spread differences. Another problem with the spread element, which is closely aligned to the nature of the representation, is how concepts can be built up for viewing spread as a dispersion from the median, which according to Bakker (2004) is a big transition. When the teacher compared the overlap of the boxes with drawn lines, she was taking into account some of the variability. The question is whether such a comparison could be conceived as an intuitive beginning for confidence interval ideas for true population medians and for viewing spread as dispersion from the median.

The sampling reasoning element is presented by the teacher via thought-simulations rather than by empirical simulations in which students could actually experience the variability of samples drawn from populations (Pfannkuch, 2005). Both verbalization and experience of sampling behavior are necessary if teachers and students are truly to grasp the nature of sampling reasoning. Moreover, this element is key to bridging students towards formal inference. The game to be played is game two whereby the reasoning involves making inferences about populations from samples, not making inferences about samples, game one. The teacher did not resolve which game she was playing and therefore a large part of inferential reasoning eluded her and her students. In order to play game two, activities, such as “growing a sample” (Bakker & Gravemeijer, 2004), bootstrapping (Finzer, personal communication, 7 July 2005), and experiencing and building concepts about sampling behaviour (Pfannkuch, 2005) could assist in developing her and her students’ sampling reasoning. Reasoning about samples also includes how the sample was selected and sample size (Watson, 2004). Although the teacher referred to sample size, she did not discuss how the sample was selected as that information was not presented as background information, but she was careful in acknowledging who was sampled and on whom she could draw an inference. In other words she paid attention to inference space judgment.

For the explanatory element a way of perceiving her reasoning is to consider that the distributions are a statistical model of a real world situation. Since contextual knowledge is essential for seeing and interpreting any messages in data, a continuous dialogue should exist between the statistical models and the real world situation. Hence features seen in data produce queries about context, which in turn suggest questions for the data (Wild & Pfannkuch, 1999). This continuous shuttling between the contextual and the statistical is present in the teacher’s reasoning. Her choice of learning task for teaching episode one, however, illustrated how lack of background information about data leads to speculation about the data rather than further exploration. Context is used by the teacher as an integral part of the interrogation of data, as a factor in determining whether confounding variables are present, and for determining whether there are alternative explanations for the findings. Friel et al. (2001, p. 140) also highlight that the contextual frame of data is necessary for comprehending and making judgments on graphs although it increases “the number of elements to which the graph reader must attend.”

The moderating elements of reasoning, the evaluative and referent elements, act as anchors for weighing the evidence and for interpreting an abstract box plot representation respectively. The evaluative element includes making a judgment by comparing distributions and is alluded to by Friel et al. (2001) in their suggested taxonomy of judgment tasks. For actually making an informal inference this element is critical. Qualitative judgments on the whole must be made to ascertain whether one is prepared or not prepared to state Group A is greater than Group B, on average. Within each of the eight elements of reasoning, the teacher is continually making qualitative and sometimes quantitative statements as a prelude to weighing the evidence. Weighing the evidence is a matter of opinion, can be subjective, and rests on experience with data. The students’ lack of experience and seemingly innate need for a definite conclusion (see sampling element dialogue) may militate against realizing that in statistics findings may be inconclusive.

Bakker and Gravemeijer (2004) consider referents as being essential for instructional design. The symbol system, the box plot, is a new representation, and students may need to interpret it with a better-known system such as a dot plot where individual data are identifiable (Carr & Begg, 1994). Friel et al. (2001, p. 139) also note that a "major component of the graph reader's interpretation process is relating graph features to their referents." The teacher's referents are many-fold, each acting to place the abstract representation into a context as well as to imagine the data underneath the box plots. For someone with her experience there may be no problem in imagining the plot underneath, but for the students the abrupt transition from the stem-and-leaf plot to the box plot may have been too fast (Pfannkuch, 2006).

According to Moore (1990) and Wild and Pfannkuch (1999), variation is at the heart of statistical thinking. All the elements are underpinned by variation as it is noticed, dealt with, measured informally, and explained. Or as Finzer (personal communication, 7 July 2005) more succinctly stated, "distribution reasoning is the recognition and utilization of patterns in variability." Reasoning about distributions is more than reasoning about shapes (Bakker & Gravemeijer, 2004), it is about decoding the shapes (Friel et al., 2001) by using deliberate strategies such as the proposed model (Figure 3) to comprehend distributions. Furthermore, there is a weighing of evidence to form an opinion on and inference from the information contained in the comparison of distributions. Such informal decision-making under uncertainty requires qualitative judgments, which would seem to be much harder than the quantitative judgments of statistical tests.

The analysis of one teacher's reasoning from box plot distributions contributes to the research base by enhancing understanding of the reasoning processes, and raising issues about the links to formal inference, the nature of the game being played, and instructional practice. The model (Figure 3) demonstrates the richness of verbalization necessary for communicating ideas and concepts from box plot distributions, and builds on other research findings. Thus the model begins to propose a coherent framework for the nature and type of informal inferential reasoning that might be addressed when teaching students how to reason when comparing box plot distributions.

5. IMPLICATIONS FOR RESEARCH AND EDUCATIONAL PRACTICE

More research work is needed on designing instruction and building teachers' and students' concepts and reasoning about distributions towards formal inference. Research is also needed on developing teachers' and students' sampling conceptions in terms of learning to reason about populations from samples using informal inference. Since this research is based on one teacher's reasoning in a non-technological environment, there may be other reasoning elements necessary for informal inference. The challenge for future research is to move towards a prescriptive model of reasoning from box plot distributions. Such a model could specify how the different reasoning elements could be woven and sequenced together during instruction and exemplify how the elements contributed towards the development of formal statistical inferential reasoning.

At the teaching level, the implications from this research suggest that developing teacher and student talk on how to communicate ideas and on concepts represented in distributions are essential. The model suggested by this research has now been used as a guide in developing teacher reasoning and for writing down how to reason from box plots. Instruction, however, needs to adopt a gradual transition approach from dot plots to abstract box plots to improve the referent element of reasoning and to build the sampling reasoning element through giving teachers and students opportunities to experience sampling behavior. Such an opportunity was taken by the teachers in this project in 2006.

Hill, Rowan, and Ball (2005) believe that teachers' mathematical content and pedagogical content knowledge are linked to student achievement and that improving teachers' mathematical knowledge will improve students' understanding. Teachers and researchers need to collaborate to develop a coherent, deeper conceptual approach to the learning of statistics. A research agenda should be implemented since the current situation in teaching and assessment requires teachers and students to make informal inferences from the comparison of distributions. Without an underlying research base on informal inference and reasoning from distributions, this situation may lead to some unforeseen consequences in later years of schooling.

REFERENCES

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Utrecht, The Netherlands: CD-ß Press, Center for Sci. and Math. Education.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education (IASE) Roundtable,* Lund, Sweden, 28 June-3 July 2004, (pp. 163-173). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications.php]

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.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass publishers.

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

Biehler, R. (1997). Students' 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]

Biehler, R. (2004, July). *Variation, Co-Variation, and Statistical Group Comparison: Some Results from Epistemological and Empirical Research on Technology Supported Statistics Education.* Paper presented at the 10th International Congress on Mathematics Education, Copenhagen.

Carr, J., & Begg, A. (1994). Introducing box and whisker plots. In J. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics (ICOTS 4)* Marrakech, Morocco, July, 1994. Minneapolis, MN: University of Minnesota.

Friel, S. (1998). Comparing data sets: How do students interpret information displayed using box plots? In S. Berenson, K. Dawkins, M. Blanton, W. Coulombe, J. Kolb, K. Norwood, & L. Stiff (Eds.), *Proceedings of the Twentieth Annual Meeting, North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 365-370). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124-159.

Hill, H., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*(4), 259-289.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151-168). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/8/13.Konold.pdf]

Konold, C., Robinson, A., Khalil, K., Pollatsek, A. Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society*, Cape Town, South Africa, July, 2002 [CD-ROM.] Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/8b2_kono.pdf]

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

Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). Dordrecht, The Netherlands, Kluwer Academic Publishers.

Pfannkuch, M. (2006). Informal inferential reasoning. *Proceedings of the Seventh International Conference on Teaching Statistics: Working cooperatively in statistics education,* Salvador, Brazil, July, 2006 [CD-ROM.] Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf]

Pfannkuch, M., Budgett, S., Parsonage, R., & Horring, J. (2004). *Comparison of data plots: building a pedagogical framework*. Paper presented at ICME-10, Denmark, TSG11: Research and development in the teaching and learning of probability and statistics.
[Online: www.stat.auckland.ac.nz/~iase/publications/11/Pfannkuch.doc]

Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education (IASE) Roundtable,* Lund, Sweden, 28 June-3 July 2004, (pp. 163-173). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/rt04/5.1_Pfannkuch&Horring.pdf]

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Watson, J. (2004). Developing reasoning about samples. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

MAXINE PFANNKUCH
Department of Statistics
The University of Auckland
Private Bag 92019
Auckland, New Zealand

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) = $1/2 \times 1/2 \times 1/2 \times 1/2$.
(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).

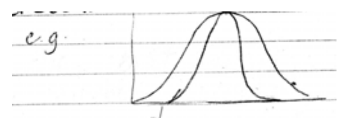
SR9 The means of the 100 sample sizes are going to have *a greater variance* than that of the whole population, the bigger the sample, the closer it is to the mean.

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.

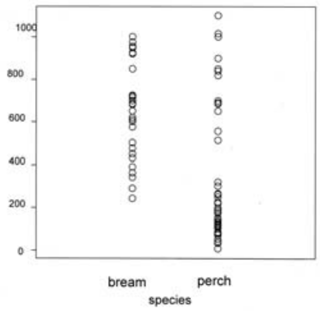
Wild, C. J., & Seber, G. A. F. (2000). *Chance encounters: A first course in data analysis and inference.* New York: John Wiley & Sons.

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

APPENDIX: MINUTE PAPER QUESTIONS

Minute Paper 1a

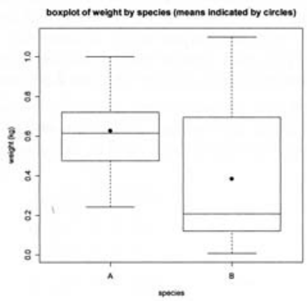
Look at the following plot. It shows the weights in grams of two species of fish (bream and perch).



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.

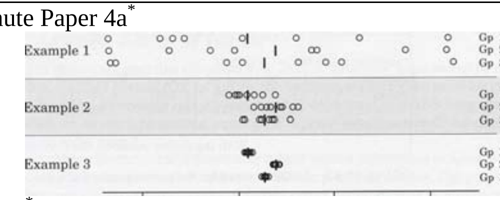
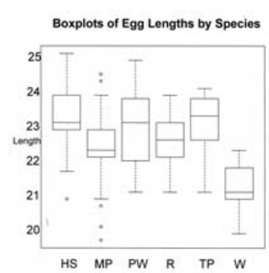


*Diagram from Wild & Seber (2000, p. 439)

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

THE ROLE OF CAUSALITY IN THE CO-ORDINATION OF TWO PERSPECTIVES ON DISTRIBUTION WITHIN A VIRTUAL SIMULATION

THEODOSIA PRODROMOU
Centre for New Technologies Research in Education, University of Warwick
t.prodromou@warwick.ac.uk

DAVE PRATT
Centre for New Technologies Research in Education, University of Warwick
dave.pratt@warwick.ac.uk

ABSTRACT

Our primary goal is to design a microworld which aspires to research thinking-in-change about distribution. Our premise, in line with a constructivist approach and our prior research, is that thinking about distribution must develop from causal meanings already established. This study reports on a design research study of how students appear to exploit their appreciation of causal control to construct new situated meanings for the distribution of throws and success rates. We provided on-screen control mechanisms for average and spread that could be deterministic or subject to stochastic error. The students used these controls to recognise the limitations of causality in the short term but its power in making sense of the emergence of distributional patterns. We suggest that the concept of distribution lies in co-ordinating emergent data-centric and modelling perspectives for distribution and that causality may play a central role in supporting that co-ordination process.

Keywords: *Distribution; Causality; Randomness, Probability; Variation; Microworld design; Emergent phenomena*

1. TWO PERSPECTIVES ON DISTRIBUTION

Distribution is commonly recognised as one of the key ideas in probability and statistics, certainly at secondary school level. For example, in the UK National Curriculum (DfES, 2000), students at lower and upper secondary level are expected to "compare distributions and make inferences, using the shapes of distributions and measures of average and range." Higher achieving students should be able to extend this to other measures of spread and understand frequency density. The assessment regime in that National Curriculum implies that the above statements refer to distributions of data, either prepared for students or generated through experiments and surveys.

The introduction of digital technology into schools has prompted interest in Exploratory Data Analysis (EDA) as a means of engaging students in statistical analysis, arguably reducing the need for a sophisticated understanding of theoretical statistical principles, demanding an appreciation of probability theory, prior to meaningful engagement. The technology is ideally suited for supporting students as they manipulate data and portray it in a range of different representations in order to infer underlying

Statistics Education Research Journal, 5(2), 69-88, http://www.stat.auckland.ac.nz/serj

trends. The EDA approach then promotes a perspective on distribution as a representation of collections of actual data, consistent with the goals of the National Curriculum.

Previous research has conceived of distribution as "an important part of learning to look at the data" (Moore, 1990, p. 106) and as an organising conceptual structure with which we can observe the aggregate features of datasets rather than just a collection of individual values (Cobb, 1999; Petrosimo, Lehrer, & Schauble, 2003). Other researchers have claimed the centrality of the concept of data as an aggregate which is characterised by core features that are invisible in any of the individual elements in the aggregate (Konold & Higgins, 2003; Mokros & Russell, 1995). Students, however, have a strong attachment to the case-oriented view; in other words, data are perceived as a collection of individual data values or cases (Wilensky, 1997; Ben-Zvi & Arcavi, 2001). To help students move beyond the case-oriented view, Hancock, Kaput and Goldsmith, (1992) claimed that it is prerequisite for students to mentally construct such an aggregate, before they can see the dataset as a whole. We consider the above approaches to be taking a *data-centric* perspective on distribution. A data-centric perspective on distribution pays attention to the variation and shape of data that has been collected, perhaps through a sampling process.

Petrosino et al. (2003) have suggested that students need to conceive of distribution "as an organizing conceptual structure for thinking about variability located within a more general context of data modelling" (p. 132). Bakker and Gravemeijer (2004), in their attempt to investigate the relationship between data as individual values and distribution as a conceptual entity, examine key aspects of both datasets and distributions such as centre, spread, density, and skewness. They propose a three-level structure; the lowest level comprises of distribution as a set of data values, and the highest level recognises the conceptual entity, distribution. Between these two levels, they position summary statistics such as centre, spread and skewness. They imagine that this structure can be read both *upwards* and *downwards*. In the upward perspective, students tend to perceive data as a series of individual cases, which they can use for calculations of any sample statistics (mean, median, etc.). In the downward perspective, students should look at the data with a notion of distribution as an organising structure, conceiving centre, spread and skewness as features of that distribution. The upward perspective leads to a frequency distribution of a dataset. In the downward perspective, alternatively, theoretically derived distributions, such as the Normal and other probability distributions, are typically used to model data. Bakker and Gravemeijer (2004) chose to deal informally and consistently with core ideas, such as variation and sampling but with distribution still being in a central position. They also envisioned that informal consideration of the shape is the basis for reasoning about distributions. Perusal of recent research literature suggests that reasoning about variation and distributions are strongly associated (Bakker, 2004; Ben-Zvi, 2004; Makar & Confrey, 2003) since "without variation, there is no distribution" (Bakker and Gravemeijer, 2004, p. 149)

However, the notion that variation generates distribution is only part of the story, and it is that part told from the perspective that recognizes what we are terming the data-centric perspective, in which distribution is seen as a collection of data results. Compare this perspective to that of the classical statistician, who accounts for unexplained variation as that part of a hypothetical model which is not apparently associated with a main effect. Here the emphasis is on a model and so we refer to this approach as the modelling perspective on distribution. Indeed, from this perspective, we might reverse Bakker and Gravemeijer's aphorism to state "without distribution there is no variation."

When we refer to theoretical distributions (for example, Normal, Uniform and Binomial), we idealise mathematical models, in which we attribute probabilities to a range of possible outcomes (discrete or continuous) in the sample space. In this modelling approach, the model gives rise to variation. Data distributions are seen as variations from the ideal model, the variations being the result of noise or error randomly affecting the signal or main effect, as reflected in the model itself. The modelling perspective on distribution pays attention to randomness and the shape of the probabilities that mould the outcomes, perhaps through some experiment. The modelling perspective reflects, in our view, the mindset of statisticians when applying classical statistical inference. Indeed, Borovcnik (2005) offered six variations on the notion of data being modelled as a main effect together with an error:

(i) Signal + Error	(ii) Pattern + Deviation	(iii) Fit + Residual
(iv) Model + Residual	(v) Explained + Unexplained Residual	(vi) Common + Specific causes

In the same vein, Konold and Pollatsek (2004) viewed the data as a combination of signal and noise, where the signal can be an average value with variation as noise around it. They argued that "the idea of distribution comes into better focus when it is viewed as the distribution around a signal" (p. 171). Bakker (2004), in turn, referred to a second type of signal in noisy processes or shape as a pattern in variability. Bakker (2004) viewed signal as a distribution, such as the shape of a smooth bell curve of the normal distribution, with which we model data. He suggested that the noise in that case is the variation around that smooth curve. The idea of "signal" and "noise" is evident in several research studies (Biehler, 1994; Wild & Pfannkuch, 1999; Noss, Pozzi & Hoyles, 1999).

In our view, a sophisticated view of signal and noise requires a co-ordination of the data-centric and modelling perspectives. We argue that the emphasis in the UK National Curriculum, and indeed as apparent in EDA approaches, is insufficient alone to nurture such co-ordination. We dream of a pedagogy which somehow enables students to appreciate the connection between the data-centric and modelling perspectives on distribution.

The development of such a pedagogy demands that we research the design of tools that aim to facilitate the co-ordination of these two perspectives. Our approach is to adopt a design perspective in which we develop a software-based task to act as a window on thinking-in-change (Noss & Hoyles, 1996). In looking to bootstrap the iterative design process, we found immediate resonance with research on emergent phenomena (Wilensky, 1997), which contain a sense of "organised randomness" (Davis & Simmt, 2003) and a tension between living within rule-defined boundaries and using the space created within those boundaries productively (Johnson, 2001). Thus, we began to think of the challenge of co-ordinating the two perspectives on distribution as one of seeing distribution as an emergent phenomenon (Prodromou, 2004). At the same time, we were alerted to the observation that there is a "centralised mindset" amongst students that may be rooted in a natural habit of interpreting phenomena in a cause-and-effect manner rather than in complex emergent terms (Resnick, 1991; Johnson, 2001; Gould, 2004). However, as we will see, we found that the tendency towards deterministic thinking was a useful resource for co-ordinating the two perspectives.

Our broad aim then is to understand better how students might conceive of data-centric and modelling perspectives of distribution. Furthermore, we aspire to develop environments in which meanings that embrace these two views of distribution might be constructed.

2. EMBRACING CAUSALITY

The modelling and data-centric perspectives on distribution offer different views of variation. In the data-centric perspective, data will spread across a range of values; in the modelling perspective, variation is portrayed as a random movement away from the main effect. In order to co-ordinate these two perspectives, we argue that it is necessary to see them as a duality that encompasses both the deterministic and the stochastic. We therefore examine research on how students apparently perceive the stochastic.

Piaget & Inhelder (1975) suggested in their seminal work that the organism eventually succeeds in inventing probability as a means of operationalising the stochastic. Prior to that achievement, random mixtures were unfathomable and the literature is abundant with examples of how even adults use various, often misleading, heuristics to make judgements of chance (for example, Kahneman, Slovic and Tversky, 1982). How is that process of operationalising the stochastic achieved? Clearly Piaget's constructivist stance would demand that we consider what students already know since therein must lie the resources for coming to appreciate distribution and other key stochastic concepts.

Pratt (2000) reported how students of age 11 years were able to articulate meanings about random phenomena which were remarkably akin to expert-like views in one respect. They understood the unpredictable, uncontrollable and unpatterned nature of randomness. These so-called *local resources* were brought to bear by these students in order to describe short-term randomness. Significantly, these same students were unable to demonstrate meanings for the predictable, controllable and patterned nature of long-term behaviour. Such *global resources* however began to emerge as these students engaged with specially designed tools, in an environment called *ChanceMaker*. This microworld consisted of mini-simulations of so-called gadgets, common random generating devices such as coins, spinners and dice. These gadgets were presented as not working properly and the challenge to the students was to mend them using tools made available within the gadgets. The students began to articulate situated versions of the Law of Large Numbers, such as "the more trials you do, the more even is the pie chart." The significance of this work for the present study is the causal nature of the students' global resources. The number of trials *determines* the state of the pie chart. Pratt (1998) discusses the notion of *phenomenalising*, the process of transforming mathematical ideas into *quasi-concrete objects* (Papert, 1996), which can be manipulated on-screen by the student, who can make sense of the mathematical concept through using it, much as most of us do, to come to appreciate everyday phenomena. By phenomenalising randomness, Pratt claimed that the students were able to exploit well-established knowledge about causality to *concretise* (after Wilensky, 1991) the Law of Large Numbers.

It is our conjecture that, given appropriate phenomenalised tools, students will be able to bridge the modelling and data-centric perspectives of distribution. In this vision, randomness becomes an agent that causes variation and in turn randomness can be "controlled" through parameters, perhaps instantiated as on-screen sliders, that experts might think of as measures of average and spread. Indeed the blurring of a distinction between a measure or representation and a control is one of the hallmarks of using technology to promote using before knowing (Papert, 1996).

It seems though that there is a paradox here. On the one hand, the work of Pratt (2000) makes a prima facie case that technologically-based environments may have the potential to offer a method of constructing meanings for distribution out of causality. On the other hand, such an approach may reinforce the centralised mindset and militate against the construction of distribution as an emergent phenomenon that bridges the data-centric and modelling perspectives of distribution. This apparent paradox lies at the heart of our work. We must design an environment that supports students in discriminating and moving smoothly between data as a series of random outcomes at the micro-level and the shape of distribution as an emergent phenomenon at the macro-level. In that respect, we conjecture that we can build an environment that enables the student (i) at the micro-level to use their understanding of causality whilst at the same time begin to recognise its limitations in explaining local variation, and (ii) at the macro-level to use the parameters as causal agents to appreciate the impact of those sliders on features of the distribution whilst appreciating their failures to completely define the distribution. We intend to use the microworld that embodies these conjectures not only to test those conjectures but further as a window on the evolution of students' thinking about the two perspectives on distribution. Through that window, we ask whether and how students co-ordinate the data-centric and modelling perspectives on distribution.

3. METHOD

Approach and tasks To elaborate this research question, we aimed first to instantiate the conjectures into a microworld that would perturb the students' thinking and act as a window on that thinking-in-change (Noss & Hoyles, 1996). Learning situations are complex ecologies in which many variables interact. Experimental methodologies are often impossible, either because of the confounded nature of the variables or for ethical reasons. We have found in previous work that the delicate process of phenomenalising a mathematical concept in order to observe thinking-in-change demands a gradual sensitising towards that complex ecology. The current ongoing study therefore falls into the category of *design experiments* (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), whereby we gain insights about both thinking-in-change and design issues from the participants' interactions during the iterative design of the microworld. Typically design experiments require several iterations. Each iteration raises new issues about the learning process and generates conjectures about how the design may better help to elaborate the research question.

In this article, we report on pupils' interactions with the third iteration of the microworld. A major issue raised by the first iteration was that the design at that time failed to generate purposeful student activity. In order for it to act as a window on students' thinking-in-change about distribution, it was essential that they were able to explore with relatively little input from the researchers. We therefore searched for a context that might stimulate such activity whilst at the same time encourage focus on distribution as a central concept. In fact, our design strategy has subsequently been influenced by the notions of *Purpose* and *Utility* (Ainley, Pratt & Hansen, 2006). We needed to provide students with a setting that would inspire a deterministic interpretation of behaviour and would find purpose in adopting a perspective that sees behaviour captured and explained in terms of emergent distributions. We have approached the problem by setting the exploration in the second iteration in the context of playing basketball. Our observations during the second iteration alerted us to the significance of the two different perspectives on distribution and the need to find a design that might support their co-ordination.

This article reports on a case study of how two pupils interacted with the third iteration of the microworld, in which they were presented with the basketball-throwing activity depicted in Figure 1. Students were first asked to throw the ball into the basket using various sliders that control the throw. Once this preliminary task was completed, they were asked whether they felt that the simulation was realistic. This normally generated the response that it seemed artificial that the ball was entering the basket every

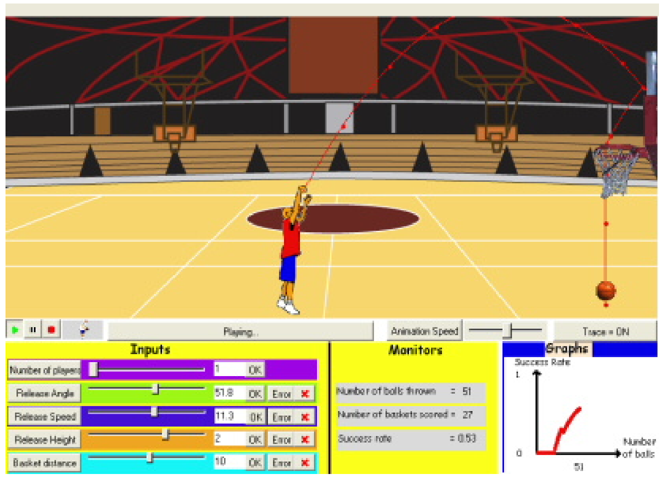


Figure 1. The player has successfully thrown the ball into the basket. The release angle, speed, height and distance can all be varied using the sliders or by entering the data directly. Once the play button has been pressed, the player continues to throw with the given parameters until the pause or stop button is pressed. The trace of the ball can be switched off. Feedback is shown in the Monitors and Graphs panes.

time. Since the system was completely determined at this point, the ball replayed faithfully its successful path on every throw. The subsequent discussion typically introduced notions such as skill-level and we showed them the error buttons as in Figure 2 which can make the situation more realistic by allowing for errors in throws.

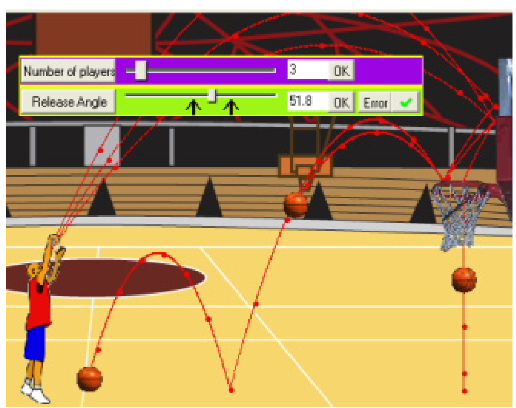


Figure 2. Three players each throw their ball simultaneously. Because the error button has been pressed, the balls vary their paths. Only one of the three throws is successful. As the three players continue throwing, the release angles will average 51.8 degrees.

The students were able to control the spread of the error through two arrows on the slider, which corresponded to points that were roughly two standard deviations above and below the mean average as in Figure 2. The students were able to move either or both of these arrows, generating values that corresponded to distributions with differing spreads and bias. The microworld also allowed the students to explore various types of graphs relating the values of the parameters to frequencies and frequencies of success. When the parameters were determined, the graphs appeared as single bar columns as in Figure 3. When the error had been introduced, the graphs appeared as histograms as in Figure 4.

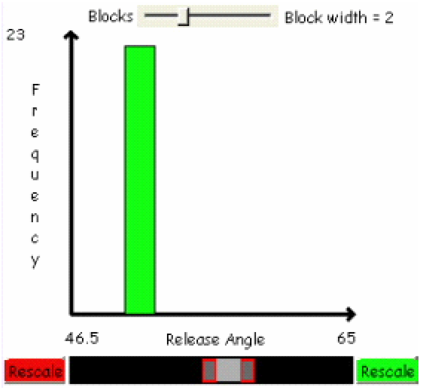


Figure 3. With no error set, the graph will appear as a single bar. The graphs can be rescaled using the button/sliders.

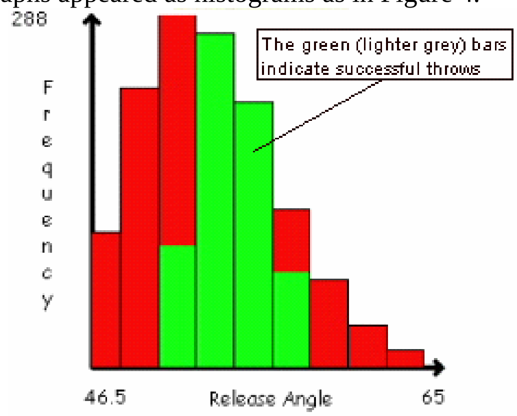


Figure 4. With error set, the green (lighter) bars in the histogram indicate the successful throws; the red (darker) bars show the remainder.

(We do not wish to enter here into the debate about whether these graphs should be referred to as histograms. With equal width bars, the matter is of no real consequence.)

Subjects The microworld was trialled with six students in a UK secondary school. The first two pairs of students were, according to their teacher, of average ability. One pair was age 14; another pair was 15. The third pair, 16 years of age, was at an early stage of advanced level study of mathematics and was of above average ability. In this paper we report on the emerging insights of the two 14 year-old students, Tom and Chris. Although we recognize the same issues as reported below from analysis of other pairs, Tom and Chris provide in our view the clearest illustration so far of the co-ordination of the two perspectives on distribution.

Procedure The pair was observed by both authors and the programmer, who had coded the microworld in Imagine Logo, a powerful version of Logo, published by Logotron (www.logo.com/imagine/). For the purposes of this paper, we refer to all three as the researchers. The episode described below took place in one session lasting about 90 minutes. The data collected included audio recording of the students' voices, video recording of the screen output on the computer, and researchers' field notes. The analysis was one of progressive focussing (Robson, 1993). At the first stage, the recordings were simply transcribed and screenshots were incorporated as necessary to make sense of the transcription. Subsequently, the first author turned the transcript into a plain account with no explicit interpretation other than through selection of the more promising sections. The less interesting sections were replaced with discursive descriptions of what happened. At third stage, an interpretative account was written by the first author and discussions about the validity of those interpretations took place with the second author. In this respect, we followed Mason's (1994) advice to make an account of the data before accounting *for* the activity. At the fourth stage of analysis, issues were extracted and turned into conjectures for use in the next and ongoing iteration of the design cycle.

4. FINDINGS

Our analysis suggests that Tom and Chris's meanings for distribution were co-ordinated through four distinctive phases, which we use below to structure the story of how the relationship between causality and variation shifted as they moved through these phases.

4.1. PHASE 1: DETERMINING A SUCCESSFUL THROW

Tom and Chris were introduced in the microworld to a single basketball player who was clearly failing to throw the ball successfully into the net. However, they were shown that his throws could be changed using the various sliders. They were challenged to improve the player's throws.

They began to vary the sliders for release speed and angle as well as height and position. They demonstrated sophisticated intuitions for altering speed and angle in such a way that the path of the ball was gradually moving nearer to the basket. Within two minutes, they had successfully set the sliders to throw the ball into the net (Figure 5). Tom and Chris continued to explore other successful throwing positions by moving the player and finding the corresponding successful release speed and angles. Although Tom and Chris had found an initial successful throw quickly, they appeared to enjoy exploring other values of the parameters that also resulted in a successful throw. It turned out later that it was important that as researchers we allowed Tom and Chris this space to become comfortable with the software, moving the sliders in a playful and exploratory way.

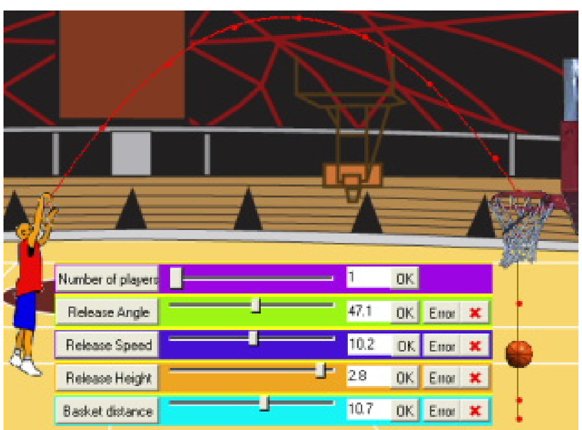


Figure 5. Within two minutes, Tom and Chris had managed to find values for the parameters that caused the player to throw the ball directly into the basket.

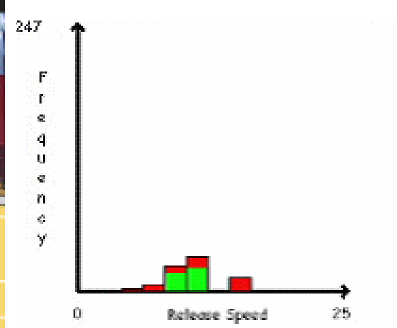


Figure 6. The graph of the release speed, based on 77 throws, shows variation caused by the two boys changing the parameters during the data collection.

Meanwhile, the data for all throws was being continuously collected by the computer since at no stage did Tom and Chris reset the data collection process by stopping the experiment. On two occasions during this phase, the researchers asked the boys to look at the histograms. Although no errors had been introduced, the graphs showed variation. This variation had been created by the manual changing of the parameters during the play (Figure 6). In fact, Tom and Chris did not comment on this variation and the researchers did not probe into the boys' understanding of this aspect. Nevertheless, we suspect that, in the light of the later developments, their own role as agents of variation was an important feature of how they later understood variation in which they were *not* the agents.

4.2. PHASE 2: EXPLORING THE ARROWS

After 24 minutes, the researchers began to introduce the notion of error. It was suggested to Tom and Chris that the previous simulation was not very realistic since, once the correct values had been discovered, the player was successful every time. The boys were introduced to the error buttons. They observed how, when the error button was pressed, two arrows appeared either side of the handle on the corresponding slider. They began to explore the effect of moving these arrows but found it difficult to make sense of what the arrows were doing. (In the transcript, Res refers to the researchers.)

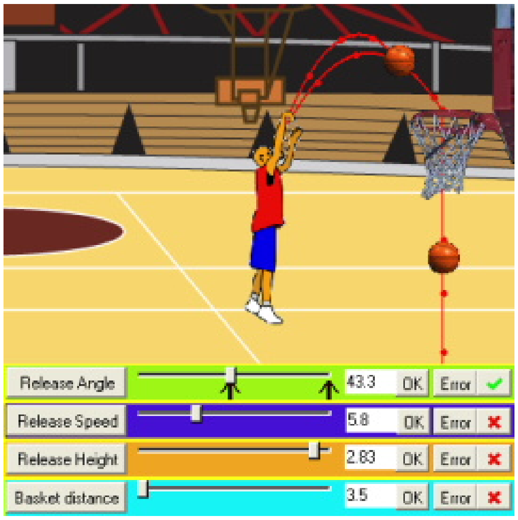


Figure 7. Tom noticed that the path was slightly different even though they had made no changes.

(1) Tom: They *(referring to the arrows)* might help our decision.
(2) *Tom and Chris spent three minutes moving the arrows around whilst the player threw the ball continuously.*
(3) Chris: It doesn't really make it any easier.
(4) Res: Why?
(5) Chris: I can't see how it is improving our chances.
(6) Res: What do you think *(looking at Tom)*?
(7) Tom: Same.
(8) Chris: Still… I'm trying to work out what these arrows are.

At first the boys appeared to associate error with something being wrong, probably associating the word itself and the cross on screen with marking of their work in class. Although confused, the boys continued to try to make sense of the role of the arrows and we did not seek to clarify. After allowing the simulation to run for about two more minutes, Tom noticed that the path of the ball changed.

(9) Tom: It's like… when it's throwing the ball, it's changing occasionally… yeah, like then *(as he was talking the ball took a different path, as in Figure 7)*… it just went differently.
(10) Chris: Could that be to do with the arrows?

A few seconds later, the researchers suggested that they look once more at the graphs.

(11) Res: Can you understand these graphs?
(12) Tom: Quite a lot of angles… spread out.
(13) Chris: We tried a lot of angles. We kept adjusting the angles.
(14) Res: Why?
(15) Chris: I don't know. We were just playing with it.

Tom and Chris were right in that they had indeed been adjusting the sliders during the simulation and that some of the variation in the angles was caused directly by them. However, at the same time, the angles were being chosen randomly by the computer and so some of the variation was due to randomness. It would appear that Tom and Chris recognised variation in this setting where they personally were the agents (lines 13-15) and were possibly entertaining the idea that the arrows may also somehow be involved (line 10).

4.3. PHASE 3: ARROWS AS AGENTS

Half an hour into the session, the researchers suggested that Tom and Chris might begin a new experiment in order to explore more systematically the role of the arrows. They began with error set for release angle but soon introduced error for speed as well (Figure 8).

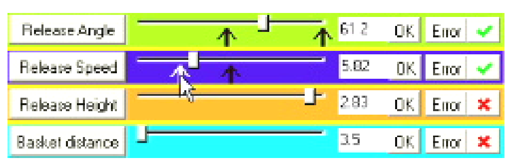


Figure 8. Tom and Chris explored the arrows further by introducing error to angle and then speed.

(16) Chris: The arrows do change it. *(They moved the two arrows closer together)*
(17) Tom: May be… might be between the two arrows… might be…
(18) Res: What might be?
(19) Chris: The release speed... might just be like random.
(20) Tom: It seems to be a bit inconsistent.
(21) Res: Why?
(22) Tom: The shots are like changing… even though that's not changing.
(23) Res: Even though what's not changing?
(24) Tom: The shots are… like that missed *(referring to the simulation in which the throw missed the basket)*… that missed, that one went in *(referring to the following throw)*… but we are not changing them from there *(pointing to the sliders)*.
(25) Res: Not changing the slider. Is that what you're saying?
(26) Tom: Yeah.
(27) Res: But even so sometimes it's missing.
(28) Tom: Yeah.

Our interpretation of this incident is that for the first time, Tom and Chris explicitly recognized that variation could occur without them acting as the agents of change (lines 22-28) and that this insight was accompanied by the preceding recognition that the angles or speeds might be chosen randomly from values between the two arrows (lines 17-20). These two ideas were themselves preceded by an acknowledgement that the arrows did in fact seem to be changing something (line 16). Perhaps such ideas had been gradually growing (line 10). A causal link between the arrows, randomness and scoring or missing seemed now to be postulated. However, such a link was not as yet explicitly established in their minds. Tom and Chris continued to explore, setting errors on and off and moving the arrows around for all of the variables, often simultaneously.

(29) Res: So what conclusions have you got so far about these arrows?
(30) Tom: A bit troublesome.
(31) Chris: Like a random number between the two arrows.

The two boys admitted that they were not yet confident about the idea that the arrows demarcated a region from which a random number would be chosen and so they continued to explore. There was now another intense period of exploration in which they moved the arrows around, sometimes close together, sometimes wide apart, sometimes symmetrically around the handle, sometimes asymmetrically. There did not appear to be much systematicity about this exploration. However, they constantly reviewed the continuing action in the simulation as they tried to make sense of the arrows. Eventually they made a breakthrough.

(32) Tom: If it's close, it's more chance of going in.
(33) Res: What do you mean?
(34) Tom: When the arrows are close together, it's got more of a chance of going into the net.

In lines 32-34, Tom and Chris seemed to have spotted this pattern of behaviour by looking closely at the effect on the animation of moving the arrows. They did not refer to any graphs during this period. A few minutes later however (line 35), they did decide to look at the graphs. Initially, all the histograms appeared to consist of a single bar (Figure 9).

Line 35 is a revealing remark. Chris was content that the graphs apparently revealed no variation since he was still relating variation to changes that they personally had generated (line 35). The causal link relating the arrows to changes in the animation had not been extended to the distribution in the graphs.

The researchers however noticed the small bar to the right of the main bar in the speed graph (see bottom left graph in Figure 9). They helped the boys to rescale the histogram of release speed, at the same time changing the block width (Figure 10).

(35) Chris There's only one because we haven't altered it.
(36) Res: So what does that graph tell you?
(37) Tom: Lines in the middle *(referring to the green (lighter) success bars).*
(38) Res: What about the reds and greens?
(39) Tom: There's an area which is green.
(40) Chris When we get the release speed to… whatever value that is *(pointing to the start of the green area)*… about 11… they're all successes… and when we change it, it's going to miss.
(41) Res: Did you change it?
(42) Chris Did you *(looking at Tom)*?

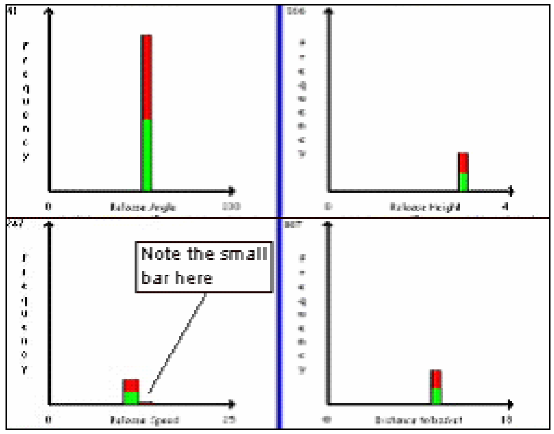


Figure 9. Tom and Chris supposed that the graphs were single bars because they had not made any alterations. They had not yet linked variation in the graphs to variation caused by the arrows. Careful inspection of the speed histogram would have revealed a tiny amount of variation.

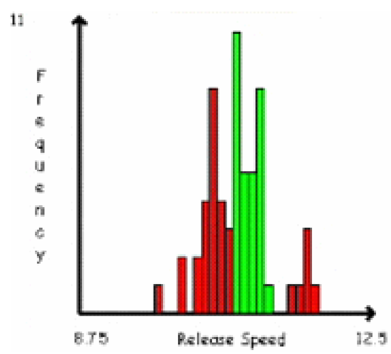


Figure 10. Tom and Chris needed to explain the variation in the histogram even when they had not themselves changed the speed slider.

(43) Tom: Which one?
(44) Chris Speed.
(45) Tom: That was the one with the arrows, wasn't it?
(46) Res: Why have we got different speeds on here?
(47) Tom: Because the arrows change it.
(48) Res: Explain that to him. I'm not sure he understands it yet.
(49) Tom: The arrows make it, like, random. So it's a random number between the two, I think.

It now seemed that Tom had abstracted a causal link between the variation in the histogram and the arrows (lines 47-49). Variation could occur even when they had not acted as agents. Instead the arrows acted as agents through some sort of unknown random mechanism.

One of the difficulties with design experiments is that the researchers are often unable to anticipate activity. Indeed, it is these unexpected outcomes that are often the most influential in shaping the design in the subsequent iteration. Our probing in lines 35-49 was certainly unplanned and as such leaves much unanswered. The precise manner by which Tom and Chris arrived at their conclusion remains a little mysterious. Of course, the strength of design research is that it allows, indeed encourages, such unexpected behaviour and is able to respond later by building such issues into the next design.

4.4. PHASE 4: MODELLING WITH THE ARROWS

Tom and Chris returned to their earlier notion that when the arrows were closer together, the chance of a successful throw was increased. They sought to create a realistic simulation of a player who, perhaps, was not professional but was pretty skilful. They only allowed error on release speed and placed the arrows close together (Figure 11).

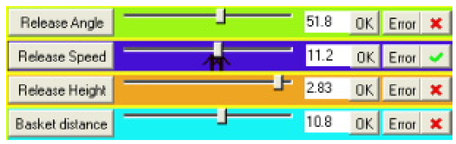


Figure 11. Tom and Chris began to model a skilful but not professional player. They used these values for their parameters.

After 93 throws, they looked at the graphs (Figure 12). The researchers were interested in how the two boys interpreted the histogram of release speed.

(50) Res: When he missed, why did he miss?
(51) Chris Because the speed wasn't enough.
(52) Tom: Most of the reds are at the lower side.

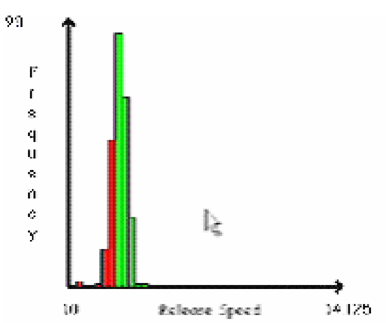


Figure 12. Tom and Chris produced this graph after 93 throws.

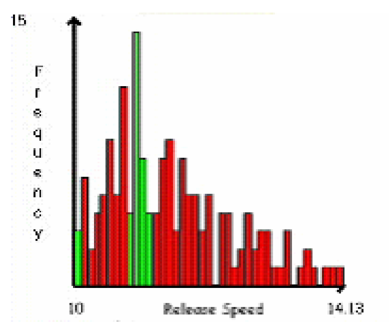


Figure 13. Tom and Chris were surprised to see two green (lighter) areas.

The researchers asked what would have happened if the player had been less skilful. Tom and Chris predicted that the red (darker) and green (lighter) areas would be swapped. It is difficult to understand what they meant by this but we think they meant that the red bars would be higher since there would be more misses (near to the height of the green bars in Figure 12) and the green bars would be lower (near to the height of the red bars in Figure 12). They did not refer to the spread of the graph.

They set the experiment up so that the arrows on the speed slider were wider apart than in Figure 11. After 225 throws, they looked at the graphs. Tom and Chris admitted some surprise at the speed histogram (Figure 13).

(53) Res: Is that what you expected?
(54) Tom: Not really… though there's lots of reds. It's kind of what I expected. I don't know… I don't know… it is kind of what I expected. There's a green there. I don't know why.

The green bar on the leftmost part of the histogram was a surprise to the boys though they were not surprised to see much more red than green since they knew that this player was less skilful. The researchers probed further.

(55) Chris: I think that green area bit will be where he hit off the side *(referring to the more central green area)* and that *(referring to the single green bar)* will be where he got it in straight away.
(56) Res: Ah, so that's why there are two separate areas of green, and why do you think the higher one is where he hit the backboard?
(57) Chris: Because there's more of it to hit… more area.

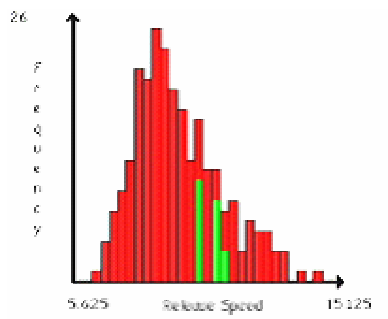


Figure 14. Did Chris's prediction that the results would be more to the left fit the speed histogram shown here?

Chris was able to explain the bimodal green distribution in terms of what he had witnessed in the simulation. The larger spread had in fact allowed the possibility of two distinctive ranges of value of speed that would generate successful throws. (Since then the researchers have found that it is possible to create situations in which there are four distinctive success regions since occasionally it is possible to score after the ball has bounced on the floor!)

5. DISCUSSION

5.1. SUMMARY

We should like to begin by summarising the four phases described above. In phase 1, Tom and Chris worked in an entirely deterministic fashion. They were comfortable with the idea that the path of the ball would be affected by the velocity and position of the throw. They were at ease in changing those parameters in order to determine a path in which the ball moved directly into the basket. This is evidenced by the brief time (about two minutes) needed to generate successful scores. Subsequently during this phase, Tom and Chris observed variation in the graphs though this was not randomly generated but the result of they themselves changing the values of the variables during the running of the simulation. We believe this appreciation enabled them to create a connexion between the variation in the histogram and causal actions in the simulation. During this phase, the agent of change was, of course, the boys themselves.

In phase 2, we used realism as an excuse to introduce the notion of error. As reported above, there was in fact some confusion over our use of this term as the boys initially expected that the computer would report an error, perhaps the ball would be somehow thrown incorrectly. This confusion though was transient. What was less transient was the sense-making process involved in gaining mastery over the arrows. Tom and Chris explored the arrows non-systematically, changing values of the parameters for variables, which also had error set as on. When they looked at the graphs, they believed that the variation was due to the changes that they had made. In this phase, they did not tend to attribute variation to randomness.

Phase 3 was marked by three key insights. First, they recognised that the throws were being chosen randomly from values between the two arrows. Though this is not exactly correct, since the arrows represent values roughly two standard deviations above and below the mean if the arrows are symmetrically placed around the handle, it is a reasonable understanding of the situation. Secondly and almost simultaneously they saw the arrows as agents of the variation in how the player threw the ball in the simulation. Indeed, they articulated this relationship rather concisely, “When the arrows are close together, it’s got more of a chance of going into the net.” We see this statement as a fine example of what Noss and Hoyles (1996) have called a situated abstraction, a heuristic that characterises the general behaviour of certain phenomena within a specific system. Thirdly, Tom and Chris were able to connect that relationship to the histograms. They were able to discuss how the variation in the histograms was itself caused by the arrows, thus co-ordinating the causal relationships between the simulation, the arrows and the graphs.

In phase 4, Tom and Chris began to use the co-ordinated understanding of the causal role of the arrows to model distributions. They saw this in a situated way. Their aim was to simulate a professional level player or one who was rather less skilful, and they set about that task by moving the arrows nearer or further away from each other. To their surprise, they encountered a bimodal distribution but were able to explain this in terms of the distinctive ways in which a player might throw the ball into the basket, namely directly or off the backboard.

5.2. CAUSALITY AND THE TWO PERSPECTIVES ON DISTRIBUTION

We began with the conjecture that we would be able to build an environment that enables the student to appreciate the limited explanatory power of causality to capture the essence of local variation. At the same time, we ventured that this environment would allow students to use causality to articulate features of distribution.

In fact, we have demonstrated, in Phases 1 and 2, the potential to use notions of skill-related error in a simulated sports context to perturb thinking away from a deterministic mindset towards one of randomly occurring events. Furthermore, in Phases 3 and 4, we have demonstrated how the sliders and arrows can become agents of change, in effect replacing the human agent. Moving the slider changes the position of the distribution. Moving the arrows changes the spread of the distribution. These ideas are articulated in situated ways such as “when the arrows are close together, it’s got more of a chance of going into the net.” We note the deterministic nature of this situated abstraction. While at the micro-level, causality is shown to have limited explanatory power, at the global level, causality can be harnessed to articulate the relationship between the parameters in the model (average, spread) and the shape of the distribution.

We regard this paradox of seeing the limitations of causality at one level while recognising its power at another level is at the heart of co-ordinating the two perspectives on distribution. We asked whether and how do students co-ordinate the data-centric and modelling perspectives on distribution. We are not able at this stage of our research to elaborate this aim to our complete satisfaction but we believe that insights into the role of causality are significant. We conjecture that as students pay attention to the sliders and arrows they are considering the modelling perspective on distribution, though of course they would not articulate it in that way. The aphorism “without distribution, there is no variation” regards distribution as the agent of variation and in effect comments on the ontology of distribution. We have tried to instantiate that perspective on distribution in the virtual world by offering students direct manipulation over the generational powers of distribution through instantiations of average and spread. In contrast, when students pay attention to the emerging data, they are considering the data-centric perspective on distribution.

In this sense, we find support in this study for the model proposed by Bakker and Gravemeijer (2004) in which distributions can be read from or towards the collection of data. However, whereas Baker and Gravemeijer refer to the student moving bi-directionally between the collection of data and a conceptual entity, we portray the journey as between a modelling and data-centric perspective. For us, the conceptual lies in the co-ordination of these two perspectives. Indeed, we wish to emphasise the equal status of those two perspectives.

Nor is this difference in emphasis merely playing with words. We believe it has teaching implications. There is much excitement about EDA as a modern method for exploring statistics. We share much of that excitement. However, the approach places emphasis upon a data-centric perspective and so far has not offered a coherent statement about how students might abstract from that perspective a rich concept of distribution, which co-ordinates both modelling and data-centric perspectives. Our studies are suggesting that the modelling perspective may need to be given equal status if such a co-ordination is to be encouraged.

Our data so far suggest that causality may be acting as the co-ordinating agent since, not only is it an idea that feels comfortable to students, but it also plays a critical role in helping them to make sense of the relationship between the parameters of the model and the shape of the data. If our ongoing design research continues to support this finding (which in the spirit of design research now becomes a conjecture to be tested in the next iteration), it throws new light on earlier research. The claims that centre around the role of causality in making sense of the stochastic are to some extent out of line with common thinking, which tends to make clear and distinct separations between the two. For example, Piaget and Inhelder (1975) portray randomness as inconceivable within operational thinking, at least until resolved by the invention of probability at a later stage of development. One of consequences of phenomenalisation (Pratt, 1998), turning mathematical ideas into quasi-concrete objects (Papert, 1996), is that mathematical concepts can be expressed in causal terms through the use of situated abstractions, as we have seen in this study. Even it seems statistical ideas, apparently separated from the deterministic world, are accessible to some extent through causal meanings.

We believe the role of causality in bridging the two perspectives on distribution may also have teaching implications. Fischbein (1975) has proposed that some of the difficulty that students have with probabilistic ideas is at least reinforced by a curriculum that emphasizes the deterministic. Indeed, commenting on this assertion by Fischbein, Langrall and Mooney (2005) state, “Children (as well as adults) need to recognize that situations involving chance can be examined and described logically and rationally” (p. 115). If it is true that causality plays the role we are suggesting, the key may lie not so much in reducing the emphasis on determinism as harnessing the power of causality towards the teaching of probability, perhaps through the use of technology.

5.3. LIMITATIONS

The reader must consider the limitations of this research to elaborate the conjectures and research questions. We have reported in some detail only on one pair of students, the clearest illustration of the emerging ideas. Even had it been possible to elaborate the activity of all three pairs, the findings must be regarded as tentative and in a sense interim. We may have, of course, further evidence after the following iteration but design research does not always follow such a smooth path.

It is quite feasible that the role of causality is directly linked to the virtual nature of the setting for this study. Perhaps it is only possible, or at least far simpler, to instantiate these ideas in a technological environment where it is possible to phenomenalise mathematical notions. It is reasonable then to suppose that access to the ideas is understood through the manipulation of the mathematical concepts as articulated through situated abstractions that link causally the inputs and outputs on screen. This is a limitation in so far as we can no more claim that our findings relate to the co-ordination of the two perspectives on distribution in other settings than can other researchers, who unavoidably work in particular settings, though sometimes ill-advisedly in our opinion, ignore the critical role of setting in abstracting. (See Pratt & Noss, 2002, for detailed elaboration of this issue.)

5.4. IMPLICATIONS FOR FURTHER RESEARCH

We have discussed data from a fairly early stage of our work-in-progress. Although we have moved through two previous iterations in order to reach this design, we recognise there are some further design changes to be made. Nevertheless, we believe our results so far indicate support for our conjecture that it is possible to design an environment in which students' well-established causal meanings can be exploited to co-ordinate data-centric and modelling aspects of distribution. Tom and Chris began to appreciate how not only might they themselves be agents of variation, but also how randomness, instantiated in the form of the quasi-concrete arrows, can create histograms in which variation is apparent. In this sense, randomness might become understood as reality once removed. What we have called "letting go of determinism" might be seen as delegating control to a quasi-concrete object that exercises that power through random effects.

It is in the nature of design research that the researchers gradually become sensitised to the ecology of the domain being investigated. We now feel that we have gained a handle on how to support the use of causal meanings in understanding distribution. In that respect we are close to having a design which can be used systematically to test out that conjecture.

- We shall remove the confusion introduced by the term, *error*. In the next iteration we shall simply refer to the arrows and explore what the students make of their role.
- We shall explore in more detail the notion of agency. We expect that agency will become an analytical category varying at least across human, slider and arrows.
- We intend to introduce a graphical representation of the modelling distribution accessed by clicking on the relevant variable such as release angle or speed. We conjecture that access to both the modelling distribution and the data-centric

distribution will enable us to explore more systematically some of the issues described above that still appear relatively mysterious.

- The introduction of a graphical representation of the modelling distribution allows us to introduce a new form of agency. We will hope to allow the students the facility to edit the modelling distribution as a means of transforming the modelling distribution directly but the data-centric distribution indirectly. We ask how will students articulate the chains of agency and how will that impact their co-ordination of the two perspectives on distribution.

Thus, the above outlines our own research programme for the near future. There are however important research questions which our programme will not address. In raising the idea that causality may be a significant agent in constructing a bridge between the data-centric and modelling perspectives, we acknowledge at the same time the possibility that technology is playing a key role in this process. There is fascinating research to be done in exploring the role of causality when other materially-based methods of supporting the co-ordination of the two perspectives on distribution are deployed.

There is much current interest (for example, Pfannkuch, 2005) in researching informal inference. (Informal inference is to be the focus of the fifth conference on Statistical Reasoning Thinking and Literacy to be held at the University of Warwick, August 2007.) EDA is developing interesting pedagogic approaches towards informal inference but we ask whether students can develop an appreciation of the robustness or power of their inferences without constructing a modelling perspective alongside their data-centric perspective. We see this question as one that should tax researchers of informal inference.

REFERENCES

Ainley, J., Pratt, D. and Hansen, A. (2006). Connecting engagement and focus in pedagogic task design, *British Educational Research Journal*. *32*(1), 21-36.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64-83.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Bakker.pdf]

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., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, *45*(1-3), 35-65.

Biehler, R. (1994). Probabilistic thinking, statistical reasoning, and the search for causes-Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.), *Research papers from ICOTS 4*, Marrakech, Morocco, July, 1994, (pp. 20-37). Minneapolis: University of Minnesota.

Borovcnik, M. (2005). Probabilistic and statistical thinking. *Proceedings of the Fourth Conference of the European Society for Research in Mathematics Education*. Sant Feliu de Guíxols: Spain.

Cobb, P. (1999). Individual and collective development: The case of statistical data analysis. *Mathematical Thinking and Learning*, *1*(1), 5–43.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research, *Educational Researcher*, *32*(1), 9–13.

Davis, B., & Simmt, E. (2003). Understanding learning systems: mathematics education and complexity science. *Journal for Research in Mathematics Education, 34*(2), 137-167.

DfES. (2000). Mathematics: The National Curriculum for England. London: Department for Education and Employment.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*: Dordrecht, The Netherlands: Reidel Publishing Company.

Gould, R. (2004). Variability: One statistician's view. *Statistics Education Research Journal*, *3*(2), 7-16.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Gould.pdf]

Hancock, C., Kaput, J. J. & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337-364.

Johnson, S. (2001). *Emergence: The connected lives of ants, brains, cities, and software.* New York: Scribner.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

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

Konold, C., & Pollatsek. A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 169-199). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Langrall, C. & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic reasoning. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 95-119). 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.

Mason, J. (1994). Researching from the inside in mathematics education - Locating an I - you relationship. In J.P. da Ponte & J.F. Matos (Eds.), *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education*, Vol. 1 (pp. 176-194), University of Lisbon, Portugal.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, *26*(1), 20-39.

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

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers*. London: Kluwer Academic Publishers.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics, 40*, 25-51.

Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers For Mathematical Learning*, *1*(1), 95-123.

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

Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 267-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children.* (L. Leake Jr., P. Burrell, & H. D. Fishbein, Trans.). New York: Norton. (Original work published in 1951).

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602-625.

Pratt, D, (1998). *The construction of meanings IN and FOR a stochastic domain of abstraction*. Unpublished doctoral dissertation, Institute of Education, University of London.

Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: The case of randomness. *Journal of the Learning Sciences, 11*(4), 453-488.

Prodromou, T. (2004). Distribution as emergent phenomenon. *Proceedings of the British Society for Research into Learning Mathematics*, *24*(1), 49-54.

Resnick, M. (1991). Overcoming the centralized mindset: Towards an understanding of emergent phenomena. In Harel & S. Papert (Eds.), *Constructionism* (pp. 205-214). Norwood, NJ: Ablex Publishing Corp.

Robson, C. (1993). *Real World Research*. Oxford: Blackwell.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223-266.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 193–204). Norwood, NJ: Ablex.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics*, *33*, 171-202.

THEODOSIA PRODROMOU
Centre for New Technologies Research in Education (CeNTRE)
Institute of Education, University of Warwick
Coventry CV4 7AL, United Kingdom

USING DATA COMPARISON TO SUPPORT A FOCUS ON DISTRIBUTION: EXAMINING PRESERVICE TEACHERS' UNDERSTANDINGS OF DISTRIBUTION WHEN ENGAGED IN STATISTICAL INQUIRY

AISLING LEAVY
Mary Immaculate College, University of Limerick
Aisling.Leavy@mic.ul.ie

ABSTRACT

This exploratory study, a one group pretest-posttest design, investigated the development of elementary preservice teachers' understandings of distribution as expressed in the measures and representations used to compare data distributions. During a semester-long mathematics methods course, participants worked in small groups on two statistical inquiry projects requiring the collection, representation, analysis and reporting of data with the ultimate goal of comparing distributions of data. Many participants shifted from reporting descriptive exclusively to the combined use of graphical representations and descriptive statistics which supported a focus on distributional shape and coordinated variability and center. Others gained skills and understandings related to statistical measures and representations yet failed to utilize these when comparing distributions. Gaps and misconceptions in statistical understanding are discussed. Recommendations for supporting the development of conceptual understanding relating to distribution are outlined.

Keywords: *Statistics education research; preservice teacher education; distribution; statistical inquiry, data comparison, teacher knowledge*

1. INTRODUCTION

The teaching of statistics in elementary schools has received increased attention and priority over the past three decades. The release of the *Curriculum and Evaluation Standards for School Mathematics* by the National Council of Teachers of Mathematics (1989), which incorporated a strand focusing on data analysis and probability, and the publication of the *Guidelines for the Teaching of Statistics K-12* (1991) by the American Statistical Association, are two important landmarks. The increased focus on elementary level data analysis and statistics is evident in the proliferation of curricula designed specifically for younger students, such as the *Used Numbers* Project (Technical Education Research Centers and Lesley College, 1989), *Mathematics in Context* (National Center for Research in Mathematical Sciences at the University of Wisconsin/Madison and Freudenthal Institute at the University of Utrecht, 1997-1998), the *Investigations in Number, Data, and Space* (TERC, 1998), and the *Connected Mathematics* Project (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002). We also see computer software, minitools, and tutorial tools developed for elementary and middle grade students, such as *Tabletop* and *Tabletop Jr.* (Hancock, 1995; Hancock, Kaput & Goldsmith, 1992), *Statistical Minitools* (Cobb, Gravemeijer, Bowers, & McClain, 1997),

Statistics Education Research Journal, 5(2), 89-114, http://www.stat.auckland.ac.nz/serj

Authentic Statistics Stack (Lajoie, 1997), and *Tinkerplots* (Konold, 1998; Konold & Miller, 2001).

The focus of recent curricula and software has been on the notion of distribution and how to support students in understanding distribution. One area which is lacking, however, is an analysis of the ways in which teachers understand and are prepared to teach fundamental notions associated with distribution.

2. SCIENTIFIC BACKGROUND

2.1. DESCRIBING DISTRIBUTION

Distribution refers to the arrangement of values of a variable along a scale of measurement resulting in a representation of the observed or theoretical frequency of an event. Descriptive statistics are indices of distribution: they summarize complex data into measures that can be compared against each other to ascertain the nature of a dataset, and the degree to which two or more datasets are similar. Central elements in the development of a concept of distribution are notions of central tendency, variability, symmetry (skew) and relative frequency (kurtosis), each of which can be modeled using descriptive statistics.

One way to get a handle on distribution is through identifying landmarks and trends (Friel, Mokros, & Russell, 1992) in data. Taken together, identifying data landmarks (outliers, gaps) and generating measures that index certain characteristics of the data (for example, measures of center and variability) provide insights into properties of any given distribution. While descriptive statistics are central components of any treatment of distribution, a focus on them alone can, as cautioned by Makar and Confrey (2005), "aggravate the focus on individual points" (p. 28). Graphical representations serve as useful tools to communicate aspects of a distribution as they facilitate a focus on aspects of the data that may be missed with the use of descriptive statistics alone. Graphics have been described as *revealing* data (Tufte, 1983) and as being superlative to statistical computations in revealing information about data. However, little is known about the ways in which learners use graphical representations to communicate aspects of a distribution.

Research reveals an emphasis on measures of central tendency as a means to describe data distributions resulting in an overemphasis on centers and the corresponding neglect of the variation found within a given distribution (Shaughnessy, 1992, 1997). Variation is a critical component of, and inextricably linked to, the concept of distribution and has been found to play a central role in children's thinking (Cobb, 1999; Konold & Pollatsek, 2002; Watson & Kelly, 2002). An understanding of distribution requires an awareness of the propensity of a variable to vary and comprehension of how that variability contributes to the notion of the distribution as an aggregate rather than a collection of individual data points.

2.2. USING DATA COMPARISON TO SUPPORT THE DEVELOPMENT OF UNDERSTANDINGS RELATED TO DISTRIBUTION

A critical statistical notion for learners is that of dataset as an entity, in other words developing a 'statistical perspective' (Konold, Pollatsek, Well, & Gagnon, 1997). Holding a statistical perspective requires a focus on the dataset as a collective rather than focusing on individual data values. By focusing on comparing distributions students are provided with a conceptual structure that facilitates a focus on aggregate (Cobb, 1999).

More specifically, when comparing datasets the activity leads to consideration of the shape, center and variability of a distribution of data, in turn, providing a context for examination of the distribution as an entity.

Another reason for engaging students in the activity of comparing datasets is due to the focus on variability that is nurtured. Unexplained variation in data creates noise and the primary purpose of many statistical techniques is to unearth the signal within the noise (Konold & Pollatsek, 2002; Wild & Pfannkuch, 1999). Comparing datasets requires a learner to examine the variation both within and between distributions of data. This requires identifying signals or patterns within a dataset worthwhile of attention, and comparing these signals against those emitted by the comparison dataset. Identification and communication of these signals reveals aspects of an individuals' understanding of the notion of distribution.

2.3. PRESERVICE TEACHERS' UNDERSTANDING OF DISTRIBUTION

Given the relatively recent election of data analysis as a focus of mathematics instruction at the K-12 level, it is conceivable that many teachers may be teaching statistical content that they themselves have little experience with as learners. Lajoie and Romberg (1998) comment that statistical concepts may be as new a topic for teachers as for the students they teach and recommend that "teachers must be provided with appropriate preservice and in-service training that will give them the knowledge base they need to feel comfortable teaching about data and chance" (p. xv).

Much of the current research focuses on preservice teachers' understandings of measures used to index distributions of data (Canada, 2004; Gfeller, Niess, & Lederman, 1999; Heaton & Michelson, 2002; Makar & Confrey, 2002). Many of these studies converge on the same finding – preservice teachers' understanding of measures of center tends to be procedural rather than conceptually-based. Leavy & O'Loughlin (2006) report on elementary preservice teacher's fluency in using the mean algorithm but identified gaps in conceptual understanding. Indicators of poor conceptual understanding were lack of understanding of the mean as a ratio, difficulty solving weighted means problems, and poor analog knowledge of the mean (a concept akin to Skemp's (1979) concept of relational knowledge). The prevalence of procedural understandings is further supported by Gfeller, Niess, & Lederman's (1999) finding that computational algorithms were the most prevalent method used by preservice teachers for solving problems related to the mean.

Studies examining variability indicate that the provision of coherent and meaningful statistical activities can lead to gains in understandings of variability. Canada (2004) found that following activities involving chance and computer-generated simulations, preservice teachers' predictions of variability moved from expectations of way too little or too much variation to more realistic expectations of variation. It has also been found that preservice teachers are more likely to use measures of variability to represent a distribution if the data are presented graphically (Makar & Confrey, 2005). This suggests that when choosing methods to represent a distribution of data, merely presenting the data graphically may draw attention to the variability of the data and make variability " ... perhaps more compelling than any measure of center" (pp. 36-37). Other advantages of using graphical representations are identified by Hammerman and Rubin (2004) who reveal that having access to a visual representation of a distribution may influence the value(s) that one chooses to represent that distribution. The authors comment on the low occurrence of comparisons based on means or other measures of central tendency and assert that "seeing a distribution makes it harder to accept a measure of center, especially a mean, as being representative of the entire distribution" (pp. 36-37).

A recent analysis of the literature revealed that only 2% of published research in mathematics education was devoted to probability and statistics (Lubienski & Bowen, 2000). Within this group, there is a visible absence of studies focusing on understandings of *elementary* preservice teachers, resulting in an inadequate picture of the preparedness of our elementary teachers to teach this new and emerging field of study. Unlike their secondary counterparts, elementary teachers are not expected to possess as strong or as broad a foundation in mathematics. However, given their role as primary mathematics educators of young children, a study of preservice elementary teacher's mathematical understanding is warranted. Subject matter knowledge in the preparation of teachers has been identified as a fundamental component of teacher education programs (Ball & McDiarmid, 1990). We now understand that poor mathematics content knowledge may lead to an overemphasis on limited truths and procedural rules (Stein, Baxter, & Leinhardt, 1990), inaccurate explanations (Borko, Eisenhart, Grown, Underhill, Jones, & Agard, 1992), and a lack of understanding of the appropriate representations to utilize when supporting the development of rich mathematical understandings in children (Borko et al., 1992). These relationships between content knowledge and instructional practices make it critical that mathematics teacher educators develop a greater understanding of elementary preservice teachers' statistical understanding.

2.4. PURPOSE OF THIS STUDY

The purpose of this exploratory study was to investigate the development of preservice teachers' understandings of distribution, expressed in the measures and representations used to compare distributions of data. Specific goals of interest were to:

(i) investigate the approaches used to compare distributions of data.
(ii) identify the statistical concepts focused on when reasoning about distributions of data, and examine the ways in which different understandings of these particular statistical concepts support or hinder the description, analysis, and comparison of datasets.
(iii) explore ways to support the development of rich understandings of distribution.

3. METHOD

3.1. PARTICIPANTS

Twenty-three participants were enrolled in a mathematics methods course in a university in the USA, as part of a one-year master's degree program leading to elementary teaching certification. Participants ranged in age from 22 to 55, seven were male. All participants held a bachelor's degree, with majors in Art, Business, Chemistry, Computer Science, Criminal Justice, Early Childhood Education, Economics, English, History, Psychology, Public Relations, Sociology, and Spanish. Three reported taking Advanced Placement Statistics in high school; almost half had no formal coursework in statistics.

3.2. APPROACH

This study employed a one-group pretest-posttest design and involved collection of baseline data or pre-test, an instructional intervention, and a post-test. The study represents a blend of components from two analogous research methodologies: Teaching experiment methodology (TEM, Steffe & Thompson, 2000) and the teacher development experiment (TDE, Simon, 2000). Using a blend of both methodologies supported the documentation of participants' understandings of distribution on entry to the study, the observation of changes in understanding over time, and the focus on the process of student learning and the concomitant teacher actions that promoted advances in statistical understanding. The study departed, though, from a true implementation of either methodology. The dearth of research relating to distribution resulted in the absence of an empirical research base from which to inform the construction of hypothetical learning trajectories, critical components of Teacher Experiment Methodology (TEM).

The whole class teaching experiment was conducted over a 15-week semester in collaboration with two teachers who were members of the four-person research team. Two or more research team members were present in the classroom during teaching sessions and were involved in the daily organization and evaluation of classroom mathematical practices. It was through juxtaposing these multiple perspectives that we gained rich and accurate insights into the development of statistical understanding. Weekly meetings of the research team focused around "taken-as-shared" (Cobb, 1999) interpretations of classroom activity, taken as shared meanings constructed as a result of cycles of construct development that supported the refinement of our own statistical knowledge. Meetings were used primarily as contexts within which to share interpretations of events, devise contexts in which to test these interpretations, and eventually refine and extend interpretations. This researcher-level data analysis supported the development of refined understandings of the process of statistical learning in addition to providing a focus on the development of teachers and their pedagogical understandings, an important component of the Teacher Development Experiment (TDE).

3.3. DESIGN AND TASKS

This study was organized in three parts (see Table 1): Pre-test (P), intervention with instructional components (I), and post-test (PT). The baseline data collection (P) was intended to probe participants' understandings of distribution. Central to this phase were the data collection and representation phases of *the bean experiment* investigation.

During the intervention phase (I), instructional activities supported the development of statistical reasoning, and made use of two tasks described later in more detail, the Beans investigation and Popcorn experiment. Firstly, students worked together to compare distributions of beans grown in different conditions. Secondly, several weeks of instruction focused on the stages of statistical inquiry and what was involved in moving through a statistical investigation (also known as data modeling, see Lehrer & Romberg, 1996). The foci of instruction were derived from analysis of the pre-test activities and related directly to distribution (representativeness, dataset as an entity). In an effort to maintain coherence between the areas of instruction, instructional themes were anchored within the umbrella concept of 'carrying out a statistical investigation'. Instruction was related primarily to the ongoing statistical investigations with the beans; the ability to ask participants to examine and assess their own work facilitated us in highlighting events that occurred in participants' own investigations. In preparation for the instruction, we had digital images of participants work (for example, the graphical representations constructed to compare datasets) or transcriptions of conversations or comments, which we then placed in our power point presentations. These records often became the focus of instruction and presented an opportunity for participants to reflect on and assess their work in light of research being examined as part of course experiences. The primary topics of instruction are presented on table 1. A third component of the instructional phase engaged participants in making journal reflections on statistical understandings related to distribution. The fourth component was involvement in focus group discussions relating to the strategies and approaches used to analyze bean data.

Finally, in the post-instruction phase (PT), we examined changes in participants' understandings about distribution. To that end, during the penultimate class session participants engaged in the popcorn investigation that involved comparison of the effect of refrigeration on the popping of kernels.

Engaging in statistical inquiry was a critical component of the study. From a pedagogical perspective, engaging preservice teachers in statistical investigation provides the potential to highlight statistical (and other) issues they may face when approaching statistical inquiry in their own classrooms. Secondly, engaging in 'real life' statistical investigation engages participants in the activities of statisticians and exposes them to real world 'messy' data. It was our goal that we set up statistical contexts that supported prospective teachers in learning 'statistical concepts in an environment much like the one recommended for students – one that is active … involving authentic data, and offering plenty of opportunities to build their conceptions through experiences with data' (Makar & Confrey, 2005, p. 30). Research focused on activities surrounding two investigations:

The bean investigation A semester-long statistics research project, *investigating optimal conditions for growing beans sprouts* (Appendix A), was one context within which the research was carried out. Participants were presented with an experimental design to determine which growing conditions supported the best growth.

Participants were divided into groups and provided with a bag of 25 lentil beans, a solution (lemon or water), a paper towel, a plastic sealable bag, and a card identifying the light intensity (light/dark) in which the beans would be placed. There were eight groups and four conditions: water/light, water/dark, lemon/light, lemon/dark. Participants were instructed to spray the solution on a paper towel, place the bean on the towel, fold the towel, and place in the plastic bag. The bags were sealed and placed in the labeled light intensity for seven days. The following week the beans were brought to class and the sprouts measured and recorded by the groups. Each group was then responsible, over the course of the semester, for constructing a hypothesis regarding the optimal condition for bean growth and determining what statistical measures or approaches they might utilize to test the hypothesis. Each group was instructed to analyze and compare the data collected by all eight groups and prepare a presentation of the findings. It was by engaging participants in this data comparison scenario that we established the need for individual distributions of data to be represented and indexed in a way that required a focus on distribution and in turn facilitated the comparison of the distributions.

The popcorn experiment The purpose of the post-intervention statistical investigation, *the popcorn investigation* (Appendix B), was to provide a context similar to the bean experiment where participants were once again engaged in a data comparison activity. The requirements of the task were similar to those of the pre-intervention task thus allowing identification and examination of changes in statistical understanding.

The investigation involved two samples of popcorn kernels (n=100 in each sample), one kept at room temperature and the other refrigerated, being popped in an open-top popcorn popper for 4 minutes. The position that each popped corn kernel landed was marked on the plastic 'target sheet' which had been placed beneath the popper. Groups were then presented with the data, asked to examine the distribution, and determine whether refrigerating the kernels influenced the distance that they reached when popped.

Table 1. Outline of statistical experiences over the course of the semester

Phase	Wk	Activity	Specific (teaching) experiences
P	4	Bean investigation: Setting up the experiment–planting	See section 3.3
P	5	Bean investigation: Data collection–measuring beans	See section 3.3
I	6	Instruction relating to asking research questions, collecting data	Instruction focused on overview of data modeling (Lehrer & Romberg, 1996), supporting children in formulating research questions, examples/analysis of questions children construct, types of data generated from questions, overview of categorical and numerical data types, overview of data collection methods (surveys, experiments, observations), supporting children in choosing appropriate data collection methods, identifying and overcoming obstacles and common difficulties faced by children when collecting data.
P	7	Bean investigation: Data representation phase	See section 3.3 and Appendix A
I	7	Bean investigation: Data analysis and comparison	The session was considered instructional in that students were presented the opportunity to learn when working in groups. Activity focused on the analysis and comparison of datasets.
I	9, 10	Reflection on statistical understandings arising from the bean experiment	Revisiting the stages of data modeling in the context of the bean experiment: What have we learned? What difficulties did we face (procedural, content understandings etc.) and how might this apply to classroom teaching of data analysis and statistics?
I	11	Focus group discussion relating to the strategies and approaches used during the bean investigation	The following are examples of questions asked during focus group discussion which focus on statistical misconceptions identified in weeks 1-10: What do descriptive statistics *not* tell us about a distribution? What does standard deviation mean in the context of a distribution of data? What is the relationship between the mean and sample size of the distribution? What role did zero values play when comparing distributions of data? Describe why some groups found the mean of group means? Why did others not? When are graphical representations useful? When/why might a box and whisker plot be used to represent a distribution? What does relative size of quartiles in a box and whisker plot communicate about the data? When/why might you use (or not use) a scatter plot?
I	11	Instruction relating to representing data (using graphical representations), analyzing and comparing data	Instruction on representing data focused on advantages of using graphs, guidelines from research about elements to attend to when constructing graphs, relationship between graphs and the data they represent (categorical/numerical), features of specific graphical representations and inherent advantages and disadvantages of their use (tables,

			pictograms, pie charts, bar graphs, line plots, stem and leaf plots, box and whisker plots), examples of graphs constructed by children, common errors children make when constructing graphs, categorizing representations of data. Instruction on analyzing and comparing data focused on defining distribution, important features of distributions, examining distributional shape (landmarks, bumps, gaps, outliers), locating measures of center on a distribution of data, examining models of the mean (leveling out, balance, fair share), locating indicators of variability on a distribution, thinking about skew when examining distributions, engaging in data comparison (Ben-Zvi, 2003; The “basketball problem,” McGatha, Cobb & McClain, 2002), examining children’s thinking about distribution, common errors children make when using, generating, and describing measures of center and variability.
I	13	Instruction on graphical representations; measures of central tendency and variability; means and weighted means	Instruction on graphical representations focused on revisiting features of graphical representations (Friel, Curcio & Bright, 2001), relationship between box and whisker plots and variability, what is a quartile?, distinguishing between univariate and bivariate data, scatter plots and the data they represent, why we use particular graphs, thinking about our bean data and graphs we used to represent the data (error analysis). Instruction on measures of central tendency and variability distinguishing between measures of central tendency, what research tells us about children’s understanding of the mean and median, features of the mean (e.g., Strauss & Bichler, 1988), revisiting models of the mean, identifying when measures of central tendency are most appropriate, examining use of the mean in the bean data experiment, identifying and indexing variability, representing variability using graphs, looking at use of variability in the bean experiments. Instruction on the weighted mean focused on what is a weighted mean, examining the elevator problem (Pollatsek, Lima, & Well, 1981), analysis of responses on weighted means problems (Leavy & O’Loughlin, 2006), examining the GPA problem (Pollatsek, Lima, & Well, 1981), analysis of responses on the GPA problems (Leavy & O’Loughlin, 2006), thinking about the bean data and weighted means.
PT	14	Popcorn investigation: Representing, analyzing, and comparing data	See section 3.3 and Appendix B
PT	15	Self report on growth in understanding	

3.4. DATA SOURCES

Several types of data were collected: videotape data, observational data, audio taped data of small group interactions, and student artifacts. Weekly sessions were videotaped. During statistical investigations two cameras were used: One focused on the whiteboard at the front of the room (where groups presented the outcomes of their investigation and where the instructor was located); the other moved between groups and focused on particular students during large group discussions. Teacher and researchers observations were recoded during the weekly sessions. There were always a minimum of two researchers in the room during each session, and for four weeks three researchers were present. The third data source was audio taped records of small group conversations during the investigations. Student artifacts were the fourth data sources and took several forms: student journals focusing on the development of statistical understanding, small group reports completed during the bean and popcorn investigations, and large presentation posters used at the culmination of the statistical investigations.

3.5. ANALYSIS OF DATA

In line with TEM and TDE there were two cycles of analysis: ongoing and retrospective analyses. Following each teaching episode, researchers examined the data and met as a group to discuss interpretations of classroom events. Each researcher had responsibility for examining in-depth a subgroup of students over the course of the semester. Discussion focused primarily on the types of understandings about distributions that were evident from videotape and audiotape recordings of participant activity during the class session. Each researcher shared the findings of her analysis and observations and when necessary situated the activity in relation to previous class sessions. The group then identified themes that emerged from the individuals' analysis of the data and constructed assertions relating to the themes. Individual researchers then revisited the data relating to their subgroup of participants with a view to finding supporting or contradictory evidence for the assertions. If the assertion was found to hold for a number of subgroups the cumulative class data were revisited in an effort to seek out evidence to triangulate the data. In cases where there was no supporting evidence we constructed a task or activity to present in the next teaching session (or in a focus group) that would test the assertion. On occasions where misconceptions relating to distributions were identified, statistical tasks were constructed to address misconceptions. It was through these cycles of hypothesis construction, data mining, triangulation, and hypothesis testing that understanding of the participants' understandings of distribution were developed. Figure 1 illustrates the interaction between data collection, analysis teaching episodes.

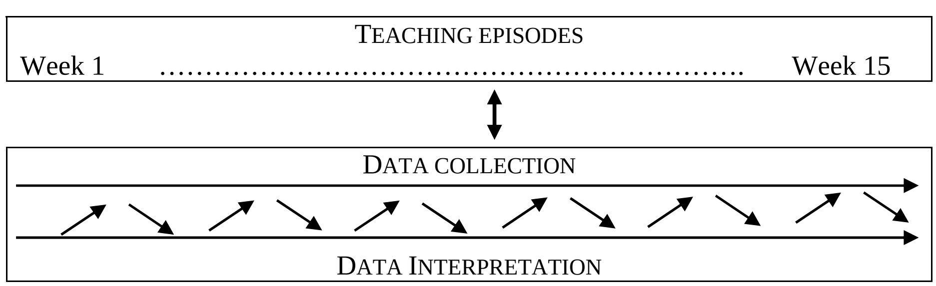


Figure 1. The interaction between data collection and data analysis in the context of a teaching experiment. Adapted from Lesh and Lehrer (2000).

Retrospective analysis at the end of the study involved re-examining participant activity over the entire semester. This allowed for a reanalysis of the findings and identification of supporting and contradictory evidence for the major claims arising form the ongoing weekly analysis of data.

4. RESULTS

4.1. INITIAL UNDERSTANDINGS RELATING TO DISTRIBUTION

Engaging in the bean experiment (Appendix A) provided insights into the types of distributional understandings held by preservice teachers. Participants compared lengths of germinated beans and presented their conclusions for feedback from their peers.

A lack of attention to distributional features of the data was apparent in the dominance of numerical methods for making data comparisons. Three of the groups (A-C) used descriptive statistics, alone, on which to base judgments about data. While use of the mean is appropriate as a comparative measure, absolute reliance on descriptive statistics is limiting as they provide merely one perspective on the data, that of centers, and do not take into account other features of the data (e.g., shape, variability). These groups did not invoke the use of graphical representations as a way to explore the datasets nor did they provide alternative perspectives not immediately apparent through the use of descriptive statistics. Group A's justification of their data comparative method follows:

> Our hypothesis was that the beans in lemon water would grow longer than the ones in water. We compared means because it sort of cancelled out the real high ones and the real low ones [data values representing bean heights] but incorporated every single piece of data. The main limitation is because there are high and low values they skew the data. … our answer was the sprouts in water were 20.2mm and in lemon water were 1.7mm so we were wrong in our hypothesis.

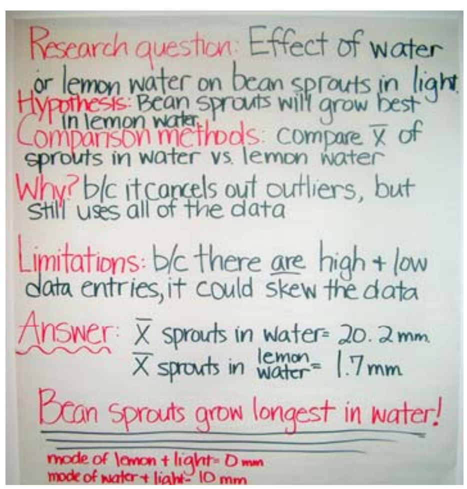


Figure 2. Illustration of Group A's approach to comparing datasets

The remaining five groups used a combination of graphical representations and descriptive statistics as data comparative tools. Interestingly, these students who used graphical representations in the pre-assessment did not have more statistical experiences with data prior to taking the course than their peers who did not construct graphs.

Examination of the graphical representations constructed by these groups, however, reveals that the graphs were used merely to illustrate summary statistics rather than illustrate distributional features. Four of these five groups used bar charts to represent group means resulting in a representation of subgroup means, rather than presenting a picture of the distribution of values along a scale of measurement (see Figure 3). The following is the response of Group D:

> We were interested in how lemon or no lemon affected the growth of beans in light or in the dark. Our hypothesis was that dark and lemon would yield the longest beans so … we were close in the sense that lemon was up there as number 2 and water and dark were number 1. So why did we do a bar graph? Cause it was easiest to show everything. So we got the averages for the four conditions. We didn't count zero cause they didn't germinate. The limitation is that outliers skew the average.

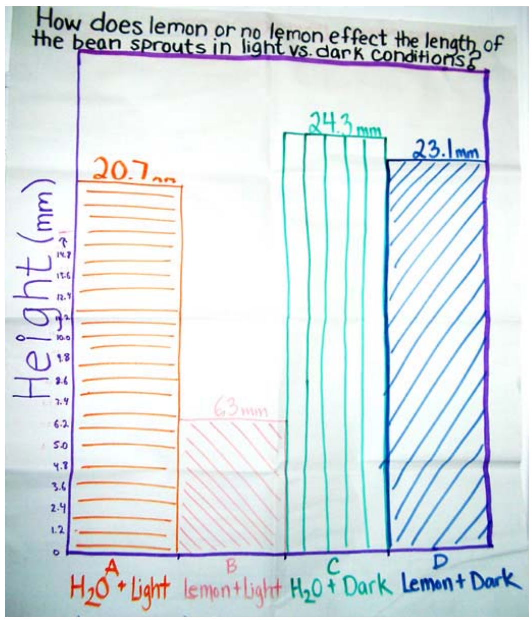


Figure 3. Illustration of group D's use of bar charts to illustrate sub group means

There was evidence of an attempt to represent the variability of the data. This is illustrated in the graphical representation of group G (see Figure 4). The representation resembles a double bar graph consisting of eight groups arranged in pairs; each pair corresponds to the two sets of data collected for each of the four conditions. The height of each bar represents the mean height of beans within each group (rather than the frequency of elements within the group). The line superimposed on each bar represents the range of heights of beans within each sample, thus the 'whiskers' represented the lower and upper limits of bean height. Thus the graphical representation was used as an instrument to report descriptive measures with the bars reporting group means and the whisker lines reporting the range of the data; indicating an attempt to coordinate both center and variability. Examination of group G's justification for inclusion of the range line reveals that they decided to report the variability given the large discrepancies in the sample statistics for beans grown under the same conditions. It seems that the unexplained variation in the measurement of bean heights was creating so much *noise* (Konold &

Pollatsek, 2002; Wild & Pfannkuch, 1999) that it was causing the group to refute the presence of a common signal across samples representing identical conditions. As Stan reported:

> We decided to go with a double bar chart cause of all the inconsistencies in measurement. So we did 8 groups and not 4 we also decided to do a box and whisker plot to show the full range … it is the orange line on the bar graph.

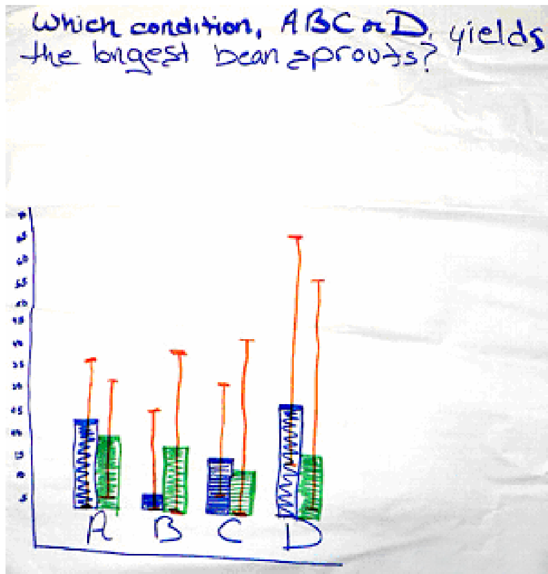


Figure 4. Illustration of Group G's attempt to coordinate variability and center

Analysis of pretest data did not reveal a relationship between participant's experiences with statistics prior to the study and the nature of the analyses carried out in the pretest.

4.2. END-OF-SEMESTER UNDERSTANDINGS RELATING TO DISTRIBUTION

We shift now to present some findings related to how groups coped with the Popcorn experiment, which as explained earlier was used to examine post-intervention performance of the groups. As compared to findings at the beginning of the semester outlined in the previous section, six of the groups shifted to using graphical representations to provide a picture of the distribution of data values. Groups reported that the selection of graphical representations to compare popcorn data was based on the capacity of the representations to highlight distributional features of the data, in contrast to the graphs used at the beginning of the semester which functioned merely to represent group means. This attention to global patterns in distributions of data was evident in the use of representations that highlighted distributional features in the data in particular the use of stem and leaf plots by five of the groups, and to a lesser extent box and whisker plots which were utilized by one group. The increased use of representations did not lead to the neglect of descriptive statistics. Figure 5 shows the data comparison strategy used by one group. Reporting the sample means and the range on the stem and leaf plots maintained the focus on measures of center and variability evident at the beginning of the study. As compared with their previous strategy (see Figure 3) this response represented a coordinated and detailed approach to analyzing data.

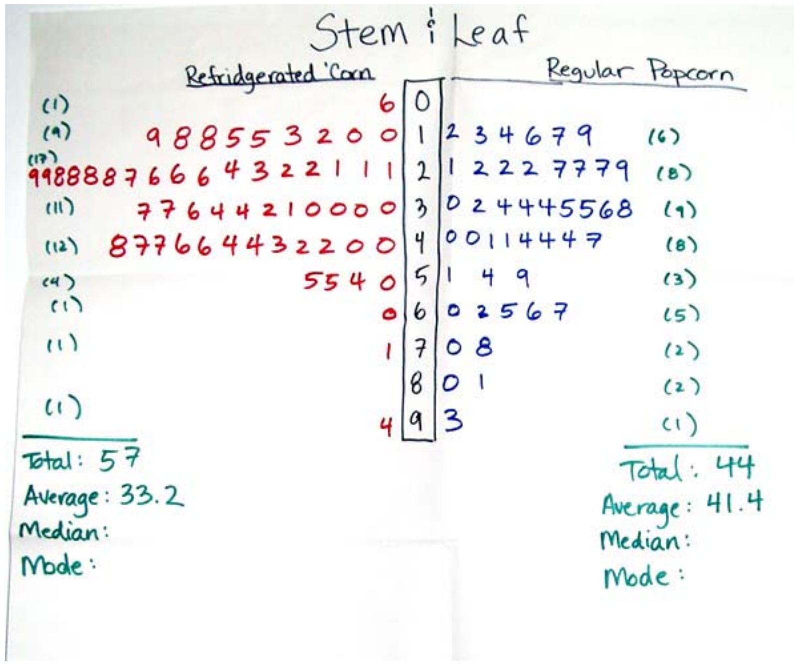


Figure 5. A comparison of distances by graphically illustrating distributional features in conjunction with the calculation of measures of central tendency

Tom: We chose this [stem and leaf plot] cause it gives you a way to see all the points at the same time and it gives you a sense of the distribution while getting all the data points there. And em … we thought that it was pretty representative. We predicted our mean would be higher for the room temp popcorn because of the way the tail skewed and we were right about that (pointing to the means)… and although the refrigerated popcorn did get a better yield it seems like the room temp popcorn did fly a little farther.

Barry: Melissa and I talked a lot about the kind of analysis we could use to think about the distribution curve. ... Well the curve for refrigerated has a bulk (pointing to the values at 10-40mm) where the curve for room temperature doesn't seem to have a bump it seems to be a much smoother curve. And we noticed that the number of kernels that popped even though it was different it wasn't all that big of a deal because the sizes were large enough that we could see what the curve would do if there were a ton more data points. You could kinda visually ... kind of .. visualize what the data would look like … how things would fall into that distribution curve.

R: During the discussion, you were trying to decide to do a box and whisker plot and you decided not to. Can you explain why?

Barry: Cause even though I like box and whisker it is difficult for me to verbalize exactly what that [box and whisker plot] represents. And I like how I can look at it and say “okay that’s cool” but to try and make descriptions about what that represents is difficult.

Mia: With stem and leaf plots you can see gaps and bumps in the data. While it looks nice on the box and whisker to see where medians lie against the boxes, is difficult to see what numbers are in between the quartiles and see how the distribution looks.

Barry: Even though in box and whisker you can see that the quartile is large you can’t see why. When you look at the stem and leaf you can see it is because 60 has the bulge. With the double stem and leaf you can see there is a bump coming out that skews the data one way or another ..

While increased attention to the distributional features of data is a welcome finding, examination of the specific groups who used representations reveals a disconcerting pattern. Five of these groups had used graphs in the bean experiment, while the sixth group had used means. Two of the three groups who relied exclusively on descriptive statistics at the end of the study to make comparisons had not used graphs to represent the bean data in the initial weeks of the study. This finding indicates the development of a broader understanding of the functionality of graphical representations, but only for those who were already inclined to use such representations. Those who set out using descriptive statistics exclusively demonstrated stability in strategy use. Thus we were successful in helping participants understand the differential limitations and advantages of particular graphical representations, as indicated in their reflections at the end of semester; however the stability in strategy use for those who used descriptive statistics indicates that for these participants we were less successful in communicating the functionality of graphical representations as exploratory data analysis tools.

Figure 6 shows the data comparison strategy used by Group A who had used similar descriptive statistics to support their argument in the bean investigation (see Figure 2).

The following transcript is Group A’s reasoning for *not* constructing a graph:

Valerie: Cause the question was more about distance and you can’t compare individual kernels, we weren’t concerned with how they clustered. That’s why we didn’t do a graph. It was more in terms of the average distance that we looked at it. We didn’t do the graph .. cause frankly the double stem and leaf is very messy for me. I understand you can see the bell curve but em it is very … it is too much. I’d prefer more concise data and more gearing towards the average. If it had asked perhaps how would you show the data or …

Robyn: How would you visually represent it?

Anne: Right.

Valerie: That’s why we chose average cause the question wasn’t so specific.

Closer examination of the dialogue indicates that participants in Group A may have interpreted the Popcorn task as estimating the average distance from the refrigerated to the non-refrigerated kernels. This involves a comparison of averages and does not require the construction of a graphical representation of the data. In this case, participants may have decided against constructing a graph as it did not represent a useful or efficient strategy in the context of this particular task.

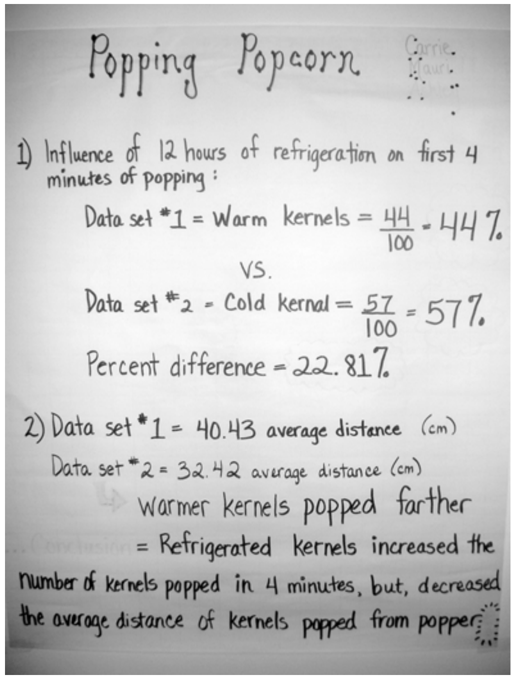


Figure 6. The use of descriptive statistics to describe a distribution of data

Examination of the reports accompanying the popcorn experiment reveals one common theme underlying the choice of descriptive measures: the ability of descriptive statistics to include all data values in the comparison. As Group B stated "averages were used because it is too difficult or rather impossible to compare individual kernels between the two [conditions]." This comment indicates the realization that unequal sample size would make a one-to-one comparison of values impossible, and in any event any one-to-one comparison of values would result in the loss of data (of the larger dataset). The upshot of this is twofold. On the positive side, participants understood the mean as a proportional measure appropriate to use in comparison of unequal size datasets. On the other hand it seems that the group was not prepared, or able, to examine global aspects of the distribution. Their comments indicate that a focus on individual data values would be the alternative had the datasets been equal in size. Thus it may be that this group was still focused on individual values in the datasets rather than possessing what Konold et al. (1997) term a statistical perspective.

When analyzing these groups' responses across the semester we can see that participants have demonstrated growth in several areas. Firstly, they are now attuned to sample size, something that they did not consider in the bean experiment. Secondly, their choice of descriptive statistics is now grounded in an attempt to include all data values particularly given the realization that datasets are different in size. It seems that this newfound attention to sample size disparity provided greater support for use of the mean and median. Finally, two groups demonstrated some awareness of the limitations of their approaches. Following the group presentations, the act of revisiting their own analyses in light of the presentations (see Step 5d Appendix A, and Step 4b Appendix B) led them to reflect critically on the limitations of not providing a picture of the distribution. The following quotes highlight the discomfort that two groups felt with the data reduction that takes places with the use of descriptive measures.

Group A: After looking at the other groups' presentations we now see that with this method [using means to make group comparisons] we cannot determine if there are clusters of distances

Group C: We liked our method. But the limitation of it [using the mean, median and mode] is that it reduces all data into one number. If we had the opportunity we would re-do our analysis again because seeing other people's graphical representations, it is much easier to see all data represented and to see how much farther it popped in either scenario.

4.3. OTHER ISSUES IMPACTING UNDERSTANDING OF DISTRIBUTION

The mean as a proportional measure Despite focused instruction on the concept of the mean, Amanda posed a question on the last day of class that revealed her difficulty understanding the mean as a data comparative measure.

Jay: Okay then it looks like we'll just find the means of the refrigerated ones and the regular ones. Then we can see if there is any difference between the groups.

Amanda: But if we have different N's doesn't each one [popped corn] in the smaller set get more significance?

This comment indicates that Amanda is grappling with the issue of how to best compare two datasets of unequal sample size. What is evident at the same time is her difficulty understanding the mean as a proportional measure. At this juncture, the instructors recapped on a conversation held several weeks previously in which particular features of the mean were explored and reminded Amanda of the ways in which the mean deals with unequal sample size. Amanda's final comment indicates that she has accepted that the mean is a measure that can be used to compare datasets of unequal sample size however she is of the understanding that there may be some cut-off point representing magnitude in sample size difference when the mean is no longer an appropriate comparison measure.

Amanda*:* I wonder how big does the difference [in sample size] have to be before you can't use the mean to compare them?

The role of zero values when examining and describing distributions Uncertainty regarding whether zero values should be included in calculation of the means for particular conditions surfaced during the bean investigation. During the group presentations Andrew stated his group had "manipulated" the data so as to generate a finding that supported their original hypothesis. In an attempt to further probe this issue, the question of whether it was valid and justifiable not to count zeros was posed in the focus group discussion.

R: Andrew, you said you manipulated the data – how did you do it?

Andrew: We left out all data points that were zero – if we had included them it [the subgroup means] would have averaged out differently –

R: Did you make the decision ahead of time?

Maria: I didn't think it would matter if we counted [the zeros] cause the total number was smaller anyway.

R: The number you divided by didn't count the zeros either?
Maria: No.
R: Is it valid not to count the zeros?
Maria: It would have skewed the data so badly
Stan: I would go back to what the original question was that we wanted to ask. You may care about those that don't germinate. But there may be another scenario where it wouldn't matter.
Andrew: Yeah if you just wanted to find the longest one then yeah those that don't grow don't matter.

We can see in the transcript a number of understandings related to zero. Andrew understands that inclusion of the zeros will change the outcomes of the comparison. Whereas Maria's statement implies the initial belief that not counting zero values would not matter given that the denominator of the mean algorithm adjusts to reflect the number of values being considered. We see, however, that Maria's second comment indicates a change in her thinking. Her comment on skew indicates that the inclusion of zeros will be influential. It was not until Stan commented on the meaning of zero within the context of the research question that the groups' consideration of zero shifted from thinking of zero merely as a number devoid of meaning to consideration of zero as a measure within the context of the investigation.

Examination of the work carried out by groups in both statistical investigations shows that, in general, values of zero were eliminated and not considered as valid data. This has repercussions for reasoning about distributions in that within the context of statistical investigations a measure of zero is a value and its consideration, whether graphical or quantitative, influences the outcome of deliberations. While we never probed the reason behind why zero values were considered inconsequential by a large proportion of the participants, two explanations come to mind. The first hypothesis suggests that a conflation of one type of mathematical understanding of zero and the experimental situation resulted in zero not being considered. From a mathematical perspective participants may have been considering zero as representing the absence of elements, that is, a set of zero objects. In the context of the bean and popcorn experiments, beans that didn't grow and kernels that didn't pop were assigned the value zero, however participants may not have considered the value zero as a quantity but rather as the absence of growth or distance, if a bean didn't grow or a kernel didn't pop then it has no measure and shouldn't be considered. The second hypothesis suggests that participants' actions resulted from over generalizing the property of zero as the identity element in addition. This notion, that zero as an identity element leaves a set unchanged, is true for addition but is not true in the context of the mean. While the numerator of the mean is an addition context, the quantity derived from the addition is then divided by another quantity – in this case zero as the identity element does not hold.

5. DISCUSSION

The first goal of the study was to identify the statistical concepts preservice teachers focus on when analyzing and comparing distributions of data. The findings of this study suggest that elementary preservice teachers' focus is on *summary* rather than *exploration* of datasets resulting in a focus on summary statistics such as measures of central tendency. This overemphasis on centers and corresponding neglect of variation has also been highlighted by Shaughnessy (1992, 1997) in the undergraduate student population. The focus on summary was also evident in that participants did not use graphical representations to support the description and comparison of distributions of data. It seems that when presented with a distribution of data participants did not attempt to represent the distribution in a way that highlighted structural features, thus limiting the accuracy of the conclusions that can be drawn from the data. Shifting attention to exploration *prior* to summarization was not an insurmountable task. Once participants were made aware of the merits of data exploration and their attention drawn to distributional shape, they were eager to utilize a number of alternative measures (for example variation) and representations when comparing and analyzing datasets. Our finding that the increased attention to graphs resulted in revealing aspects of distributions, supporting the findings of Hammerman and Rubin (2004) and Makar and Confrey (2005) who noted a particular emphasis on variability.

The second goal of the study was to examine the ways in which different understandings of these particular statistical concepts support or hinder the description, analysis, and comparison of datasets. As mentioned in the previous paragraph, the overemphasis on measures of central tendency went hand in hand with the neglect of graphs and variability. For many participants, their *lack of exposure* to statistical ideas and statistical inquiry lead to the blanket implementation of measures they were familiar with – the mean invariably. However, once participants' attention was drawn to variation, in concert with an emphasis on how variation is modeled in graphical representations, variation quickly became a central component of participants' understanding of distribution. Similarly, providing experiences which highlighted the functionality of graphical representations, as tools to explore and reveal aspects of distributions, supported a focus on graphs. This resulted in a concentration on the selection of particular representations according to their propensity to reveal features of distributions.

Finally, this research reveals a number of ways to support the development of rich understandings of distribution. Firstly, it is critical that we draw preservice teachers' attention to the notion of distribution; many participants did not hold a distributional view of data. The use of the experimental context supported the construction of distributional perspectives due to the emphasis drawn by the context on the variation of data values along a scale of measurement (i.e., how height of the bean varied within the range of possible heights). Once this notion of distribution was established participants could see the interrelationships between measures of center and variability and the underlying structures of data that they emulate, and recognize the importance of graphs in revealing aspects of data. Secondly, a focus on the dataset as an entity was essential as it provided a meaning and context for the construction of representative values (see also Mokros & Russell, 1995) – values that were initially applied without an underlying rationale. The act of comparing datasets forced the entity view in that the act of comparison required the search for comparison values, each of which needed to be representative of the body of data. Finally, once the notion of distribution is established and the concept of dataset as an entity developed, understandings of distribution can be further nurtured though exposure to the range of measures and representations that support the continued effort of describing, analyzing and comparing distributions. It was surprising to find that many preservice teachers were not adept at constructing, or in some cases even aware of, stem and leaf plots and box and whisker plots. This lack of experience as learners with representations and measures that they may be required to teach in the future is a worrying, yet not surprising, finding as highlighted by Lajoie & Romberg (1998) in their call for teachers to be provided with content experiences in data analysis and statistics. In essence, the study highlights that once provided the *opportunity* to engage in statistical inquiry in conjunction with instruction focusing on data analytic techniques, preservice elementary teachers develop understanding of statistical measures and techniques and utilize them in statistically sound and justifiable ways.

5.1. LIMITATIONS

There are a number of limitations of this study which are related primarily to the research design and the nature of the experiences presented in the class. When considering the research design, the research participants are not representative of the general population of preservice teachers. Given their undergraduate degrees this study may overestimate the mathematical content knowledge of the general population of elementary preservice teachers. Other design limitations, related to the complexities of carrying out research in educational settings, carry with them implications related to the study validity. For example, there not was random assignment of participants to the study and there was not a control group. These factors pose threats in terms of the internal validity of the findings.

5.2. IMPLICATIONS

One challenge prospective teachers face is examining the structure of their own mathematical knowledge in an effort to come to know what it means to understand a particular concept. So when thinking about teaching statistics a preservice teacher may ask what it means for a 6^{th} grade student to understand the mean. This requires examining what it is we know about the mean, how we came to know "it," what "it" contributes to our understanding, and how we might embark on supporting students in developing similar understandings. So, we might agree, if asked, that the mean is a value that is not necessarily represented in the dataset. But if asked to list all we know about the mean we may not list this particular piece of knowledge – so how do we know what we know? How do we start to decompress (Ball & Bass, 2000) our mathematical knowledge? These types of understandings are what Ball and Bass (2000) call knowledge packages and Ma (1999) calls knowledge bundles – they are the fundamental understandings, connections, ideas that teachers need to develop (or may already possess) so that our knowledge becomes more accessible as a resource for teaching. This study provides insights into the ways in which engaging in statistical inquiry, wherein one is accountable for justifying the tools used to explore a distribution of data, challenges learners to question what it is they know and how this knowledge can be used in ways that are mathematically sound and justifiable. In their current form, traditional methods of teaching the pedagogy of data analysis and statistics fail to engage prospective teachers in examining their own knowledge for teaching.

This study also suggests that when considering the mathematical preparation of teachers we cannot assume that preservice teachers have sufficient exposure to statistical measures and techniques, in particular the construction and selection of appropriate graphical representations. Even when preservice teachers demonstrate knowledge of how to generate particular measures; this study shows that they may not recognize situations in which to use these measures, their understandings were what Skemp (1979) would categorize as primarily instrumental with poor relational understandings. It seems that participants did not have an adequate picture of the landscape of data analysis – in other words, what measures are available, when we might use them, and why. These findings suggest strongly that when considering the mathematical preparation of teachers that the focus be placed on what Hiebert and Lefevre (1986) define as knowing *how-to* (procedural understanding) and *why* (conceptual understanding) we use particular statistical techniques and representations.

There are a number of ways in which future research relating to preservice teacher statistical knowledge can build on and extend the findings of this study. It is important to investigate whether similar studies using different types of statistical investigations result in similar conclusions as this study. This would have implications not only for the ways in which we research preservice teacher's statistical understandings but also for the ways in which we structure our instruction in elementary schools. An extension of this work would be to investigate the ways in which providing experiences in data modeling at the preservice level influences the teaching practices of prospective teachers when they enter classrooms.

5.3. RECOMMENDATIONS

The final paragraph of the discussion poses recommendations for ways to support the development of rich understandings relating to distribution. The remainder of this section poses recommendations for ways to support learners build conceptual understanding and skills.

Firstly, while all students seemed to be good consumers of statistical information, in other words they demonstrated skills in graphical interpretation and comprehension (reading the data, reading between the data, and reading beyond the data), the majority exhibited difficulties constructing graphical representations. Efforts should be made to provide preservice teachers opportunities to work with real data and engage in activities related to constructing graphical representations. Secondly, of those students who demonstrated skills in deriving descriptive statistics and constructing graphical representations, relational understanding of these measures was absent. For example, some participants wished to construct scatter plots of the univariate data and persisted in trying to manipulate the data so that it would be amenable to presentation on a scatter plot. This again relates to their primary experiences as consumers of statistics – they have not been in the position of having to select appropriate statistics and representations for particular purposes. This finding calls for a coordinated effort to provide experiences that allow preservice teachers to consider the appropriateness of particular measures and representations for the purposes of data analysis. Thirdly, and not unrelated to the previous points, was the poor conceptual understanding of descriptive statistics and graphical representations, calling for a more conceptual focus in mathematics and methods courses. It also cautions teacher educators against drawing conclusions about conceptual understandings based on demonstrated proficiency in generating and applying measures (such as the mean for data comparison purposes). Finally, our observations lead to the conjecture that gains in understanding demonstrated over the course of the semester were influenced primarily by having access to strategies used by peers when engaging in data description and comparison activities. While classroom teaching experiences supported the development of skills and conceptual understanding, what they seemed not to do was convince participants of the utility of such measures when engaged in data analysis. In other words, our analysis indicated that factors other than classroom teaching were more influential in convincing students to apply new concepts and skills when engaging in statistical inquiry. It became apparent that engagement in small group statistical inquiry acted as a conduit whereby prospective teachers observed and gained access to the complex decision making processes of others when engaged in exploratory data analysis and then compared those decisions against their own. Such experiences provide opportunities for participants to learn in practice, to develop communities of learners who engage in authentic statistical inquiry, and who continuously seek to find more efficiently and statistically justifiable ways of thinking about distributions of data.

ACKNOWLEDGEMENTS

I would like to acknowledge the contribution of Christy Graybeal, Toni Smith and Carolina Napp to the design and implementation of this research.

REFERENCES

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 83-104). Westport, CT: Abex.

Ball, D. L., & McDiarmid, G. W. (1990). The subject matter preparation of teachers. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 437-449). New York: Macmillan.

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

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*(3), 194-222.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology*, *31*(6), 839-855.

Canada, D. (2004). *Preservice teachers*' *understanding of variation.* Unpublished doctoral dissertation, Portland State University.
[Online: www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php]

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, *1*(1), 5-43.

Cobb, P., Gravemeijer, K. P. E., Bowers, J., & McClain, K. (1997). *Statistical Minitools.* Designed for Vanderbilt University, TN, USA. Programmed and revised (2001) at the Freudenthal Institute, Utrecht University, The Netherlands.

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

Friel, S. N., Mokros, J. R., & Russell, S. J. (1992). *Used numbers. Statistics: Middles, means, and in-betweens.* Palo Alto, CA: Dale Seymour Publications.

Gfeller, M. K., Niess, M. L., & Lederman, N. G. (1999). Preservice teachers' use of multiple representations in solving arithmetic mean problems. *School Science and Mathematics, 99*(5), 250-257.

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/serj/SERJ3(2).pdf]

Hancock, C. (1995). *TableTop Software.* Novato, CA: Broderbund Software Direct.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337-364.

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35-59.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Konold, C. (1998). *Tinkerplots: Tools and curricula for enhancing data analysis in the middle school.* Grant proposal submitted to the National Science Foundation (# ESI-9818946). University of Massachusetts, Amherst.

Konold, C., & Miller, C. (2001). *Tinkerplots* (version 0.23). Data Analysis Software. University of Massachusetts, Amherst.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics: 1996 Proceedings of the 1996 IASE Round Table Conference* (pp. 151-167). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/8/13.Konold.pdf]

Lajoie, S. P. (1997). The use of technology for modelling performance standards in statistics. In J. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 57-70). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/8/5.Lajoie.pdf]

Lajoie, S. P., & Romberg, T. A. (1998). Identifying an agenda for statistics instruction and assessment in K–12. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. vii–xxi). Mahwah, NJ: Lawrence Erlbaum Associates.

Lappan, G., Fey, J. T., Fitzgerald, W. F., Friel, S. N., & Phillips, E. D., (2002). *Connected Mathematics Project*. Needham, MA: Prentice Hall.

Leavy, A. M., & O'Loughlin, N. (2006). Moving beyond the arithmetic average: Preservice teachers understanding of the mean. *Journal of Mathematics Teacher Education, 9*(1), 53-90.

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

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education*. (pp. 665-708). Mahwah, NJ: Lawrence Earlbaum Associates.

Lubienski, S. T., & Bowen, A. (2000). Who's counting? A survey of mathematics education research 1982-1998. *Journal for Research in Mathematics Education, 31*(5), 626-633.

Ma, L. (1999). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States. Mahwah, NJ: Lawrence Erlbaum Associates.

Makar, K., & Confrey, J. (2002, July). Comparing two distributions: Investigating secondary teachers' statistical thinking. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/10_18_ma.pdf]

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

Mathematical Sciences Advisory Committee, 1984-1985. Academic preparation in mathematics: Teaching for transition from high school to college. Cited in: Beres, R.J. (1988). Statistics for college bound students: Are the secondary schools responding? *School Science and Mathematics*, *88*(3), 200-209.

McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical: Developing a conjectured learning trajectory. *Journal of Mathematical Behavior, 21*, 339-335.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics*, *14,* 415-429.

Mokros, J., & Russell, S. J. (1995). Children's concept of average and representativeness. *Journal for Research in Mathematics Education*, *26*(1), 20-39.

National Center for Research in Mathematical Sciences at the University of Wisconsin/Madison and Freudenthal Institute at the University of Utrecht (1997-1998). *Mathematics in context*. Chicago: Encyclopedia Britannica.

National Council of Teachers of Mathematics Commission on Standards for School Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Pollatsek, S. J., Lima, S., & Well, A. D. (1981). Concept or understanding: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematical teaching and learning* (pp. 465-494). New York: McMillan.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Vol. 1, pp. 6-22). Waikato, New Zealand: Mathematics Education Research Group of Australasia.

Simon, M. A. (2000). Research on the development of mathematics teachers: The teacher development experiment. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 335-361). Mahwah, NJ: Lawrence Earlbaum Associates.

Skemp, R. R. (1979). *Intelligence, learning and action: A new model for the theory and practice in education*. New York: Wiley.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Lawrence Earlbaum Associates.

Stein, M. K., Baxter, J., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal*, *27*(4), 639-663.

Strauss, S., & Bichler, E. (1988). The development of children's concept of the arithmetic average. *Journal for Research in Mathematics Education, 19*(1), 64-80.

Technical Education Research Centers and Lesley College (1989). *Used numbers: Real data in the classroom*. Menlo Park, CA: Dale Seymour Publications.

TERC (1998). *Investigations in number, data, and space*. Menlo Park, CA: Dale Seymour Publications.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire: Graphics Press.

Watson, J. M., & Kelly, B. A. (2002, August). *Can grade 3 students learn about variation?* Paper presented at The Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/2a1_wats.pdf]

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

AISLING LEAVY
R124 Mary Immaculate College, South Circular Road
University of Limerick
Limerick, Ireland

APPENDIX A: SMALL GROUP REPORT GUIDELINES FOR THE BEAN INVESTIGATION

THE BEAN EXPERIMENT: SMALL GROUP ACTIVITY

Group members: ____

Step 1: Re-examine the primary research scenario and experiment design

> Ben likes bean sprouts in his salads and sandwiches. Lately he has been unhappy with the quality of bean sprouts available at the grocery store so he has decided to grow his own. He suspects that lighting conditions and the addition of lemon juice to the water may affect the length of the bean sprouts. Ben wants to grow the longest bean sprouts possible. Briefly describe how he could determine which growing conditions (Light vs. Dark, Plain water vs. Water with lemon juice) support the best growth.

		Solution	
		Water	Lemon
Light Intensity	Light	A	B
	Dark	C	D

Step 2: Construct <u>research questions</u> you may examine in an effort to investigate Ben's suspicion.

Step 3: What <u>comparison method(s)</u> would you utilize to examine the data *and consequently answer your questions.*

Step 4: Construct a hypothesis describing what you will believe will be the outcome of the data analysis.

Now use the data from the bean sprouts to answer your question. Prepare a poster to present your analysis and be prepared to discuss the questions below.

Step 5:

a) Explain why you chose your particular comparison method(s):

b) Explain what (if any) limitations there are of this method(s):

c) Present the answer to your research question. How did you come to this conclusion?

Time to reflect!

d) You have now viewed the presentations and approaches your peers took when approaching the same task. What would you do if you had the opportunity to complete your analysis again – would you change your approach?

If so, why? And what would you do differently?

APPENDIX B: SMALL GROUP REPORT GUIDELINES FOR THE POPCORN INVESTIGATION

USING AN EXPERIMENT TO TEST A CONJECTURE – IT'S ALL ABOUT POPCORN!
SMALL GROUP ACTIVITY

Group members: ____

There are some people who say that refrigerating popcorn kernels prior to popping changes certain characteristics of the kernels. In this experiment we are going to examine one such conjecture.

Step 1: Make a list of factors that you think could be influenced by refrigeration.

Today, we are going to investigate:

> *Does refrigerating corn for 12 hours prior to popping influence either (a) the number of corn that pop over a 4 minute period, or (b) the distance that the corn falls/jumps from the popper?*

Step 2: <u>Record</u> the data for the first experiment: popping kernels for 4 minutes in an uncovered popper in the space below:

Step 3: <u>Record</u> the data for the second experiment: popping <u>refrigerated</u> kernels for 4 minutes in an uncovered popper in the space below:

Step 4:
Now use the data from both experiments to answer the research question. Prepare a poster to present your analysis and be prepared to discuss the questions below.

a) Explain why you chose your particular comparison method(s):

b) Explain what (if any) limitations there are of this method(s):

c) Present the answer to the research question. How did you come to this conclusion?

Time to reflect!

d) You have now viewed the presentations and approaches your peers took when approaching the same task. What would you do if you had the opportunity to complete your analysis again – would you change your approach?

If so, why? And what would you do differently?

PAST IASE CONFERENCES

ICOTS-7: WORKING COOPERATIVELY IN STATISTICS EDUCATION
Salvador (Bahia), Brazil, July 2-7, 2006

The International Association for Statistical Education (IASE) and the International Statistical Institute (ISI) organized the Seventh International Conference on Teaching Statistics (ICOTS-7) which was hosted by the Brazilian Statistical Association (ABE) in Salvador (Bahia), Brazil, July 2-7, 2006.

Information about ICOTS-7 is on the website: http://www.maths.otago.ac.nz/icots7

Abstracts for all papers (plenary, invited and contributed) and a list of all posters are on the website. ICOTS-7 papers are available at http://www.stat.auckland.ac.nz/~iase/publications.php

OTHER PAST CONFERENCES

JOINT STATISTICAL MEETINGS 2006
Seattle WA, USA, August 6 – 10, 2006

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the American Statistical Association, the International Biometric Society (ENAR and WNAR), the Institute of Mathematical Statistics, and the Statistical Society of Canada.

The conference produced an impressive abstract book of 530 pages, which is available at the conference website or directly. The abstract book is packed with interesting abstracts of papers dealing with statistics education, statistical literacy, and other issues of interest. Most of the interesting papers can be found by use of keyword "Statistical Education."

Website: http://www.amstat.org/meetings/jsm/2006/

Statistics Education Research Journal, 5(2), 115-123, http://www.stat.auckland.ac.nz/serj

FORTHCOMING IASE CONFERENCES

SRTL-5
THE FIFTH INTERNATIONAL RESEARCH FORUM ON STATISTICAL REASONING, THINKING, AND LITERACY
Coventry, UK, August 11 - 17, 2007

Reasoning about Statistical Inference:
Innovative Ways of Connecting Chance and Data.

The Forum's focus will be on informal ideas of inference rather than on formal methods of estimation and tests of significance. This topic is emerging from the presentations and discussions at SRTL-3 and 4 and is a topic of current interest to many researchers as well as teachers of statistics. As new courses and curricula are developed, a greater role for informal types of statistical inference is anticipated, introduced early, revisited often, and developed through use of simulation and technological tools. We encourage research papers that address reasoning about statistical inference at all levels of education including the professional development of elementary and secondary teachers.

TOPICS

Research presented at the conference will address questions such as the following:

1. What are the simplest forms of statistical inference that students can understand?
2. How does reasoning about statistical inference develop from the simplest forms (informal) to the more complex ones (formal)?
3. How can instructional tasks and technological tools be used to promote the understanding of statistical inference?
4. What are sequences of activities that can help student develop a conceptual understanding of statistical inference?
5. What types of misconceptions are found in students' reasoning about statistical inference?
6. What types of foundational knowledge and reasoning are needed for students to understand and reason about statistical inference?
7. How do students develop an understanding of the language used in describing statistical inference (e.g., significance, confidence)?
8. How does an understanding of statistical inference connect and affect understanding of other statistical concepts?
9. What are useful items and questions to use to assess understanding of statistical inference?

THE LOCAL SRTL-5 ORGANIZERS

Janet Ainley, janet.ainley@warwick.ac.uk
Dave Pratt, dave.pratt@warwick.ac.uk
For more information visit the SRTL-5 website: http://srtl.stat.auckland.ac.nz/

IASE SATELLITE CONFERENCE ON ASSESSING STUDENT LEARNING IN STATISTICS
Guimaraes, Portugal, August 19-21, 2007

THEME

This satellite conference invites papers on all aspects of assessing student learning in statistics. For example, we expect to have papers on writing effective exam questions, on exam implementation strategies, and on alternative assessment methods such as projects, lab assignments, and writing assignments. We also encourage submissions on how to use assessment to improve student learning, and on developing and administering assessments items to conduct research into student learning. Proceedings will be available free at the publication page of IASE

CONFERENCE COMMITTEE

Brian Phillips (Australia) (Joint Chair and Joint Chief Editor) bphillips@swin.edu.au
Beth Chance (USA) (Joint Chair) bchance@calpoly.edu
Allan Rossman (USA) arossman@calpoly.edu
Ginger Rowell (USA) rowell@mtsu.edu
Gilberte Schuyten (Belgium) gilberte.schuyten@UGent.be
Larry Weldon (Canada) (Joint Chief Editor) weldon@sfu.ca
Local Organiser: Bruno C. de Sousa (Portugal) bruno@mct.uminho.pt

For more information visit the website:
http://www.stat.auckland.ac.nz/~iase/conferences.php?show=iasesat07

ISI-56
THE 2007 SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE
Lisboa, Portugal, August 22 – 29, 2007

The 56th Session of the International Statistical Institute (ISI) will be held in Lisboa, Portugal. As it does at each major ISI conference, IASE will be organizing about 10 statistics education sessions for ISI-56. Please check the website at http://www.isi2007.com.pt/ for more information, and contact the session organizers below if you would like to offer to speak in one of the sessions.

IASE SPONSORED IPMS

IPM37 *Research on Reasoning about Distribution,*
Joan Garfield (jbg@umn.edu)

IPM38 *How Modern Technologies Have Changed the Curriculum in Introductory Courses,*
Lucette Carter (lucette.carter@gmail.com)

IPM39 *Preparing Teachers of Statistics,*
Allan Rossman (arossman@calpoly.edu)

IPM40 *Research on the Use of Simulation in Teaching Statistics and Probability,*
Rolf Biehler (biehler@mathematik.uni-kassel.de)

IPM41 *Optimizing Internet-based Resources for Teaching Statistics* (cosponsored by IASC),
Ginger Holmes Rowell (rowell@mtsu.edu)
IPM42 *Observational Studies, Confounding and Multivariate Thinking,*
Milo Schield (milo@pro-ns.net)
IPM43 *Teaching of Official Statistics* (cosponsored by IAOS) ,
Sharleen Forbes (Sharleen.Forbes@stats.govt.nz)
IPM44 *Teaching of Survey Statistics* (cosponsored by IASS),
Steve Heeringa (sheering@isr.umich.edu)
IPM45 *Studying Variability through Sports Phenomena* (cosponsored by Sports Statistics),
TBD
IPM46 *Use of Symbolic Computing Systems in Teaching Statistics* (cosponsored by IASC),
Zaven Karian (Karian@Denison.edu)
IPM92 *Statistical Education and Literacy in the 21st Century* (cosponsored by INE),
Pedro Campos (pcampos@fep.up.pt)

IASE ORGANIZING COMMITTEE:

Allan J. Rossman (USA) arossman@calpoly.edu
Gilberte Schuyten (Belgium) gilberte.schuyten@UGent.be
Chris Wild (New Zealand) c.wild@auckland.ac.nz

For more information visit the ISI 56 website at http://www.isi2007.com.pt/ or contact members of IASE OC.

JOINT ICMI /IASE STUDY
STATISTICS EDUCATION IN SCHOOL MATHEMATICS: CHALLENGES FOR TEACHING AND TEACHER EDUCATION
Monterrey, Mexico, June 30 to July 4, 2008

The International Commission on Mathematical Instruction (ICMI, http://www.mathunion.org/ICMI/) and the International Association for Statistical Education (IASE, http://www.stat.auckland.ac.nz/~iase/) are pleased to announce the Joint ICMI /IASE Study Statistics Education in School Mathematics: Challenges for Teaching and Teacher Education.

Following the tradition of ICMI Studies, this Study will comprise two parts: the Joint Study Conference and the production of the Joint Study book. The Joint Study Conference will be merged with the IASE 2008 Round Table Conference.

The Joint Study Conference (ICMI Study and IASE Round Table Conference) will take place at the Instituto Tecnológico y de Estudios Superiores, Monterrey, Mexico (http://www.mty.itesm.mx/), from June 30 to July 4, 2008. Participation in the Conference is only by invitation, based on a submitted contribution and a refereeing process. Accepted papers will be presented in the Conference and will appear in the Proceedings that will be published by ICMI and IASE as a CD-ROM and on the Internet.

The second part of the Joint Study – the Joint Study book – will be produced after the conference and will be published in the ICMI Study Series. Participation in the Joint Study Conference does not automatically assure participation in the book, since a second selection and rewriting of selected papers will be made after the conference.

Proposed papers for contributions to the Joint Study Conference should be submitted by e-mail no later than October 1, 2007, to the IPC Study Chair (Carmen Batanero, batanero@ugr.es). Papers should be relevant to the Joint Study focus and research questions, as described in the Discussion Document (which is available at the Joint Study Website (http://www.ugr.es/~icmi/iase_study/). Guidelines for preparing and submitting the paper are also available in the Discussion Document. Please address questions to Carmen Batanero, batanero@ugr.es or Joan Garfield, jbg@umn.edu .

INTERNATIONAL PROGRAMME COMMITTEE:

Carmen Batanero (Spain)
Bernard Hodgson (Canada, ICMI representative)
Allan Rossman (USA, IASE representative)
Armando Albert (México, ITSM representative)
Dani Ben-Zvi (Israel)
Gail Burrill (USA)
Doreen Connor (UK)
Joachim Engel (Germany)
Joan Garfield (USA)
Jun Li (China)
Maria Gabriella Ottaviani (Italy)
Maxine Pfannkuch (New Zealand)
Mokaeane Victor Polaki (Lesotho)
Chris Reading (Australia)

LOCAL ORGANISING COMMITTEE:

Blanca Ruiz (Chair)
Tomás Sánchez
Armando Albert

More information is available from Carmen Batanero, batanero@ugr.es or from http://www.ugr.es/~icmi/iase_study/

ICOTS-8
DATA AND CONTEXT IN STATISTICS EDUCATION: TOWARDS AN EVIDENCE-BASED SOCIETY
Ljubljana, Slovenia, 11-16 July 2010

We are pleased to announce that the IASE Executive accepted the proposal made by the Statistical Society of Slovenia to hold ICOTS-8 in 2010 in Slovenia.

The decision was announced at the ICOTS-7 farewell dinner. The first steps towards organising have already been taken by the IASE Executive: The conference theme has been chosen and the Scientific and Local Committees have been appointed.

The conference theme *Data and Context in Statistics Education* emphasises two concepts that are key concepts at nearly all levels of statistics education. The subtitle *Towards an Evidence-Based Society* offers a gateway to reflections about the past, present and future status of statistics in society and about the impact of statistics education on learning objectives. We shall learn about statistics and see how we learn through the use of statistics.

THE INTERNATIONAL PROGRAMME COMMITTEE EXECUTIVE:

IPC Chair: John Harraway
Programme Chair: Roxy Peck
Information Manager: John Shanks
Scientific Secretary: Helen MacGillivray
Editor Proceedings: Alan McLean

THE LOCAL ORGANISING COMMITTEE

LOC Chair: Andrej Blejec.

OTHER FORTHCOMING CONFERENCES

THE 11TH ASIAN CONFERENCE IN MATHEMATICS
ATCM 2006
Hong Kong SAR, China, December 12-16, 2006

CONFERENCE THEME

The aim of this conference with the theme *Advancing and Fostering Mathematical Sciences and Education through Technology* is to provide a forum for educators, researchers, teachers and experts in exchanging information regarding enhancing technology to enrich mathematics learning, teaching and research at all levels. English is the official language of the conference.

TOPICS OF INTERESTS

The conference will cover a broad range of topics on the application and use of technology in mathematics research and teaching. Though statistics can be recognized in many proposed themes, a special theme, *Statistics using Dynamic Statistics Software*, might be of particular interest.

Website: http://www.atcminc.com/mConferences/ATCM06/index.shtml

THE 6TH ANNUAL HAWAII INTERNATIONAL CONFERENCE ON STATISTICS, MATHEMATICS AND RELATED FIELDS
Honolulu HI, USA, January 17 – 19, 2007

The 6th Annual Hawaii International Conference on Statistics, Mathematics and Related Fields will be held at the Renaissance Ilikai Waikiki Hotel in Honolulu, Hawaii. The 2007 Hawaii International Conference on Statistics, Mathematics and Related Fields will be the gathering place for academicians and professionals from statistics and mathematics related fields from all over the world.

The main goal of the 2007 Hawaii International Conference on Statistics, Mathematics and Related Fields is to provide an opportunity for academicians and professionals from various statistics and/or mathematics related fields from all over the world to come together and learn from each other. An additional goal of the conference is to provide a place for academicians and professionals with cross-disciplinary interests related to statistics and mathematics to meet and interact with members inside and outside their own particular disciplines.

Website: http://www.hicstatistics.org/

USCOTS 2007
UNITED STATES CONFERENCE ON TEACHING STATISTICS
Columbus OH, USA, May 17-19, 2007

The second biennial United States Conference on Teaching Statistics (USCOTS 2007) will be held on May 17-19, 2007 at the Ohio State University in Columbus, Ohio, hosted by CAUSE, the Consortium for the Advancement of Undergraduate Statistics Education. The target audience for USCOTS is teachers of undergraduate and AP statistics, from any discipline or type of institution. Teachers from two-year colleges are particularly encouraged to attend.

The theme for USCOTS 2007 is *Taking Statistics Teaching to the Next Level*. “Next level” has many interpretations, such as developing a second course, gaining more confidence in teaching statistics, moving students beyond statistical literacy to statistical thinking, and using the latest technology to improve teaching and learning. USCOTS is a "working conference" with many opportunities for hands-on activities, demonstrations, networking, idea sharing, and receiving the latest information on research and best practices in teaching statistics. Leaders in statistics education will give plenary talks, including Jessica Utts, Paul Velleman, Dick DeVeaux, Allan Rossman, and Mike Shaughnessy.

For more information and registration, visit the USCOTS web page http://www.causeweb.org/uscots/

2007 JOINT STATISTICAL MEETINGS
Salt Lake City UT, USA, July 29 - August 2, 2007

The 2007 Joint Statistical Meetings will be held July 29 - August 2, 2007 at the Salt Palace Convention Center located at 100 South West Temple, Salt Lake City, Utah 84101.

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the American Statistical Association, the International Biometric Society (ENAR and WNAR), the Institute of Mathematical Statistics, and the Statistical Society of Canada. Attended by over 5000 people, activities at the meeting include oral presentations, panel sessions, poster presentations, continuing education courses, an exhibit hall (with state-of-the-art statistical products and opportunities), career placement service, society and section business meetings, committee meetings, social activities, and networking opportunities. Salt Lake City is the host city for JSM 2007 and offers a wide range of possibilities for sharing time with friends and colleagues. For information, contact jsm@amstat.org

Website: http://www.amstat.org/meetings/jsm/2007/

JOINT SOCR (STATISTICS ONLINE COMPUTATIONAL RESOURCE) CAUSEWAY CONTINUING EDUCATION WORKSHOP 2007
UCLA, Los Angeles CA, USA, 6-8 August 2007

The 2007 joint SOCR/CAUSEway continuing education workshop aims at demonstrating the functionality, utilization and assessment of the current UCLA, SOCR and CAUSEweb resources. This workshop will be of most value to AP teachers and college instructors of probability and statistics classes who have interests in exploring novel IT-based approaches for enhancing statistics education. The workshop will provide an interactive forum for the exchange of ideas and recommendations for strategies to integrate computers, modern pedagogical approaches, the Internet and new student assessment techniques.

For further information:
http://wiki.stat.ucla.edu/socr/index.php/SOCR_Events_Aug2007/

9TH INTERNATIONAL CONFERENCE OF THE MATHEMATICS EDUCATION INTO THE 21ST CENTURY PROJECT MATHEMATICS EDUCATION IN A GLOBAL COMMUNITY
Charlotte NC, USA, September 7 - 13, 2007

The Mathematics Education into the 21st Century Project has just completed its eighth successful international conference in Malaysia, following conferences in Egypt, Jordan, Poland, Australia, Sicily, Czech Republic and Poland. Our project was founded in 1986 and is dedicated to the planning, writing and dissemination of innovative ideas and materials in Mathematics and Statistics Education. The next conference is planned for September 7 - 13, 2007 in Charlotte, North Carolina. The chairman of the Local Organising Committee is Dr. David K. Pugalee, of the University of North Carolina Charlotte. The conference will open with an evening welcome reception on Friday, September 7th and finish with lunch on September 13th. The title of the conference is *Mathematics Education in a Global Community*. Papers are invited on all innovative aspects of mathematics education. Our conferences are renowned for their friendly and productive working atmosphere. They are attended by innovative teachers and mathematics educators from all over the world, 25 countries were represented at our last conference for example.

More information: Alan Rogerson, arogerson@inetia.pl

Website: http://math.unipa.it/~grim/21project.htm

STATISTICS EDUCATION RESEARCH JOURNAL REFEREES DECEMBER 2005-NOVEMBER 2006

Acknowledgement: We are grateful to the following who acted as referees for the *Statistics Education Research Journal* in the past year.

Carmen Acuna, USA
Janet Ainley, UK
Arthur Bakker, UK
Mustafa Baloglu, Turkey
Andrej Blejec, Slovenia
Carol Joyce Blumberg, USA
Marjorie Bond, USA
Manfred Borovcnik, Austria
Lea Bregar, Slovenia
Tim Burgess, New Zealand
José María Cardeñoso, Spain
Beth Chance, USA
Helen Chick, Australia
Matt Ciancetta, USA
José Luis Cortina, Mexico
Cileda Coutinho, Brazil
Carolyn Cuff, USA
Geoff Cumming, Australia
Bob delMas, USA
Kirsten Doehler, USA
Felix Famoye, USA
Mary Fraire, Italy
Christa Fratto, USA
Joan Garfield, USA
Linda Gattuso, Canada
Randall Groth, USA
Corinne Hahn, France
Jim Hammerman, USA
John Harraway, New Zealand
Diane Johnson, USA
Peter Johnston-Wilder, UK
Ruben Klein, Brazil
David Lane, USA
Cynthia Langrall, USA
Hollylynne Stohl Lee, USA
Nancy Leech, USA
Li Jun, China
Jiajuan Liang, USA
Kay Lipson, Australia
Marsha Lovett, USA
Dasen Luo, USA
Rhonda Magel, USA
Katie Makar, Australia
Claudine Mary, Canada
Elsa Medina, USA
Jamie Mills, USA
Edward Mooney, USA
Mari Murtonen, Finland
Nyaradzo Mvududu, USA
Irena Ograjenšek, Slovenia
Tony Onwuegbuzie, USA
Gabriella Ottaviani, Italy
Elena Papanastasiou, Cyprus
Lionel Pereira-Mendoza, Singapore
Susan Peters, USA
Peter Petocz, Australia
Maxine Pfannkuch, New Zealand
Mokaeane Victor Polaki, Lesotho
Dave Pratt, UK
Jackie Reid, Australia
Ernesto Sanchez, Mexico
Juana Sanchez, USA
Richard Scheaffer, USA
Mary Jane Schmitt, USA
Gilberte Schuyten, Belgium
Carl Schwarz, Canada
Norbert Schwarz, USA
Gary Smith, USA
Eric Sowey, Australia
Juhani Tahtinen, Finland
James Tarr, USA
Kostas Triantafyllopoulos, UK
Jessica Utts, USA
Barbara Ward, USA
Jane Watson, Australia
Sharon Weinberg, USA
Joseph Wisenbaker, USA
Roger Woodard, USA