Statistics Education Research Journal

Volume 8 Number 1 May 2009

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@jcu.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

TABLE OF CONTENTS

EDITORIAL

Welcome to Volume 8 of *SERJ*! My term as co-editor will expire at the end of this year. At that point, I am pleased to note that I will have been co-editor for half of *SERJ*'s existence. The end of my term signals another important transition for *SERJ*. Please see the "Call for Nominations" beginning on page 3 for a description of the job of co-editor, and please feel free to self-nominate. I think that Iddo Gal, who overlapped with me as co-editor during the first half of my term, would agree that he and I are not true statistics education researchers. We served as "caretaker" co-editors, developing the vision and maintaining the processes for *SERJ* as the statistics education research field continues to mature.

The current continuing co-editor, Peter Petocz, is an established statistics education researcher. Peter will provide the leadership to publish a special issue on "Qualitative Issues in Statistics Education Research" in 2010. See page 5 for more information on this special issue.

It is vital that a second leader in the discipline come forward to join Peter in the role of co-editor. Please consider self-nominating or encouraging qualified researchers you know to consider the position.

SERJ has grown in two notable dimensions during my term as co-editor. First, we are receiving more manuscripts than in the past. As I write, we have received 18 manuscripts in 2009. In 2008 we received our 18th manuscript in June, and in 2007 the 18th manuscript arrived in August. Clearly, the field is growing and the number of articles *SERJ* publishes will grow as well. Second, *SERJ* is now indexed in high-visibility databases such as EBSCOhost and PSYCinfo. We are currently in discussions with ERIC to list *SERJ* as well.

Of course, *SERJ* continues to face challenges. Our review process is often unacceptably slow. As the co-editor managing the review process, I take responsibility for the delays that have frustrated authors and reviewers. The Associate Editors and reviewers must also accept the workload that accompanies increasing submission rates. Thanks to Joan Garfield and her graduate students at the University of Minnesota, we have a new database of *SERJ* reviewers available. The new database will not only expand the list of potential reviewers but will also facilitate matching reviewers with manuscripts relevant to reviewer expertise. It is difficult to adequately blind the identities of authors and reviewers, in light of immediate Web search capabilities and the inclusion of non-print technologies in some submissions. As Iddo has noted in the past, one of the the biggest challenge we face is to mentor researchers, authors, reviewers, and Editorial Board members from developing countries.

As the end of my term approaches, I thank Carmen Batanero, Flavia Jolliffe, Iddo Gal, Peter Petocz, Chris Wild, and especially Joan Garfield and Beth Chance for their advice, guidance, prodding, and support over the past four years. I have found the position of *SERJ* co-editor to be educational and rewarding.

I hope that you enjoy the articles published in this first issue of 2009. Thank you for the opportunity to serve you as co-editor.

TOM SHORT

Statistics Education Research Journal, 8(1), 2, http://www.stat.auckland.ac.nz/serj

CALL FOR NOMINATIONS FOR NEW CO-EDITOR: *STATISTICS EDUCATION RESEARCH JOURNAL*

DEADLINE FOR SUBMISSION OF NOMINATIONS: 15 JULY 2009

The International Association for Statistical Education (IASE) is starting a search for the next co-editor of *Statistics Education Research Journal* (*SERJ*), its peer-reviewed electronic journal. The new editor will serve a four-year term starting 1 January 2010, replacing Tom Short (John Carroll U., USA), who will end his four-year tenure at that time, and ending 31 December 2013. The new editor will join Peter Petocz (Macquarie U., Australia), the continuing co-editor until December 2011.

1. ABOUT SERJ

SERJ was established in 2002 by IASE to advance research-based knowledge that can help to improve the teaching, learning, and understanding of statistics and probability at all educational levels and in both formal and informal contexts. *SERJ* presently publishes two issues per year but is likely to move up to three issues per year in coming years. The scope of submitted manuscripts represents the growing interest in research and in new knowledge that can inform practice in statistics education.

The *SERJ* organization includes two co-editors who serve for four years (one is replaced every two years), an Assistant Editor in charge of copy-editing and production, and an Editorial Board presently comprised of 16 Associate Editors from 10 countries. *SERJ* issues and materials are published on the IASE website, presently hosted by the University of Auckland (www.stat.auckland.ac.nz/~iase/). The journal maintains autonomy regarding content and process, although some activities are coordinated with IASE and its parent organization and *SERJ* co-publisher, the International Statistical Institute (ISI). All journal activities are conducted electronically. Board members meet during key international conferences such as ICOTS or ISI Biennial Sessions. *SERJ* is a virtual organization and it operates on the basis of voluntary work by all board members and editors.

2. THE CO-EDITORS ROLE AND EXPECTED QUALIFICATIONS

The co-editors are responsible for overall management of all journal operations. They manage peer-review and editorial processes, determine the composition of the Editorial Board and the reviewer pool, and initiate and conduct communication with prospective authors, reviewers, associate editors, and external stakeholders. The co-editors are expected to establish editorial policies, set scholarly and quality expectations, and uphold acceptance criteria regarding manuscripts. The co-editors should have a forward-looking vision and initiate new features and structures, if needed in consultation with Board members and others, so as to enable *SERJ* to respond to the evolving knowledge needs in the dynamic area of statistics education. Overall, the co-editors should lead the journal to make an important contribution to research and practice in statistics education.

The qualified individual will have a strong research background in areas which are part of statistics education, and be familiar with educational practice in this area. He or

Statistics Education Research Journal, 8(1), 3-4, http://www.stat.auckland.ac.nz/serj

she should possess the skills necessary to work with prospective contributors in a supportive yet critical spirit, be able to maintain and strengthen international professional networks of authors and reviewers, and enhance the *Journal*'s reputation and impact.

3. THE SEARCH PROCESS AND HOW TO MAKE NOMINATIONS

Review of nominations will begin on 15 July 2009, but nominations should be submitted as soon as possible. IASE encourages both nominations of suitable candidates and self-nominations from interested individuals. All nominations and self-nominations will be considered by the Search Committee, which can also propose additional nominees. Candidates or self-nominees are asked to send an academic vita or professional resumé, together with a brief statement describing their vision for continuing the growth and development of the *Journal*, and their qualifications for the position. Candidates might also be asked to respond to additional questions by the Search Committee.

Please send nominations (with all supporting materials listed above) or questions to the Chair of the Search Committee, Iddo Gal, (U. of Haifa. Israel), at: <iddo@research.haifa.ac.il>. (Questions about the practicalities of the editorship can also be sent either to the continuing co-editor, Peter Petocz <ppetocz@efs.mq.edu.au> or to the departing co-editor, Tom Short <tshort@jcu.edu>).

CALL FOR PAPERS: QUALITATIVE APPROACHES IN STATISTICS EDUCATION RESEARCH

The *Statistics Education Research Journal* (*SERJ*), a journal of the International Association for Statistical Education (IASE), is planning a special issue for November 2010, focused on research on the topic of *Qualitative Approaches in Statistics Education Research*. **Submission deadlines: Letters of intent by Sept. 14, 2009; Full papers by Nov. 2, 2009**. The Guest Editors for this issue will be Sue Gordon (University of Sydney, Australia, <s.gordon@usyd.edu.au>) and Anna Reid (Macquarie University, Australia, <anna.reid@mq.edu.au>).

1. RESEARCHING STATISTICS EDUCATION: QUANTITATIVE OR QUALITATIVE

Education is a social activity with strong elements of unpredictability. Students enter with a variety of different educational and life experiences, and many of these aspects are difficult to analyse using statistical methods – even when the students are studying statistics. An obvious way forward is to use research approaches that focus on the social and personal aspects of learning. Qualitative research approaches can capture and explain the more experiential dimensions of learning, illuminate pedagogical issues, and complement and enrich results that may be obtained from quantitative studies. Much mainstream research in education utilises qualitative methods to examine learning situations, often through analysing linguistic or socio-cultural elements. As with quantitative research, qualitative methods must demonstrate rigour and validity according to recognised criteria appropriate to these methods.

Although previous editions of *SERJ* contain examples of qualitative approaches to research, these approaches seem less utilised compared with other areas of pedagogical research. For statisticians, the quantitative approach to research problems seems to be the natural one, even when examining aspects of statistics pedagogy. However, different research questions are often addressed – and sometimes answered – by different research approaches, and there are cases where a qualitative approach, or a ‘mixed methods’ approach, is more fruitful than a quantitative one. For some statisticians, qualitative approaches to research can seem alien, or not in their area of (statistical) expertise, and hence they may try to avoid them. Within statistics, rigour is usually represented by the use of statistical techniques; in some cases, this approach is even supported by law in the form of government policies for accountability of research funding. Nevertheless, there are many situations where a statistical approach contains qualitative elements, for instance, the wording or selection of questions for a survey, the decision of what to measure and how to measure it, and the very notion of classification.

In his editorial in volume 6, number 2 of *SERJ*, Iddo Gal referred to “dynamic data,” which he defined as the information that is collected when research is carried out in situations where students are using dynamic software packages or interactive applets (though the term has a more common use in referring to situations where information is used to update a data set). Such data consist of information about what students actually did, what they said during the process, and how they interpreted the results. In other disciplines, this is sometimes referred to as “observational ethnography” and can embrace issues of ethnic diversity, life experience, narrative as inquiry and theory development.

Statistics Education Research Journal, 8(1), 5-6, http://www.stat.auckland.ac.nz/serj

This special issue will focus on qualitative approaches to research in statistics education with the aim of illuminating practice or theory, highlighting pedagogical issues and critically examining and reflecting on diverse perspectives in the teaching and learning of statistics at all levels.

2. POSSIBLE TOPICS

Examples of topics that would be relevant for this special issue on *Qualitative Approaches in Statistics Education Research* include, but are not limited to:

a. Papers reporting the results of empirical studies in statistics education carried out using particular qualitative methodologies, or with ‘mixed methods’ approaches.
b. Papers reporting on dynamic learning situations where students learn through using statistical packages, for example, in research in statistics education.
c. Articles that critically appraise the use of qualitative methods of any type in statistics education research.
d. Reviews of qualitative research in statistics education.

3. SUBMISSION GUIDELINES

Authors are advised to aim for papers in the range of 4000-6000 words of body text (not counting abstract, tables and graphs, references, appendices). Manuscripts for the special issue will be limited to a *maximum* of 7500 words of body text, but shorter, concise papers are encouraged. All manuscripts will be refereed following *SERJ*’s regular double-blind peer-review process. Manuscripts should be submitted in accordance with *SERJ*’s standard Author Guidelines and using the Template file found on the *Journal*’s website: www.stat.auckland.ac.nz/serj.

4. DEADLINES AND CONTACT INFORMATION

Interested authors should send *a letter of intent by Sept. 14, 2009*, but preferably earlier, with a 150-250 word abstract describing key aspects of the research. This letter should be sent by e-mail to *SERJ* co-editor Peter Petocz: <Peter.Petocz@mq.edu.au>, and authors can expect to get a response within two weeks. Authors wishing to send informal queries regarding the suitability of a planned paper can also contact Peter.

Full manuscripts must be submitted by Nov. 2, 2009 at the latest to Peter Petocz at the address above, in accordance with the submission guidelines listed earlier.

Decisions about the suitability of proposed papers and the allocation of accepted papers to the special issue or to a regular *SERJ* issue will be made jointly by the *SERJ* Editors and Guest Editors.

MODELING THE GROWTH OF STUDENTS' COVARIATIONAL REASONING DURING AN INTRODUCTORY STATISTICS COURSE

ANDREW S. ZIEFFLER
University of Minnesota
zief0002@umn.edu

JOAN B. GARFIELD
University of Minnesota
jbg@umn.edu

ABSTRACT

This study examined students' development of reasoning about quantitative bivariate data during a one-semester university-level introductory statistics course. There were three research questions of interest: (1) What is the nature, or pattern of change in students' development in reasoning throughout the course?; (2) Is the sequencing of quantitative bivariate data within the course associated with differences in the pattern of change in reasoning?; and (3) Are changes in reasoning about foundational concepts of distribution associated with differences in the pattern of change? Covariational and distributional reasoning were measured four times during the course, across four cohorts of students. A linear mixed-effects model was used to analyze the data, revealing some interesting trends and relationships regarding the development of covariational reasoning.

Keywords: *Statistics education research; Growth modeling; Topic sequencing*

1. THE IMPORTANCE OF UNDERSTANDING COVARIATION

Reasoning about *association* (or *relationship*) between two variables, also referred to as *covariational reasoning*, or reasoning about *bivariate data*, involves knowing how to judge and interpret a relationship between two variables. Covariational reasoning has also been defined as the cognitive activities involved in coordinating two varying quantities while attending to the ways in which they change in relation to each other (Carlson, Jacobs, Coe, Larsen, & Hsu, 2002). This type of reasoning may take a very mathematical form (e.g., a linear function), a statistical form (reasoning about a scatterplot), or a more qualitative form (e.g., causal predictions about events, based on observed associations, such as spending more time studying seems to lead to better test grades, as described in causal model theory in psychology). Covariational reasoning is also viewed as playing an important role in scientific reasoning (Koslowski, 1996; Schauble, 1996). Although covariation between events is a necessary but not sufficient basis for inferring a causal relationship, it is a basis for making causal inductive inferences in science (Zimmerman, 2005).

The concept of covariation may be unique in that it is an important concept in the different fields of psychology, science, mathematics, and statistics, and that covariational reasoning is described somewhat differently in each discipline. Statisticians may be

Statistics Education Research Journal, 8(1), 7-31, http://www.stat.auckland.ac.nz/serj

surprised that reasoning about covariation, which they think of as a statistical topic focusing on bivariate distributions of data, is much more complex than the familiar caution that "correlation does not imply causation," and beyond reasoning about scatterplots, correlation, and regression analyses. Indeed, cognitive psychologists McKenzie and Mikkelsen (2007) wrote that covariational reasoning is one of the most important cognitive activities that humans perform.

Davis (1964) summed up the goal of education well when he wrote, "The primary object of teaching is to produce learning (that is, change), and the amount and kind of learning that occur can be ascertained only by comparing an individual's or a group's status before the learning period with what it is after the learning period" (p. 234). This idea of measuring "change" is even more salient in the current era of educational research (e.g., see *No Child Left Behind Act of 2001*; United States Department of Education, 2005).

The study described in this paper attempted to examine the development, or change, in students' reasoning about quantitative bivariate data over the span of an entire introductory statistics course. Furthermore, this study examined whether students' development of reasoning about quantitative bivariate data can be explained by other factors that have been identified in the research literature.

2. REVIEW OF THE LITERATURE

Because of its important role in so many disciplines, covariational reasoning has been the focus of research in psychology, science, and mathematics education, in addition to statistics education. The research studies related to covariational understanding and reasoning are quite diverse, and vary according to the disciplinary field of the researchers. Therefore, we summarize in the following section the main contributions from each of these different disciplines.

2.1. RESEARCH STUDIES IN PSYCHOLOGY, MATHEMATICS EDUCATION, AND SCIENCE EDUCATION

Research by psychologists provides much of the foundational work in covariational reasoning. Since the early studies by Inhelder and Piaget (1958), psychologists have documented the importance of covariational reasoning in the day-to-day lives of people. These studies document that people are surprisingly poor at assessing covariation and that prior beliefs about the relationship between two variables have a great deal of influence on their judgments of the covariation between those variables (e.g., Jennings, Amabile, & Ross, 1982; Kuhn, Amsel, & O'Loughlin, 1988). The psychological research also shows that one particular shortcoming that people have when intuitively assessing covariation, is to believe that there is a correlation between two uncorrelated events, because they believe they are related. Referred to as an illusory correlation, this phenomenon has been offered as a cognitive explanation for stereotypic judgments (see Hamilton & Gifford, 1976; McGahan, McDougal, Williamson, & Pryor, 2000).

Many of the psychology studies examined how people reason about covariation of data in contingency tables (e.g., Kao & Wasserman, 1993). Some of the results have found that people have difficulty when the relationship is negative (e.g., Beyth-Marom, 1982), and that peoples' covariational judgment of the relationship between two variables tends to be less than optimum (i.e., smaller than the actual correlation presented in the data or graph) especially when they believe there is a relationship between the two variables in question (e.g., Jennings et al., 1982). A consistent finding in several studies is that people have a tendency to form causal relationships based on a covariational analysis in almost every situation where they have prior beliefs about the relationship (e.g., Ross & Cousins, 1993).

A different focus is found in studies conducted by mathematics education researchers on covariational reasoning, which is used extensively in both algebra (Nemirovsky, 1996) and calculus (Thompson, 1994). Many of these studies have examined students' understanding of functions, or aspects of bivariate reasoning that are commonly used in algebra and calculus (e.g., Carlson et al., 2002). In particular, studies have suggested that this type of reasoning plays a major role in students' understanding of the derivative, or rate of change (e.g., Carlson et al.), and that this interpretation of covariation is slow to develop among students (e.g., Monk & Nemirovsky, 1994; Nemirovsky, 1996). Studies from mathematics education have also shown that not only is students' ability to interpret graphical and functional information slow to develop, but that students tend not to see the graph of a function as depicting covariation (Thompson, 1994).

Research studies in science education research have examined aspects of covariation found in both the psychological studies (e.g., confusing correlation and causation; e.g., Adi, Karplus, Lawson, & Pulos, 1978) and the mathematical studies of covariation (e.g., reasoning about lines and functions in the context of science problems; e.g., Wavering, 1989). A third type of science education study focuses on more of the statistical aspects of science. For example, Kanari and Millar (2004) examined students' approaches to data collection and interpretation as they investigated relationships between variables, as part of students' ability to reason from data. The authors found that students of all ages had a much lower success rate in investigations where the dependent variable did not covary with the independent variable, than in those where it did covary. They suggested that school science investigations should include both covariation and non-covariation cases to develop students' covariational reasoning.

2.2. COVARIATIONAL REASONING AND JUDGMENTS IN STATISTICS EDUCATION RESEARCH

The newly emerging field of statistics education research includes studies of students' covariational reasoning in the context of instruction in statistics. The impact of computers in developing students' covariational reasoning was studied by Batanero, Estepa, Godino, and Green (1996) and Batanero, Estepa, and Godino (1997). They identified several misconceptions and errors students make when reasoning about covariation. For example, these studies revealed the persistence of a *unidirectional misconception*, meaning that students only perceive a relationship between two variables if it is positive.

Both studies also showed that students maintained their causal misconception throughout the duration of the experiments, and that students had problems with several aspects associated with covariational reasoning, such as distinguishing between the role of independent and dependent variables and reasoning about relationships that were negative. Finally, students realized that the absolute value of the correlation coefficient was related to the magnitude of the relationship, but did not relate that idea to the spread of scatter around the regression line.

Other studies have examined students' covariational reasoning as they study regression and reported some of the difficulties associated with this topic including problems with interpretation (e.g., Sánchez, 1999), and problems with the coefficient of determination, or R^2 (Truran, 1997). Konold (2002) presented a different view of whether or not people can make accurate covariational judgments when presented with contingency tables or scatterplots. He suggested that people are not poor at making these judgments, but rather they have trouble decoding the ways in which these relationships are displayed (e.g., scatterplots or contingency tables).

In a study of younger children, Moritz (2004) had students translate verbal statements to graphs and also translate a scatterplot into a verbal statement. The students were also given a written survey that included six or seven open-ended tasks involving familiar variables. The variables were chosen so that students would expect a positive covariation, but the data given in the task represented a negative covariation. Moritz found many of the same student difficulties as other studies have revealed: that students often focused on isolated data points rather than on the global data set (e.g., Ben-Zvi & Arcavi, 2001); that students would often focus on a single variable rather than the bivariate data; and that several students had trouble handling negative covariations when they are contradictory to their prior beliefs.

Two design experiments investigated the role of technology in helping students reason about bivariate data, and how students differentiate between local and global variation in bivariate data. Gravemeijer's (2000) results suggested that students need an idea of the global trend (prior expectation) and that students have a hard time distinguishing between arbitrary and structural covariation. He suggested that students examine and compare several univariate data sets (time series) as an introduction to examining bivariate data.

This approach was used by Cobb, McClain, and Gravemeijer (2003) to help students view bivariate data as distributed in two-dimensional space, to see scatterplots as situational texts, and to track the distribution of one variable across the other (scan vertically rather than diagonally). Using the *Minitools* software (Cobb, Gravemeijer, Bowers, & Doorman, 1997) students examined the "vertical variation" across levels of *x* in graphs of bivariate data. Students were asked to compare differences in the distribution of the *y*-variable at different levels of the *x*-variable (see Figure 1).

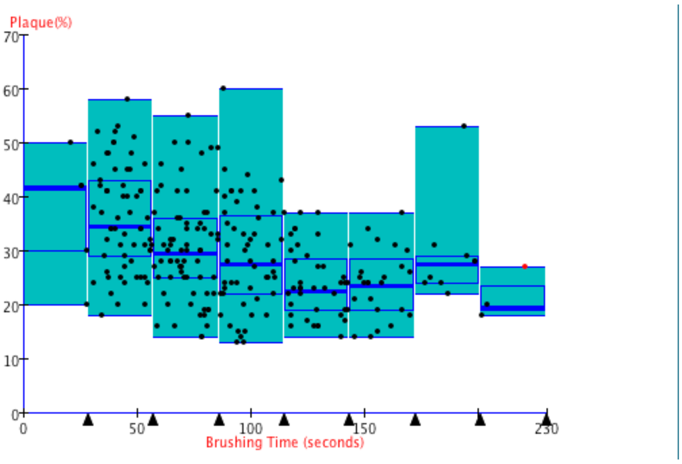


Figure 1. Minitools software allows students to start looking at the local variation for different values on the x-axis in addition to the global trend

The results of their study suggested that the shape of a distribution is a better place to start than is variability and that there be a continued focus on relative density and on the shape of the data within vertical slices. They also suggested that an emphasis on shape could lead to a discussion of strength and direction in a bivariate plot and that the focus on vertical distribution could lead to a more intuitive idea of the line of best fit.

2.3. UNIVARIATE DISTRIBUTION AS THE FOUNDATION FOR COVARIATIONAL REASONING

Recent research has pointed to the importance of building up a foundation for covariation upon the building blocks of distribution (e.g., Cobb, 1998; Cobb et al., 2003; Gravemeijer, 2000; Konold, 2002; Konold & Higgins, 2003). Cobb et al. and Gravemeijer (2000) have suggested that a deep understanding of characteristics of distribution – such as shape, center and variation – is important foundational knowledge in a complete understanding of bivariate data. Building on the ideas of distribution is also congruent with Ben-Zvi and Garfield’s (2004) recommendation of focusing on big ideas to provide a foundation for course content and develop the underpinnings of statistical reasoning.

Cobb et al. (2003) have hypothesized that a focus on graphs and shape is an important piece of statistics students’ development. They suggested that a focus on shape will make it easier for students to transition to reading a bivariate plot (scatterplot) because students were able to find it reasonable to talk about and compare the distribution within different vertical slices of the bivariate distribution. This, in turn, will “provide a basis for a subsequent focus on trends and patterns in an entire data set” (Cobb et al., p. 84). Gravemeijer (2000) also suggested that students begin by comparing univariate data sets, but instead of the focus on shape in the vertical slices, he posited that the median might be a better comparison. He purported that students can then focus on a global trend by examining the median of the vertical distribution across measures of the horizontal (*x*) variable. Still other statistics educators have suggested that variation might be the piece of pre-requisite knowledge that mandates the most attention, pointing out that in fact, covariation concerns the correspondence of variation among two or more variables (e.g., Moritz, 2004).

2.4. SUMMARY OF THE RESEARCH

Looking at the studies across the different disciplines, we note the following general findings:

- Students’ prior beliefs about the relationship between two variables have a great deal of influence on their judgments of the covariation between those variables;
- Students often believe there is a correlation between two uncorrelated events (illusory correlation);
- Students’ covariational judgments seem to be most influenced by the joint presence of variables and least influenced by the joint absence of variables;
- Students have difficulty reasoning about covariation when the relationship is negative;
- Students’ covariational judgment of the relationship between two variables tends to be less than optimum (i.e., smaller than the actual correlation presented in the data or graph); and
- Students have a tendency to form causal relationships based on a covariational analysis.

Taken as a whole, the research on covariational reasoning has examined many questions about misconceptions and difficulties that students have in reasoning about covariation, and has suggested methods for introducing and developing these ideas. However, there are many research questions yet unanswered. With enrollment in undergraduate statistics courses increasing (College Board, 2003) it is important that educators strive to understand and improve students' ability to reason with and understand covariation.

Although researchers have examined peoples' covariational reasoning on both dichotomous and continuous variables, there have been few studies that have examined the development of students' reasoning about covariation in an introductory statistics course and the optimal placement of bivariate quantitative data analysis. The literature reviewed has suggested that students' reasoning about covariation could be influenced by several factors, including students' developing reasoning about univariate distribution. Therefore, three research questions were used to frame this study:

1. What is the nature, or pattern of change in students' development in reasoning about quantitative bivariate data throughout an introductory statistics course?
2. Is the sequencing of quantitative bivariate data within a course associated with differences in the pattern of change in students' reasoning about quantitative bivariate data?
3. Are changes in students' reasoning about the foundational concepts of distribution associated with differences in the pattern of change in students' reasoning about quantitative bivariate data?

3. METHODOLOGY

3.1. OVERVIEW OF STUDY

This study took place during the fall semester of the 2005/2006 school year. It involved four cohorts of a one-semester (three credit hours), non-calculus based introductory statistics course taught in the College of Education at a mid-western university in the United States of America. Two different instructors taught these four cohorts. All four cohorts met in a computer lab two times a week for an hour and fifteen minutes each time. Each of these cohorts had an enrollment of about 30 students.

This study utilized linear mixed-effects models (LMM) to examine change in students' development of reasoning about quantitative bivariate data. Because the modeling of change requires individuals to be measured on the same concept in temporal sequence, a repeated-measures, or longitudinal design was employed. Students enrolled in a collegiate level introductory statistics course were assessed on their reasoning about quantitative bivariate data four times during a semester. Examining the change in students' reasoning about quantitative bivariate data over these four time points addressed the first research question.

To examine the association between course sequencing and the patterns of change in students' reasoning about quantitative bivariate data, the two instructors of the course used in the study were used as blocks to randomly assign each cohort of the course to one of two different course sequences (see Table 1). These two sequences both started with the topics of sampling and exploratory data analysis (EDA). Then the first sequence continued with the topic of quantitative bivariate data followed by sampling distributions, probability, and inference. The second sequence followed EDA with sampling distributions, probability, inference, and ended the course with the topic of quantitative bivariate data.

Table 1. The two sequences taught fall semester 2005

Sequence 1:					
Sampling →	EDA →	Bivariate Data →	Sampling Distribution →	Probability →	Inference
(6 Days)	(7 Days)	(4 Days)	(3 Days)	(2 Days)	(6 Days)
Sequence 2:					
Sampling →	EDA →	Sampling Distribution →	Probability →	Inference →	Bivariate Data
(6 Days)	(7 Days)	(3 Days)	(2 Days)	(6 Days)	(4 Days)

To examine whether changes in students' reasoning about the foundational concepts of distribution were associated with changes in the development of students' reasoning about quantitative bivariate data, students were also assessed on their distributional reasoning four times during the course of the semester.

3.2. SETTING

The study participants consisted of n = 113 undergraduate students. These students were typically female social science majors (84% females and 16% males) who were enrolled in the course to complete part of their graduation requirements. These students belong to the larger population of undergraduate social science majors who take an introductory statistics course in an Educational Psychology department.

This particular introductory statistics course was designed so that it was aligned with recent Guidelines for the Assessment and Instruction in Statistics Education (GAISE; see American Statistical Association, 2005a) endorsed by the American Statistical Association (American Statistical Association, 2005b). In addition, the course materials were based on what has been learned from research literature on teaching and learning statistics. The unit on quantitative bivariate data was designed to help students avoid common errors and difficulties identified in the research literature and to build a solid understanding and good reasoning based on the results of best practices and research results. Overall, the research literature guided both the structure of the course (i.e., scope and sequence) and the instructional methods (i.e., activities, technologies, and discussions) used within the course. The course included collecting and analyzing real data sets, software programs to illustrate abstract concepts, and many active learning techniques. Lesson plans for every instructional session were created during the initial design phase of the course in the summer of 2004, which included class goals, discussion questions, and a sequence of activities. These lesson plans helped provide more consistency across multiple cohorts of the course taught by different instructors. These materials were used, evaluated, and revised during the two semesters prior to the study.

The two instructors teaching the four cohorts followed identical lesson plans throughout the duration of the course and met regularly to help ensure consistency among the cohorts. Both of the instructors had helped develop the course materials and had taught the course multiple times prior to the time of this study. Both instructors were experienced teachers, having both high school and college teaching experience, and were doctoral students in the Quantitative Methods in Education (QME) program with a concentration in Statistics Education, so they were also familiar with the current statistics education guidelines and relevant research.

3.3. INSTRUMENTS

Three instruments were administered to students to collect data on their reasoning and their background characteristics. These were: 1) the Bivariate Reasoning Assessment, 2) the Distributional Reasoning Scale, and 3) the Student background survey. Descriptions of each instrument follow.

Bivariate Reasoning Assessment (BR) Students' covariational reasoning was measured using the quantitative bivariate data scale from the Assessment Resource Tools for Improving Statistical Thinking (see Appendix for the instrument; Garfield, delMas, & Chance, n.d.). The eight forced-choice items assessed reasoning and interpretation regarding the correlation coefficient and relationships between the correlation coefficient and the display of data in a scatterplot. These items seem aligned with important aspects of bivariate reasoning indentured in the statistics education literature (e.g., Mortiz, 2004).

Distributional Reasoning Scale (DR) Ten items from the Comprehensive Assessment of Outcomes in a First Statistics Course (CAOS; available from ARTIST, Garfield et al., n.d.) were used to measure students' reasoning about univariate distribution. Experts have identified these items as focusing on reasoning about univariate distribution. They included items on interpreting different graphical displays, drawing conclusions from data, and reasoning about variation.

Student Background Survey (SBS) To help determine whether the randomization process was effective, and also to identify which covariates might be important in explaining the pattern of students' development of reasoning about bivariate data, several different instruments were combined and used to gather data. These survey items assessed students' prior mathematical (10 items) and statistical (30 items) knowledge, as well as identifying students' academic background (4 items) and prior coursework in mathematics, statistics, and computer science (15 items). Each of these instruments is described in much greater detail in Zieffler (2006).

Instrument administration Each of the research instruments was administered on the first day of class (Session 1) to obtain baseline measures. The BR and DR instruments were also administered during three other class periods (Session 14, Session 25, and Session 29). These assessments were administered in Session 14 and Session 25 because those were the two classroom sessions that immediately preceded instruction of bivariate data for each of the two course sequences listed in Table 1. The assessment was also given during the last classroom session of the semester (Session 29).

The items from these two instruments were combined into one comprehensive instrument to ease the actual administration, and the items were randomized for each of the four administrations. This comprehensive instrument was administered during class time to ensure test security and integrity. Because of the difficulty associated with assessing students multiple times without feedback, students were offered extra credit to participate in the study.

3.4. DATA ANALYSIS

In this section, the analysis used to answer each of the research questions is described. Before these descriptions are offered, a brief explanation of linear mixed-effects models, the primary analysis method used, is given.

Linear mixed-effects models Researchers interested in studying change are generally interested in answering two types of questions about change (Boyle & Willms, 2001). The first of these questions of interest is how to "characterize each person's pattern of change over time," and the second asks about "the association between predictors and the patterns of change" (Singer & Willett, 2003, p. 8). The statistical models that researchers use to examine change go by a variety of names – random coefficients models, mixed-effects models, hierarchical linear models (HLM), or multilevel models are just a few. These models provide a statistical methodology that allows researchers to answer both types of questions about change, and in addition have many advantages over traditional statistical methods such as RM-MANOVA, including the accommodation of missing data (e.g., Collins, Schafer, & Kam, 2001) and flexibility in model specification which can lead to greater power and efficiency in estimation (e.g., Verbeke & Molenberghs, 2000).

The linear mixed-model (LMM) used for this study is a multi-level regression model that incorporates two components: a level-1 linear model that describes intra-individual (within subjects) change, and a level-2 conditional model that describes systematic inter-individual (between subjects) differences in change. In the level-1 model, time is used as the independent variable for predicting individual students' baselines (starting points) and trajectories (shape or pattern of the curve) in their reasoning about bivariate data. The level-2 models allow us to determine the extent that those baselines and trajectories vary as a function of one or more covariates (i.e., other measured variables, such as previous achievement, that are used to differentiate individuals). For a more detailed explanation of the LMM methodology, see Verbeke and Molenberghs (2000) or Raudenbush and Bryk (2002).

Unconditional model analysis To explore students' change in development in reasoning about bivariate data, an unconditional LMM was fitted to the data to describe the pattern of change exhibited in the data. An important piece of the mixed-effects model methodology is the correct specification of the model including both the fixed and random effects, as well as the within-group covariance structure. In the tradition of mixed-effects models analysis, diagnostic strategies such as graphs and sample statistics were employed to help provide guidance for this specification. More formal specifications to further substantiate the appropriate structure of the level-1 model were made by computing and comparing model estimates and fit statistics.

Conditional model analyses A conditional LMM was used to help provide answers for the second and third research question. A conditional model allows for predictors other than just time. To answer the second research question, the two instructional sequences were effect coded and introduced into the model for change that is adopted. To answer the third research question, the change in students' reasoning about univariate distribution was quantified and entered as a predictor in the model for change.

4. RESULTS

The data analyses and results are presented in three sections, one for each the three research questions. All analyses were carried out using *R* version 2.2.1 (*R* Development Core Team, 2008). The mixed-effects modeling utilized the `lme4` (Bates & Sarkar, 2005) and `nlme` (Pinheiro, Bates, DebRoy, & Sarkar, 2005) libraries. For more detailed descriptions of all the analyses presented in this section see Zieffler (2006).

Initial analyses of several measured covariates using the Student Background Survey (not presented) suggested that the randomization process seemed to have been effective in producing groups with equivalent student characteristics (see Zieffler, 2006 for more detail). Examination of the sample scores and responses for all instruments showed sufficient reliability, using Cronbach's coefficient alpha (Cronbach, 1951), for research purposes (all were above .71).

Table 2 shows the average student score on the four administrations of both the distributional and the bivariate reasoning assessments. It is not surprising that the students began the class with a very low mean score on the BR, but it was surprising that the largest increase came between the beginning of course and the 14th instructional session (before any formal instruction on bivariate data). It was also surprising that the mean score at the end of the class was barely over 50% correct, revealing the difficulty students have reasoning about bivariate data. The same pattern is also seen in students' DR scores. In both instructional sequences, the average distributional reasoning score increased. The greatest increase occurred between the first and second measurement occasions.

Table 2. Means (standard deviations), on the bivariate and distributional reasoning assessment for all measurement occasions for both instructional sequences

	Distributional Reasoning (DR)		Bivariate Reasoning (BR)	
Class Session	Sequence 1[a]	Sequence 2	Sequence 1	Sequence 2
Session 1	0.56 (1.04)	1.18 (1.43)	1.00 (1.03)	0.79 (1.23)
Session 14	7.31 (1.69)	7.50 (1.74)	3.84 (1.53)	4.09 (1.61)
Session 25	7.51 (1.70)	7.55 (1.41)	5.12 (1.48)	4.61 (1.63)
Session 29	7.56 (1.77)	7.50 (1.57)	4.57 (1.58)	5.02 (1.53)

Note. The DR had a possible range of 0 to 10, with higher numbers indicating a higher perceived degree of reasoning. The BR had a possible range of 0 to 8, with higher numbers indicating a higher perceived degree of reasoning.

[a]Sequence 1 taught bivariate data early and inference later. Sequence 2 taught inference early and bivariate data later (see Table 1).

4.1. RESULTS OF FITTING THE UNCONDITIONAL MODEL

To explore students' change in development in reasoning about bivariate data, a LMM was fitted to the data to describe the pattern of change exhibited in the data. Based on the results of several analyses (not presented), a quadratic level-1 model was employed to model the mean within-student change in reasoning about quantitative bivariate data. A random-effects structure with unstructured residuals was also adopted and used in all subsequent analyses. Lastly, several model comparisons seemed to suggest that the best fitting model to the data would have random-effects associated with both the linear and quadratic terms but not with the intercept term. Exploratory analysis on the residuals of the fitted models [distribution of standardized residuals against the grouping factor (i.e., the random effect) and against fitted values, separately for each level of the classification factor (i.e., the fixed effect)] revealed that the model assumptions were adequately met, according to the inspection criteria described by Pinheiro and Bates (2000). The parameter estimates for the unconditional model appear in Table 3.

Table 3. Unconditional model used to describe students' change in reasoning about quantitative bivariate data (n = 113)

		Unconditional Model
Fixed Effects		
	Intercept	0.90***
	Linear term	0.32***
	Quadratic term	-0.01***
Variance Components		
Level-1	Within-student	1.23***
Level-2	*Linear Term*	
	Variance	0.0148**
	Quadratic Term	
	Variance	0.0000124*
	Covariance with linear term	-0.000407*
Goodness-of-fit		
	-2LogLikelihood	1432.9
	AIC	1446.9
	BIC	1475.0

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.
Note. This model was fitted using Restricted Maximum Likelihood in *R*.

Interpretation of the parameter estimates for the second unconditional model The sample fixed-effects estimate the average initial score, linear rate of change, and quadratic rate of change on the BR. Each of the three fixed-effects is statistically significant ($p < 0.001$). This average within-student trajectory is plotted in Figure 2. The within-student variance component summarizes the average scatter of an individual student's observed BR score around his/her change trajectory. This estimate is statistically significant ($p < 0.001$) which suggests that there is still within-student variation to account for.

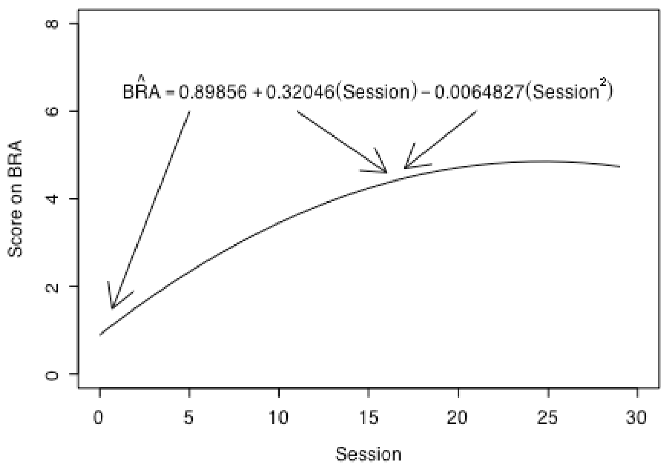


Figure 2. Predicted change in quantitative bivariate reasoning for an average student

The level-2 variance components quantify the amount of unpredicted variation in the individual growth parameters. Though the estimated variance components for the linear rate of change and the quadratic rate of change both seem to be non-zero in the population ($p < 0.01$ and $p < 0.05$ respectively), their practical significance is questionable. The covariance, which is also significant ($p < 0.05$), informs us of the relationship between linear rate of change and quadratic rate of change. Interpretation can be easier if the covariance is re-expressed as a correlation coefficient of -0.94.

We conclude that the relationship between the average linear rate of change and quadratic rate of change in students' ability to reason about quantitative bivariate data is both negative and strong and, because the hypothesis test is significant, is believably non-zero. This indicates that students who have higher linear rates of change also tend to have lower quadratic rates of change.

This model suggests that students, on average, have some ability to reason about quantitative bivariate data before any instruction on bivariate data (e.g., before Session 14) in an introductory statistics course as indicated by the significance of the intercept fixed-effect term. There also seems to be very little variability in students' baseline reasoning about quantitative bivariate data. In other words, they all seem to be starting at the same place. The significance of the positive linear fixed-effect term suggests that students, on average, are increasing their level of reasoning about quantitative bivariate data throughout an introductory statistics course, but this growth does not persist due to the negative quadratic fixed-effect term. Eventually, due to mathematical reasons alone, the quadratic term will remove more than the linear term will add, causing the trajectory to peak and then decline, assuming the relationship continues in this manner. Both of these rates of change vary from student-to-student.

4.2. RESULTS OF FITTING THE FIRST CONDITIONAL MODEL

A conditional LMM was used to help provide an answer for the second research question. To answer this research question, the two instructional sequences were effect coded and introduced into the quadratic model for change that was adopted in the previous section. A model including cross-level interaction terms between the covariate and each level-1 predictor was initially fitted to the data and refined.

Interpretation of the parameter estimates for the first conditional model This model included instructional sequence as a predictor of initial status, as well as both linear and quadratic change. Interpretation of its six fixed-effects (which are not presented) are straightforward: (1) the estimated score on the BR for all students at the beginning of an introductory statistics course is on average 0.90 ($p < 0.0001$); (2) the estimated mean difference in initial BR score between students on average and those taking a class that uses the second instructional sequence (coded 1) is -0.07 points ($p = 0.49$); (3) the estimated average linear rate of change in BR score for all students is 0.32 ($p < 0.0001$); (4) the estimated average difference between the overall average linear rate of change and students in classes that taught the second instructional sequence is -0.00004 ($p = 0.999$); (5) the estimated average quadratic rate of change for all students is -0.01 ($p < 0.0001$); (6) and lastly the estimated average difference in quadratic rate of change for students enrolled in courses that taught the second instructional sequences is 0.0002 ($p = 0.78$).

These results suggest that on average, students in both sequences have similar development in their reasoning about bivariate data throughout an introductory statistics course. In other words, the initial differences in average BR scores between students taking a course that utilized the first instructional sequence and students taking a course that utilized the second instructional sequence are indistinguishable from zero. Likewise, the differences in average linear rate of change and average quadratic rate of change are also not indistinguishable from zero.

The significant within-student variance component in the conditional model is virtually identical to that from the unconditional model. This is expected because there were no level-1 predictors that were added to this model. Both of the level-2 variance components are also essentially unchanged. These conditional variances quantify the inter-individual differences in linear and quadratic change, respectively, that remain unexplained by the predictor.

4.3. RESULTS OF FITTING THE SECOND CONDITIONAL MODEL

To answer this research question, MANOVA was initially employed to examine and summarize the change in students' reasoning about distribution. Because not all students had a measurement at the fouth timepoint, only 98 of the 113 students were used in these analyses. The results of these analyses (not presented) suggested that the difference scores between the first and last measurement occasions could be used as a proxy for describing the change in students' development in reasoning about distribution. These scores were then mean centered (DIST), to facilitate interpretations, and entered as predictors in a conditional LMM. A model including cross-level interaction terms between the covariate and each level-1 predictor was initially fitted to the data and refined. The parameter estimates for the conditional model appear in Table 4.

Table 4. Conditional model to examine students' change in reasoning about univariate data as a predictor of change in students' reasoning about quantitative bivariate data (n = 98)

		Conditional Model
Fixed Effects		
Initial Status	Intercept	0.86***
	DIST	0.13**
Linear rate of change	Linear term	0.32***
Quadratic rate of change	Quadratic term	-0.00658***
Variance Components		
Level-1	Within-person	1.10***
Level-2	*Linear Term*	
	Variance	.01**
	Quadratic Term	
	Variance	0.0000140
	Covariance with linear term	-0.000421
Goodness-of-fit		
	-2LogLikelihood	1237.2
	AIC	1253.2
	BIC	1284.3

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.
Note. This model was fitted using Restricted Maximum Likelihood in *R*.

Interpretation of the parameter estimates for the conditional model The fixed-effects for this conditional model suggest that the only parameter that seems to be influenced by students' change in reasoning about univariate distribution is their initial status in reasoning about bivariate data. The estimated average initial score for students who show average change in their reasoning about univariate data is 0.86 ($p < 0.0001$). The estimated strength of association between initial BR scores and centered DR scores is 0.13 ($p < 0.01$). This result suggests that on average, there is a positive relationship between initial BR scores and centered DR scores indicating that students who exhibit larger than average changes in their reasoning about univariate distribution also tend to have higher initial levels of reasoning about bivariate data. Differences between students' change in reasoning about univariate distribution on average tends not to be associated with either linear or quadratic rates of change in reasoning about bivariate data throughout an introductory statistics course. A visual depiction of this model is shown in Figure 3.

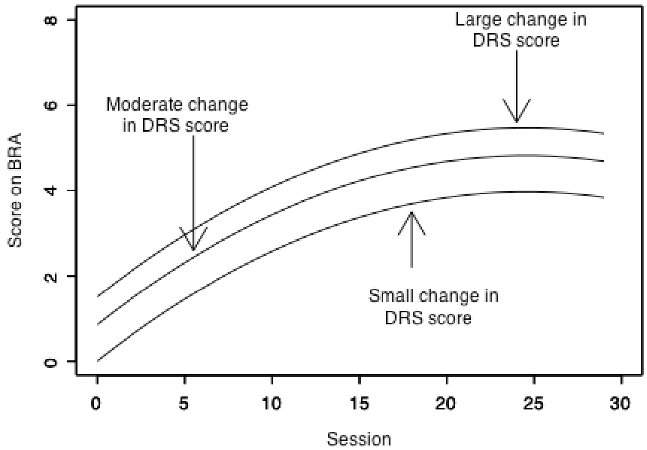


Figure 3. Predicted average change in quantitative bivariate reasoning for students with small, moderate, and large changes in their reasoning about distribution

5. DISCUSSION

This study examined the development of students' reasoning about bivariate data over a 15-week introductory college statistics course. Three research questions were examined and used to structure the collection and analysis of data. The answers to each question are summarized below.

5.1. WHAT IS THE NATURE, OR PATTERN OF CHANGE IN STUDENTS' DEVELOPMENT IN REASONING ABOUT QUANTITATIVE BIVARIATE DATA THROUGHOUT AN INTRODUCTORY STATISTICS COURSE?

Student data collected over the semester revealed marked growth in reasoning about bivariate data but this happened primarily in the first time period. The LMM that was adopted to examine this growth suggested that students exhibit both linear and quadratic growth in their development about reasoning about bivariate data and that this growth varies among individual students. A quadratic model indicates that students' reasoning about bivariate data does not increase in a constant linear fashion, but instead increases differentially over time. The significant negative quadratic term suggests that although students initially show great strides in their reasoning about bivariate data, they likely eventually plane off in this development and over time *might* actually even regress – although given the paucity of measurement occasions used in the study, this regression likely occurs after the course is over. This pattern of development, however, is consistent with several different learning theories (e.g., overlapping waves theory; Siegler, 2000), and might suggest that a saturation point in bivariate reasoning is reached by students and then decay or interference impedes any more growth in reasoning which could actually occur *during* the course (e.g., Wixted, 2004).

The model also suggested that on average students without any instruction start with very little reasoning about bivariate data and that this is true for nearly all students (at this institution). This could be because almost all of the students used in this study had never had a previous high school or college-level statistics course. However, the low initial status leaves much to be desired, especially as covariation is recognized and promoted by the National Council of Teachers of Mathematics in the mathematics curriculum at nearly every age level. This might be explained by the fact that many of these students hadn't had a mathematics course in several years prior to taking statistics, but it might also be because reasoning is not a major focus of most mathematics courses.

Although the fixed-effects and random-effects terms for intercept, linear rate of change, and quadratic rate of change were all statistically significant, the practical significance might not be as important. For instance, the variance term associated with the quadratic rate of change was statistically significant ($p < 0.05$) indicating that students vary in their quadratic rates of change. However, the actual variance term was 0.0000124. This small variance component indicates that students' quadratic rates of change are very similar. Also, comparatively, the within-student variance component still accounts for the majority of the variation in BR scores (98%).

One interesting finding is that most of the change in development in reasoning about bivariate data seemed to occur between the first two measurement occasions. This was before bivariate data was formally taught in either instructional sequence. This might indicate that students' development in reasoning about bivariate data is more an artifact of their development of statistical reasoning in general than it is a result of any formal instruction on the topic of bivariate data. However, the brevity of the unit within this particular introductory statistics class (four instructional sessions) might also inhibit an increase in development of reasoning due to instruction about this topic. It also might mean that students' reasoning about bivariate data is closely tied to their reasoning about univariate distribution as suggested by the statistics education literature (e.g., Cobb et al., 2003; Gravemeijer, 2000).

5.2. IS THE SEQUENCING OF QUANTITATIVE BIVARIATE DATA WITHIN A COURSE ASSOCIATED WITH DIFFERENCES IN THE PATTERN OF CHANGE IN STUDENTS' REASONING ABOUT QUANTITATIVE BIVARIATE DATA?

The sequencing of bivariate data within a course seemed not to be associated with changes in students' development of reasoning about bivariate data. There seemed to be no differences in either the linear or quadratic rates of change in covariational reasoning between the two instructional sequences. The fact that sequencing was not important in explaining patterns of development might not be surprising if, as stated in the last section, reasoning about bivariate data is just an artifact of reasoning about statistics in general.

Finding no differences in students' reasoning between the two sequences might suggest that the topic could be placed wherever the instructor or textbook authors decided. As a word of caution, however, even though the development in reasoning about bivariate data might not change as a result of the placement of this topic, student development of reasoning about other topics might be impacted. One of these topics could be inference. Although this wasn’t tested formally in this study, some anecdotal evidence, such as students’ complaints and discussion, suggests that students in the class where bivariate data was taught earlier seemed to be struggling with inference more than students in the other classes. It might also be that bivariate data is a topic that is more “digestible” than inference at the end of a semester.

Course sequencing has also received little attention in the statistics education literature. Although Chance and Rossman (2001) have speculated about the placement of a unit on bivariate data, there has been no research on optimal placement of this, or for that matter any other topic within an introductory statistics course. The literature on textbook usage has, however, suggested that the content and sequencing of textbooks could influence how effectively students will learn that content (e.g., Valverde, Bianchi, Wolfe, Schmidt, & Houang, 2002).

5.3. ARE CHANGES IN STUDENTS’ REASONING ABOUT THE FOUNDATIONAL CONCEPTS OF DISTRIBUTION ASSOCIATED WITH DIFFERENCES IN THE PATTERN OF CHANGE IN STUDENTS’ REASONING ABOUT QUANTITATIVE BIVARIATE DATA?

This study found that students who exhibit larger than average changes in their reasoning about univariate distribution also tend to have higher initial levels of reasoning about bivariate data. Furthermore, beyond initial status, this study has suggested that change in reasoning about univariate distribution is not associated with students’ development of reasoning about quantitative bivariate data. The findings from this research question are also somewhat novel. The research literature on students’ reasoning about bivariate data has been generally speculative. Although Cobb et al. (2003) and Gravemeijer (2000) have all suggested that students need to be able to reason about univariate distribution before they can reason about bivariate data, there have been no studies that have examined this hypothesis. Perhaps the pattern of change in reasoning exhibited by students in this study casts some doubt on these speculations. However, because most of the growth in reasoning about bivariate data seemed to occur during the instruction of univariate distribution, perhaps these two types of reasoning are inextricably connected.

5.4. LIMITATIONS TO THE STUDY

It is important to note the relatively small sample size (n = 113) in light of the use of multi-level modeling. This sample size may have resulted in less efficiency and power for the multilevel tests. This may have especially impacted the findings for the third research question (n = 98). As only 98 students had measurements on the fourth occasion, the sample was reduced due to the fact that not every student had a difference score (level-2 predictor) for this model.

A second limitation is the use of difference scores as a proxy for change in students’ reasoning about univariate distribution. The use of difference scores has long been a controversial issue, especially in regard to reliability (e.g., Cronbach & Furby, 1970; Willett, 1989b). The limited variability in scores may also have impacted the LMM coefficients.

Thirdly, teacher differences may also have affected the results. Inconsistencies due to these differences might have affected growth in such a way as to “cover up” differences due to one of the tested level-2 predictors. In larger studies this can be accounted for by using a three-level model where measurements are nested within students, which are nested within teachers. Thus, the variation can be further partitioned and accounted for. However, the small number of teachers (k = 2) did not allow this type of model to converge in this study.

Lastly, generalization may also be limited due to the type of introductory statistics students that were used in the study, namely social science students. However, they might be typical in terms of initial levels of reasoning and background for students enrolled in a non-calculus based first semester statistics course. Also, the study participants seemed typical in terms of exhibiting many of the same misconceptions that were identified in the literature (e.g., they have a tendency to form causal relationships based on a covariational analysis).

5.5. IMPLICATIONS FOR TEACHING

Despite the limitations described above concerning this study, the results suggest some practical implications for teachers of introductory statistics courses. For example, the results suggest that it is important to spend ample time developing students' reasoning about univariate distribution to provide a solid foundation for reasoning about quantitative bivariate data. This recommendation is consistent with recommendations in the statistics education literature that advocate that by covering fewer topics, a deeper conceptual understanding of the topics covered can be achieved, which translates into a greater understanding of topics that are covered at a later time (e.g., American Statistical Association, 2005b; Cobb, 1992; International Association for Statistical Education, 2005).

Although this study did not show a change in students' reasoning about quantitative bivariate data based on where the unit was placed in a course, anecdotal evidence did suggest that the sequence had an effect on students' reasoning about statistical inference. The smooth transition from normal distribution to sampling distribution to statistical inference may lead to a better understanding of statistical inference rather than inserting a unit on quantitative bivariate data between these topics.

It is also important to note that despite the use of a good research-based unit of instruction on bivariate data, students still had difficulty with many items on the bivariate reasoning assessment at the end of a 15-week course. These results confirm the finding in the research literature that ideas of covariation are often difficult for students to learn and may be counter-intuitive. Therefore, more attention should be paid to activities and instructional materials used to develop the important concepts that support covariational reasoning. Finally, the results suggest that if teachers emphasize the development of students' statistical reasoning throughout a course or curriculum, it may help students better prepare themselves to reason about quantitative bivariate data.

5.6. FUTURE RESEARCH

Additional research is suggested that examines growth of student reasoning within an introductory statistics course. One factor that continues to need investigation is the optimal placement of a unit on quantitative bivariate data and how this placement influences students' development in covariational reasoning, as well as the development of reasoning about other topics within an introductory statistics course such as inference. Questions about the best sequencing of curriculum within an introductory statistics course are important not only in how they impact students' learning and reasoning about statistics in general, but in how those sequences impact students' reasoning of sub-topics within a course.

Another suggested line of research is how foundational topics in an introductory statistics course influence students' development of reasoning about other topics. Although this study examined how changes in students' reasoning about univariate distribution influenced their reasoning about quantitative bivariate data, a different study might consider how students' reasoning about variation might influence reasoning about quantitative bivariate data or other statistical reasoning.

This study has employed a methodology that allows researchers to examine students' development of reasoning in an introductory statistics course in the context of a college classroom setting. It has also made an attempt at using randomization in classroom research. Future researchers may want to study predictors that may account for the level-2 variation.

Future research might also use a non-linear model and time-varying predictors to depict and explain student development. Non-linear models have been used to model change in student development (e.g., McArdle & Epstein, 1987). This might be more aligned with learning theories that model growth, retention and forgetting (e.g., Min, Vos, Kommers, & van Dijkum, 2000; Murre & Chessa, 2006; Wozniak, 1990). For instance, the use of the logistic curve to model population growth (introduced by Verhulst in 1845) was adapted by Pearl (1925) to model cognitive growth. Another example of non-linear growth to describe learning is the hyperbolic curve outlined by Thurston (1919). Time-varying predictors can be included in level-1 models to allow for direct effects between the predictor and outcome of interest over time.

In summary, the study of change in students' reasoning requires multiple measurements over time. The current methodologies used to study change (structural equation modeling [SEM] and multi-level modeling) require the same assessment to be used at each time point. This is generally not pedagogically acceptable to most college teachers given the time constraints that accompany collegiate courses. Even more complicated is the fact that to model a complex growth pattern requires more measurement occasions, especially during times that students are exhibiting the most change, such as near the beginning of the semester (Willett, 1989a; Willett, Singer, & Martin, 1998). This frequent testing could have a negative impact on student attitudes and cause early fatigue in study subjects. As the call for growth studies by policy makers and other interested parties increases, careful attention should be given to the methodologies and the practical problems faced by educators in their implementation.

ACKNOWLEDGEMENTS

The authors would like to thank Jeffrey Long, Michael Harwell, and Jeffrey Harring for their methodological advice.

REFERENCES

Adi, H., Karplus, R., Lawson, A., & Pulos, S. (1978). Intellectual development beyond elementary school VI: Correlational reasoning. *School Science & Mathematics, 78*(8), 675-683.

American Statistical Association. (2005a). *GAISE College Report.* [Online: http://www.amstat.org/education/gaise/GAISECollege.htm]

American Statistical Association. (2005b). *GAISE Endorsement.* [Online: http://www.amstat.org/education/gaise/ASAEndorse.htm]

Baterno, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151-169.

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' undertanding of statistical association in a computer based teaching environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference* (pp. 191-205). Voorburg, The Netherlands: International Statistical Institute.

Batanero, C., Godino, J. D., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier, & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 221-236). Stellenbosh, South Africa: University of Stellenbosh.

Bates, D., & Sarkar, D. (2005). *The lme4 package*. R package version 0.9975-13.

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

Ben-Zvi, D., & Garfield, J. (2004). (Eds.). *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishing.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition, 10,* 511-519.

Boyle, M. H., & Willms, J. D. (2001). Multilevel modeling of hierarchical data in developmental studies. *Journal of Child Psychology and Psychiatry, 42*(1), 141-162.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events. *Journal for Research in Mathematics Education, 33*(5), 352-378.

Chance, B., & Rossman, A. (2001). Sequencing topics in introductory statistics: A debate on what to teach when. *American Statistician, 55*(2), 140-144.

Cobb, G. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action*, MAA Notes No. 22, 3-43.

Cobb, P. (1998). Theorizing about mathematical conversations and learning from practice. *For the Learning of Mathematics, 18*(1), 46-48.

Cobb, P., Gravemeijer, K. P. E., Bowers, J., & Doorman, M. (1997). Statistical Minitools [applets and applications]. Nashville, TN and Utrecht, The Netherlands: Vanderbilt University, TN & Freudenthal Institute, Utrecht University.

Cobb, P., McClain, K., & Gravemeijer, K. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*(1), 1-78.

College Board (2003). *Advanced Placement Statistics course guide*. New York: Author.

Collins, L., Schafer, J., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6,* 330-351.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-333.

Cronbach, L. J., & Furby, L. (1970). How should we measure change – or should we? *Psychological Bulletin, 74*, 68-80.

Davis, F. B. (1964). Measurement of change. In F. B. Davis (Ed.), *Educational measurements and their interpretation* (pp. 234-252). Belmont, CA: Wadsworth.

Garfield, J., delMas, R., & Chance, B. (n.d.). *Assessment Resource Tools for Improving Statistical Thinking.* Retrieved April 8, 2006.
[Online: https://app.gen.umn.edu/artist/index.html]

Gravemeijer, K. P. E. (2000, April). *A rationale for an instructional sequence for analyzing one- and two-dimensional data sets*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis for stereotypic judgments. *Journal of Experimental Social Psychology, 12,* 392-407.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge and Kegan Paul.

International Association for Statistical Education. (2005). *SRTL-4 Report*. [Online: http://srtl.stat.auckland.ac.nz/]

Jennings, D., Amabile, T., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). Cambridge, England: Cambridge University Press.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748-769.

Kao, S. F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1363-1386.

Konold, C. (1999). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education, 3*(1). [Online: http://www.amstat.org/publications/jse/v3n1/konold.html]

Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics, 34*(2), 69-81.

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

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning (Learning, development & conceptual change)*. Cambridge, MA: MIT Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110-133.

McGahan, J. R., McDougal, B., Williamson, J. D., & Pryor, P. L. (2000). The equivalence of contingency structure for intuitive covariation judgments about height, weight, and body fat. *Journal of Psychology, 134,* 325-335.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54,* 33-61.

Min, R., Vos, H., Kommers, P., & van Dijkum, C. (2000). A concept model for learning: An attempt to define a proper relations scheme between instruction and learning and to establish the dynamics of learning in relation to motivation, intelligence and study-ability ('studeerbaarheid'). *Journal of Interactive Learning Research, 11*(3/4), 485-506.

Monk, S., & Nemirovsky, R. (1994). The case of Dan: Student construction of a functional situation through visual attributes. In E. Dubinsky, J. Kaput, & A. Schoenfeld (Eds.), *Research in collegiate mathematics education: Volume 1* (pp. 139-168). Providence, RI: American Mathematics Society.

Moritz, J. B. (2004). Reasoning about covariation. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227-256). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Murre, J. M. J., & Chessa, A. G. (2006). *A model of learning and forgetting II: The learning curve.* Unpublished manuscript.

Nemirovsky, R. (1996). A functional approach to algebra: Two issues that emerge. In N. Dedrarg, C. Kieran, & L. Lee (Eds.), *Approaches to algebra: Perspectives for research and teaching* (pp. 295-313). Boston: Kluwer Academic Publishers.

Pearl, R. (1925). *The biology of population growth*. New York: Knopf.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2005). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-66.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
[Online: http://www.R-project.org]

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications, Inc.

Ross, J. A., & Cousins, J. B. (1993). Patterns of student growth in reasoning about correlational problems. *Journal of Educational Psychology, 85*(1), 49-65.

Sánchez, F. T. (1999). *Significado de la correlación y regression para los estudiantes universitarios* [Meanings of correlation and regression for undergraduates]. Unpublished doctoral dissertation, University of Granada, Spain.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*(1), 102-119.

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development, 71*(1), 26-35.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.

Thompson, P. W. (1994). Images of rate and operational understanding of the fundamental theorem of calculus. *Educational Studies in Mathematics, 26,* 229-274.

Thurston, L. L. (1919). The learning curve equation. *Psychological Monographs, 26*(3), 1-51.

Truran, J. M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. In J. Garfield, & J. M. Truran (Eds.), *Research papers on stochastics educations from 1997* (pp. 205-212). Minneapolis, MN: University of Minnesota.

United States Department of Education. (2005). Raising achievement: A new path for No Child Left Behind. News Release.
[Online: http://www.ed.gov/policy/elsec/guid/raising/new-path-long.html]

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the Book. Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht: Kluwer Academic Publishers.

Verbeke, G, & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer Verlag.

Wavering, M. J. (1989). Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching, 26*(5), 373-379.

Willett, J. (1989a). Questions and answers in the measurement of change. *Review of Research in Education, 15,* 345-422.

Willett, J. B. (1989b). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement, 49,* 587–602.

Willet, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10,* 395-426.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235-269.

Wozniak, P. A. (1990). *Optimization of learning.* Unpublished master's thesis, Poznan University of Technology. Poznan, Poland.

Zieffler, A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Unpublished doctoral dissertation, University of Minnesota.

Zimmerman, C. (2005). *The development of scientific reasoning: What psychologists contribute to an understanding of elementary science learning.* Paper commissioned by the National Academies of Science (National Research Council's Board of Science Education, Consensus Study on Learning Science, Kindergarten through Eighth Grade).
[Online: www7.nationalacademies.org/bose/Corrine_Zimmerman_Final_Paper.pdf]

ANDREW S. ZIEFFLER
Educational Psychology
206 Burton Hall
178 Pillsbury Dr. SE
Minneapolis, MN 55455

APPENDIX

Bivariate Reasoning Assessment (BR) [ARTIST Quantitative Bivariate Data Scale]

1. Sam is interested in bird nest construction, and finds a correlation of 0.82 between the depth of a bird nest (in inches) and the width of the bird nest (in inches) at its widest point. Sue, a classmate of Sam, is also interested in looking at bird nest construction, and measures the same variables on the same bird nests that Sam does, except she does her measurements in centimeters, instead of inches. What should her correlation be?
   a. Sue's correlation should be 1, because it will match Sam's exactly.
   b. Sue's correlation would be 1.64(.82) = 1.3448, because you need to change the units from inches to centimeters and 1 inch = 1.64 centimeters.
   c. Sue's correlation would be about 0.82, the same as Sam's.

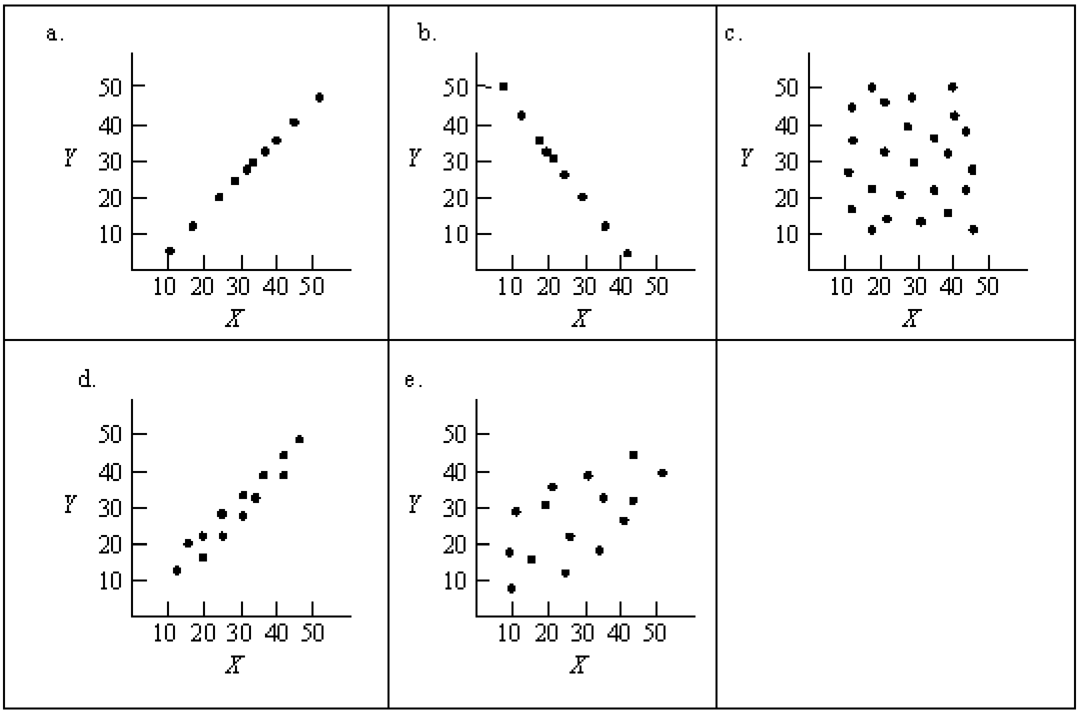
2. A student was studying the relationship between how much money students spend on food and on entertainment per week. Based on a sample size of 270, he calculated a correlation coefficient (*r*) of 0.013 for these two variables. Which of the following is an appropriate interpretation?
   a. This low correlation of 0.013 indicates there is no relationship.
   b. There is no linear relationship but there may be a nonlinear relationship.
   c. This correlation indicates there is some type of linear relationship.

3. A random sample of 25 Real Estate listings for houses in the Northeast section of a large city was selected from the city newspaper. A correlation coefficient of -0.80 was found between the age of a house and its list price. Which of the following statements is the best interpretation of this correlation?
   a. Older houses tend to cost more money than newer houses.
   b. Newer houses tend to cost more money than older houses.
   c. Older houses are worth more because they were built with higher quality materials and labor.
   d. New houses cost more because supplies and labor are more expensive today.

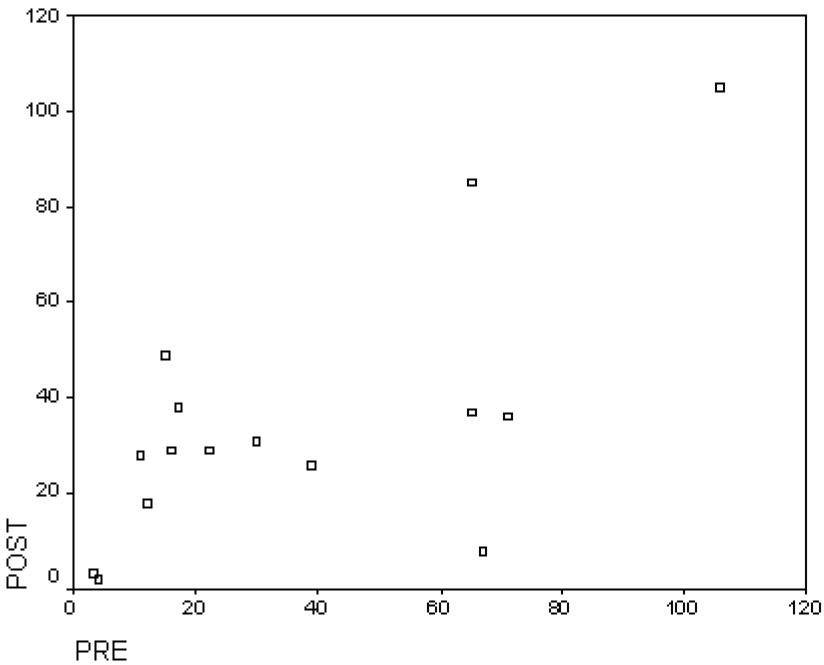
For items 4 and 5, select the scatterplot that shows:

4. A correlation of about 0.60.
   a. a
   b. b
   c. c
   d. d
   e. e

5. The strongest relationship between the *X* and *Y* variables.
   a. a
   b. b
   c. a and b
   d. a and d
   e. a, b, and d



Dr. Jones gave students in her class a pretest about statistical concepts. After teaching about hypotheses tests, she then gave them a posttest about statistical concepts. Dr. Jones is interested in determining if there is a relationship between pretest and posttest scores, so she constructed the following scatterplot and calculated the correlation coefficient.



6. Locate the point that shows a pretest score of 107. This point, which represents John's scores, is actually incorrect. If John’s scores are removed from the data set, how would the correlation coefficient be affected?
   a. The value of the correlation would decrease.
   b. The value of the correlation would increase.
   c. The value of the correlation would stay the same.

7. It turns out that John’s pretest score was actually 5, and his posttest score was 100. If this correction is made to the data file and a new correlation coefficient is calculated, how would you expect this correlation to compare to the original correlation?
   a. The absolute value of the new correlation would be smaller than the absolute value of the original correlation.
   b. The absolute value of the new correlation would be larger than the absolute value of the original correlation.
   c. The absolute value of the new correlation would be the same as the absolute value of the original correlation.
   d. It is impossible to predict how the correlation would change.

8. A statistics instructor wants to use the number of hours studied to predict exam scores in his class. He wants to use a linear regression model. Data from previous years shows that the average number of hours studying for a final exam in statistics is 8.5 hours, with a standard deviation of 1.5 hours, and the average exam score is 75, with a standard deviation of 15. The correlation is 0.76. Should the instructor use linear regression to predict exam scores from hours studied?
   a. Yes, there is a high correlation, so it is alright to use linear regression.
   b. Yes, because linear regression is the statistical method used to make predictions when you have bivariate quantitative data.
   c. Linear regression could be appropriate if the scatterplot shows a clear linear relationship.
   d. No, because there is no way to prove that more hours of study causes higher exam scores.

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

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

ABSTRACT

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

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

1. TERMINOLOGY

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

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

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

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

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

2. OVERVIEW OF THE PROBLEM AND ITS IMPORTANCE

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

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

3. BACKGROUND

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

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

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

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

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

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

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

4. RESEARCH QUESTIONS

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

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

5. METHOD

5.1. TASKS

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

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

5.2. SAMPLE

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

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

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

5.3. ANALYSIS

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

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

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

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

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

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

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

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

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

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

6. RESULTS

6.1. BOOKS

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

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

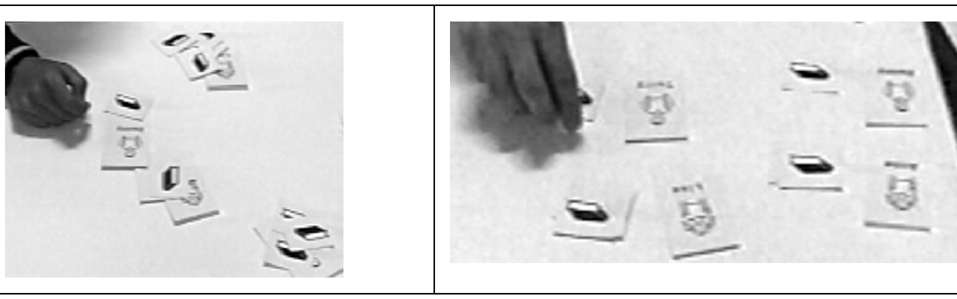


Figure 1. No visible (or very little) variation shown in representation (Level 0)

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

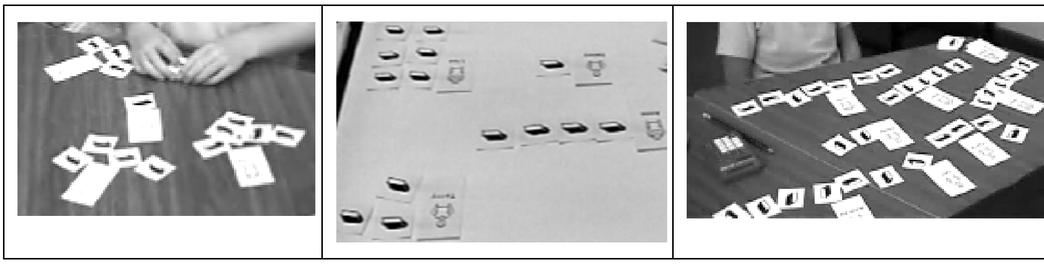


Figure 2. Variation clear but unstructured (Level 1)

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

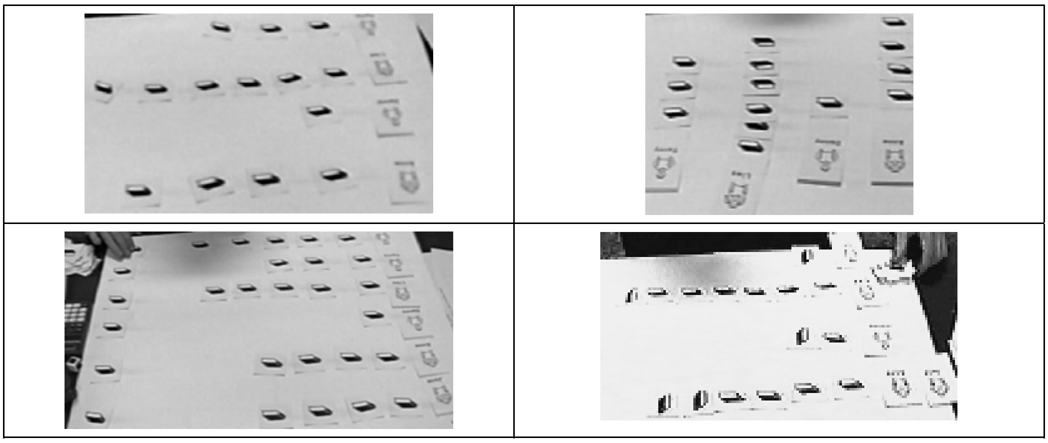


Figure 3. Variation shown in rectangular format (Level 2)

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

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

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

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

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

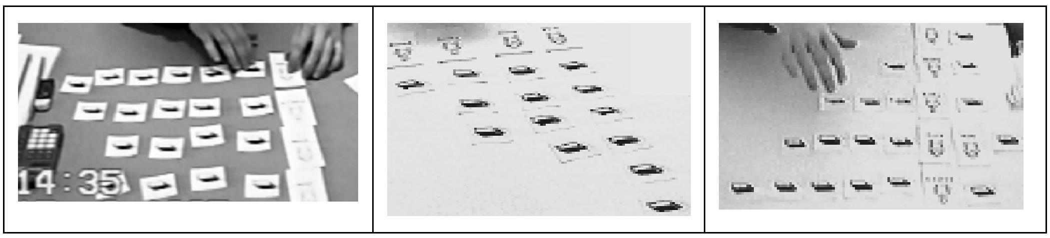


Figure 4. Ordered variation in the display (Level 3)

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

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

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

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

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

6.2. WEATHER

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

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

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

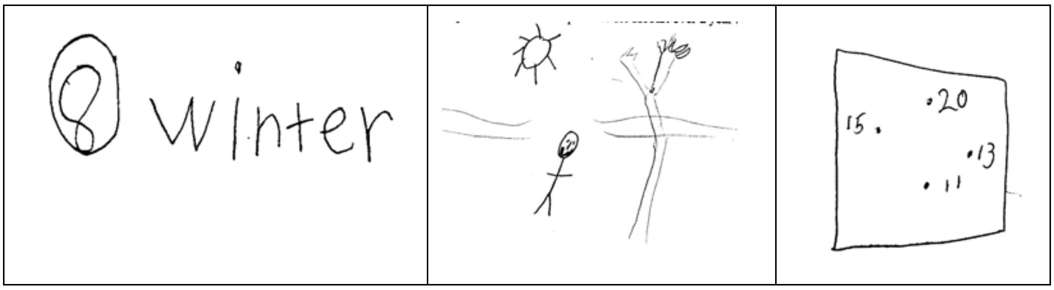
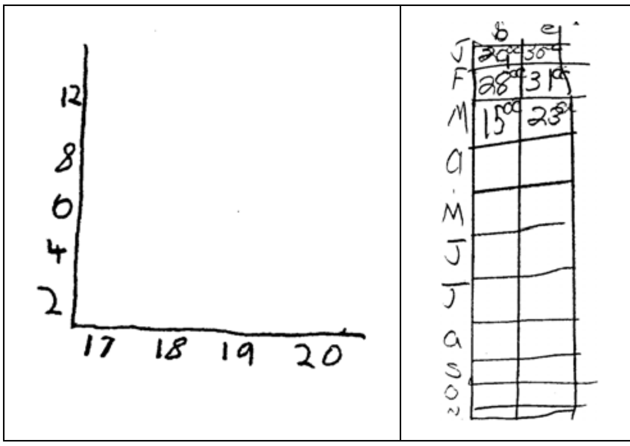


Figure 5. Static weather pictures (Level 0)

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

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

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



S: That's start, oh well beginning and the end [writes on sheet]

I: So you have got the beginning of the month and the end of the month. Is that what you are doing?

S: Yes. Start off with … so I would do like start with the highest … [writes]

I: So what are you putting in there? You are doing a table?

S: Yes. Like 29°C is the highest the start of January and the end would be about 30.

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

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

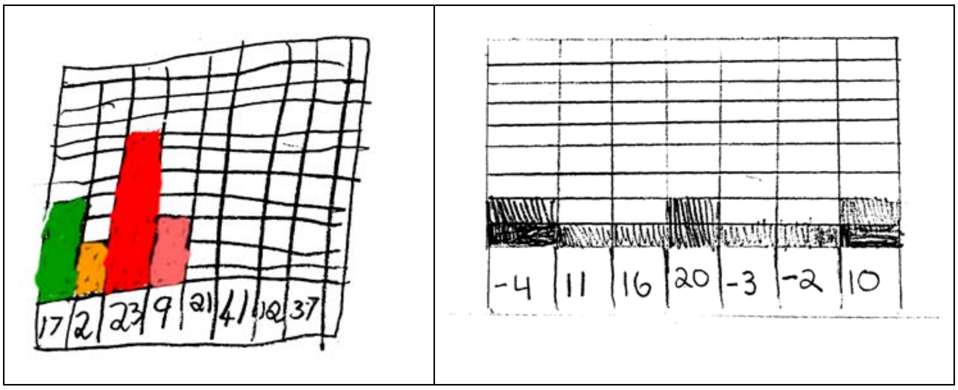


Figure 8. Frequency graphs (Level 2)

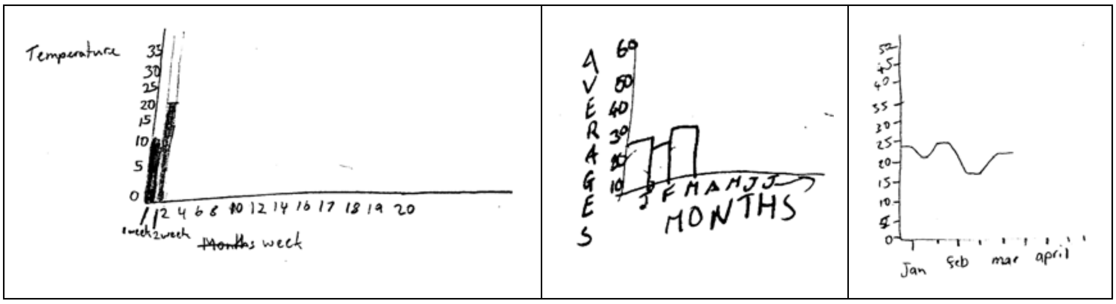
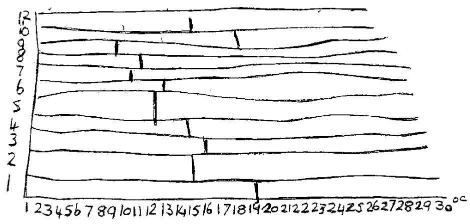


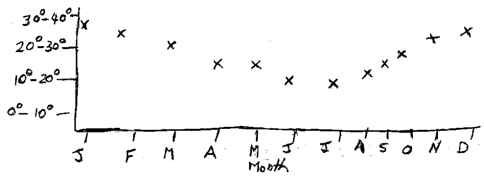
Figure 9. Beginnings of temperature graphs (Level 2)

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

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

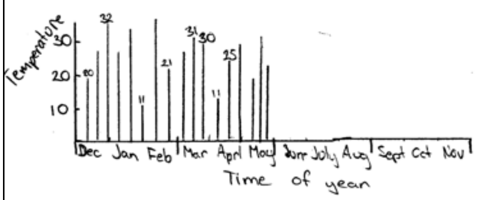


I: So if those are the averages for the months – now can you tell me why they are going up and down a bit?
S: Colder and hotter.
I: OK – so where's the coldest part of the year?
S: In the middle [points to graph].
I: Sort of in the middle there, somewhere.
I: And it gets warmer toward the end.
S: Yes,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

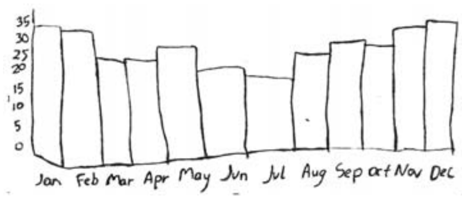


S: I think I will have a little line graph.
I: Can you explain it to me.
S: Well January is usually the hottest month and so the average what the temperature is.
I: The average of the temperature and that [graph] represents that for the month.
S: Yes each cross.
I: And why does it go down?
S: Because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

Figure 10. Graphs with seasonal variation (Level 3)



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



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

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

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

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

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

6.3. LOLLIES

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

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

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

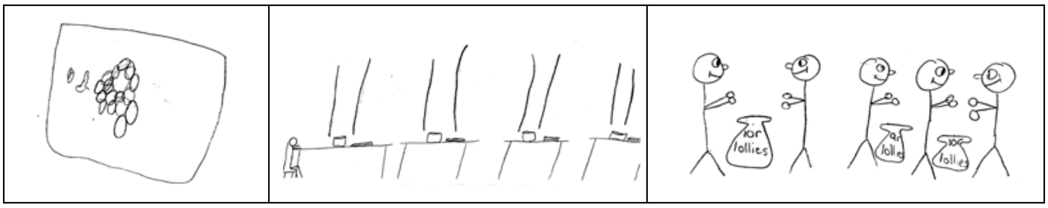


Figure 12. Sketches of contexts for drawing lollies (Level 0)

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

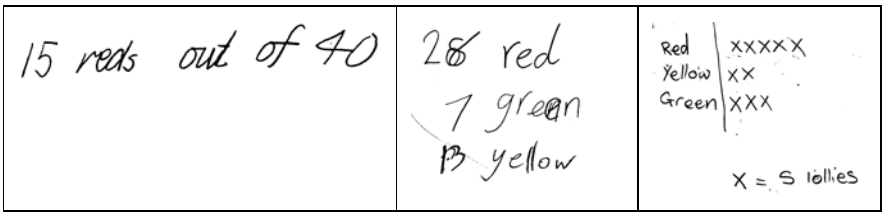


Figure 13. Individual outcomes from draws (Level 0)

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

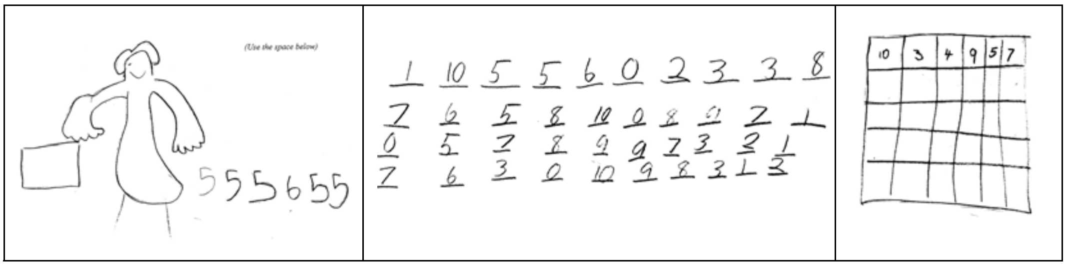


Figure 14. Multiple outcomes from draws (Level 1)

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

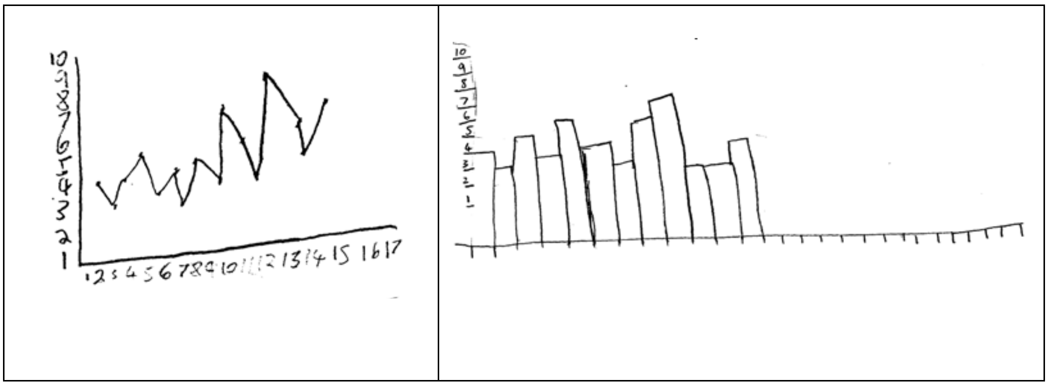
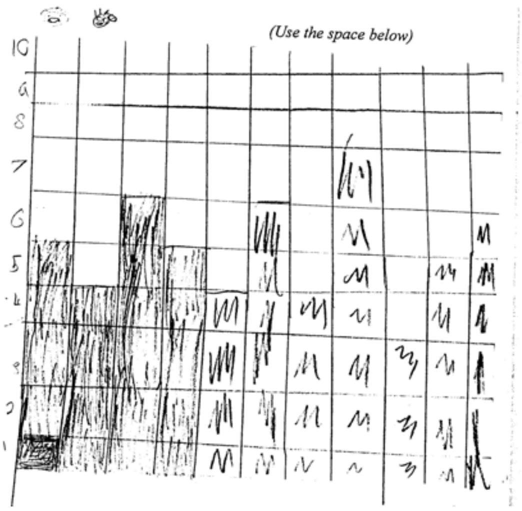


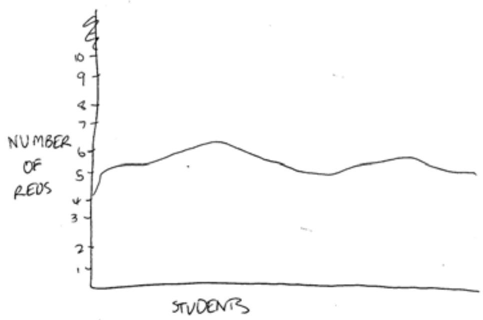
Figure 15. Time-series-like graphs (Level 2)

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

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



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



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

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

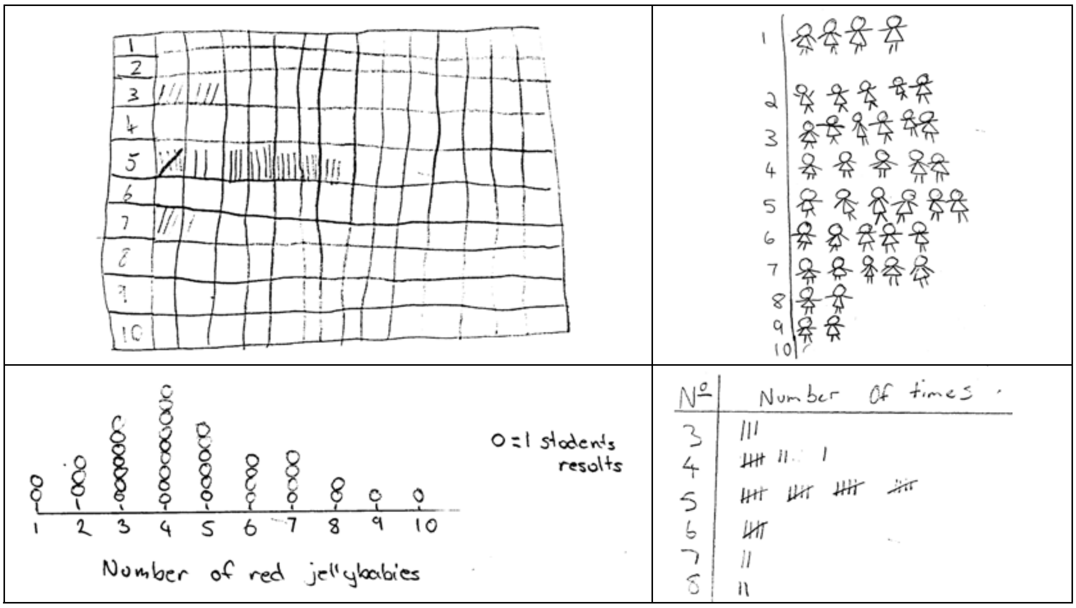
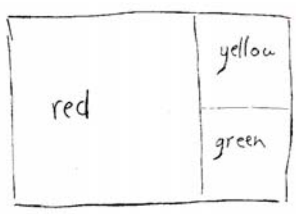


Figure 17. Frequency distribution graphs (Level 4)

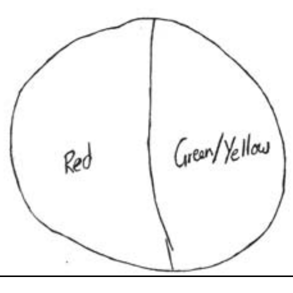
Table 6 shows the levels of response for each grade for the initial representations drawn for the LOLLIES protocol. Young students had some difficulty with appreciating the task in its original form. Only two graphs produced by Grade 7 students appeared to represent expectation to the exclusion of variation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

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

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



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



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

Figure 18. Graphs representing expectation only (Level 0)

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

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

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

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

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

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

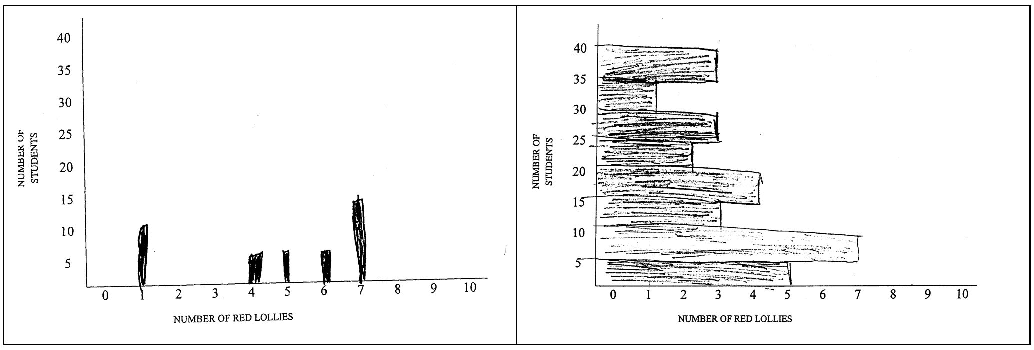


Figure 19. Frequency graphs with axes (Level 2)

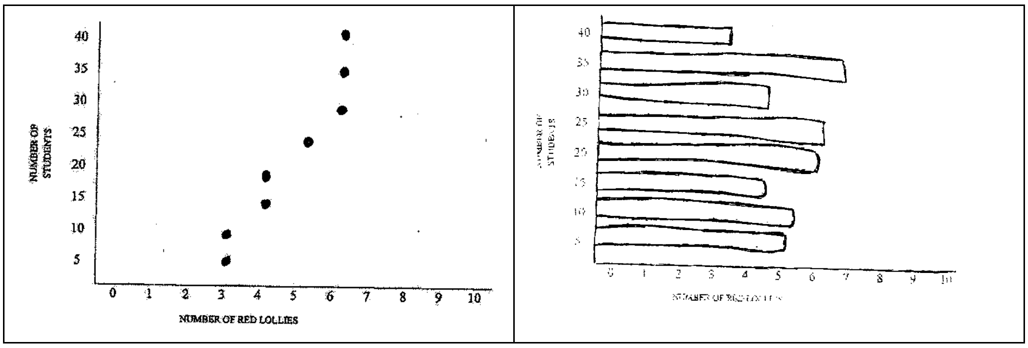


Figure 20. Frequency graphs with axes and a central expectation (Level 3)

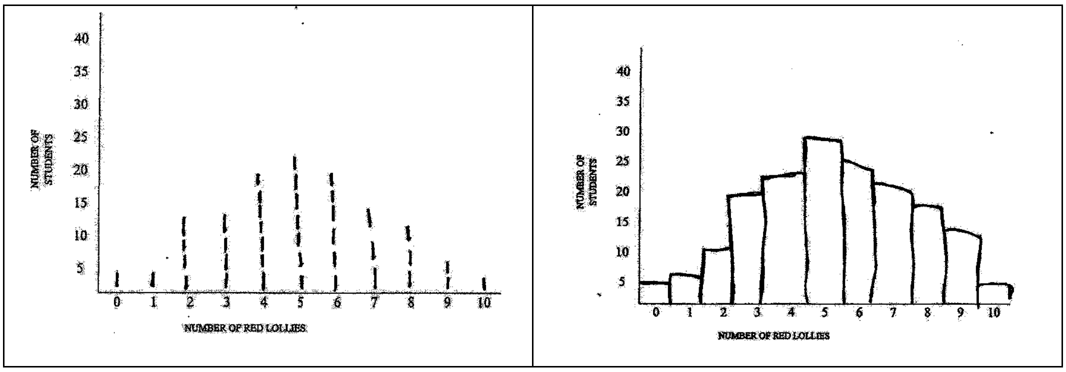


Figure 21. Frequency distributions with axes (Level 4)

6.4. SUMMARY

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

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

7. DISCUSSION

7.1. VARIATION, EXPECTATION, AND DISTRIBUTION

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

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

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

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

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

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

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

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

7.2. LIMITATIONS

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

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

7.3. IMPLICATIONS FOR RESEARCH AND THE CLASSROOM

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

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

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

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

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

ACKNOWLEDGMENTS

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

REFERENCES

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

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

Biggs, J. B. (1992). Modes of learning, forms of knowing, and ways of schooling. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 31-51). London: Routledge.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning*, *9*, 83-130.

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

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

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

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

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

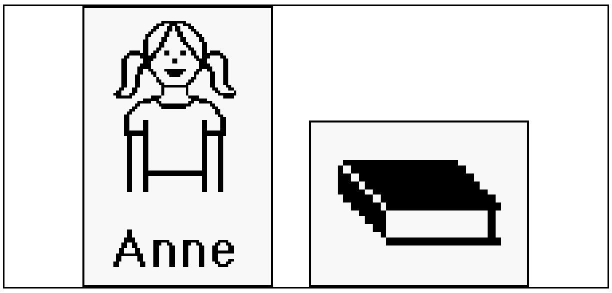
Watson, J. M., & Kelly, B. A. (2006). Expectation versus variation: Students' decision making in a sampling environment. *Canadian Journal of Science, Mathematics and Technology Education*, *6*, 145-166.

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

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

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

APPENDIX A: INTERVIEW PROTOCOL FOR THE BOOKS TASK



Sample of cards used as materials for representation.

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

APPENDIX B: INTERVIEW PROTOCOL FOR THE WEATHER TASK

1. Some students watched the news every night for a year, and recorded the daily maximum temperature in Hobart. They found that the average maximum temperature in Hobart was 17°C.

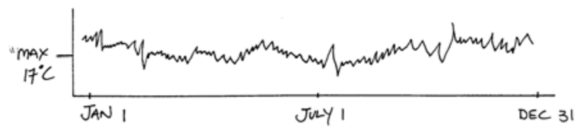
   a) What does this tell us about the temperature in Hobart?
   b) Do you think all the days had a maximum of 17°C? - Why or why not?
   c) (What do you think the maximum temperature in Hobart might be for 6 different days in the year?)* ______, ______, ______, ______, ______, ______
   d) Why did you make these choices?

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

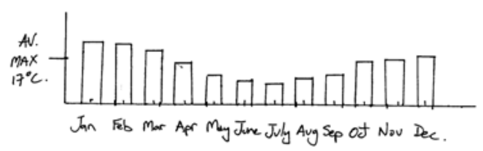
2. How would you describe the temperature for Hobart over a year in a graph?

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

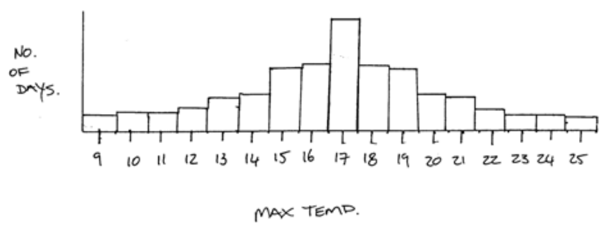
a)



b)



c)



APPENDIX C: INTERVIEW PROTOCOL FOR THE LOLLIES TASK

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

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

2. Suppose six of you do this experiment.

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

3. Look at these possibilities that some students have written down for the numbers they thought likely.

(a) 5,9,7,6,8,7 (b) 3,7,5,8,5,4 (c) 5,5,5,5,5,5 (d) 2,3,4,3,4,4
(e) 7,7,7,7,7,7 (f) 3,0,9,2,8,5 (g) 10,10,10,10,10,10

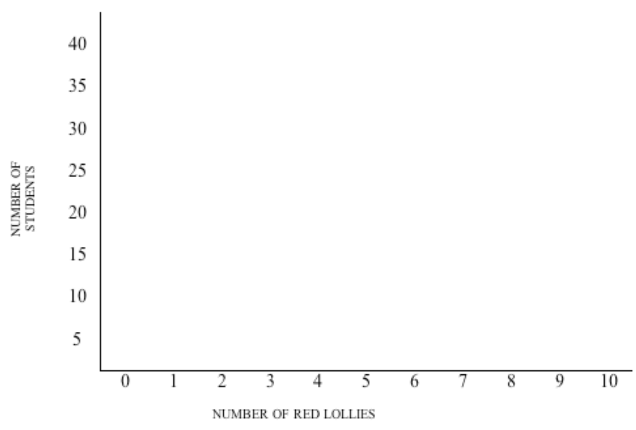
Which one of these lists do you think best describes what might happen? Why do you think this?

4. Suppose that 6 students did the experiment. What do you think the numbers will most likely go from and to?

From __________ (lowest) to __________ (highest) number of reds. Why do you think this?
Now try it for yourself: ______, ______, ______, ______, ______, ______
Given the results, do you want to change any of your previous answers?

5. Suppose that 40 students pulled out 10 lollies from the container, wrote down the number of reds, put them back, mixed them up.

a) Can you show what the number of reds look like in this case? (Use the blank space below)
b) Now use the graph below to show what the number of reds might look like for the 40 students.



FACTORS INFLUENCING THE DEVELOPMENT OF MIDDLE SCHOOL STUDENTS' INTEREST IN STATISTICAL LITERACY

COLIN CARMICHAEL
University of Tasmania
colin.carmichael@utas.edu.au

ROSEMARY CALLINGHAM
University of Tasmania
Rosemary.Callingham@utas.edu.au

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

IAN HAY
University of Tasmania
Ian.Hay@utas.edu.au

ABSTRACT

This paper reviews factors that contribute to the development of middle school students' interest in statistical literacy and its motivational influence on learning. To date very little research has specifically examined the influence of positive affect such as interest on learning in the middle-school statistics context. Two bodies of associated research are available: interest research in a mathematics education context and attitudinal research in a tertiary statistics context. A content analysis of this literature suggests that interest development in middle school statistics will be the result of a complex interplay of classroom influences and individual factors such as: students' knowledge of statistics, their enjoyment of statistics and their perceptions of competency in relation to the learning of statistics.

Keywords: *Literature review; Attitudes; Statistics education*

1. INTRODUCTION

There is currently a shortage of mathematics and statistics graduates in Australia. In their review of mathematical sciences research, the Australian Academy of Science (2006), reported that in 2003 only 0.4% of Australian graduates majored in either mathematics or statistics, which compared unfavorably with an OECD average of 1%. Further, the Australian Bureau of Statistics has reported difficulty in obtaining suitably qualified statistics graduates (Trewin, 2005). Such shortages have their origins in the secondary school context, where the number of students enrolled in higher level mathematics courses is showing a declining trend (McPhan, Morony, Pegg, Cooksey, & Lynch, 2008). In addition to this, McPhan et al. reported that students' lack of interest and

Statistics Education Research Journal, 8(1), 62-81, http://www.stat.auckland.ac.nz/serj

liking for mathematics during their middle school education was one of five factors that contributed to this decline, the other factors being their previous achievement in mathematics, their mathematics self-concept, and their perceptions regarding the usefulness and difficulty of mathematics. This paper seeks to address the issues associated with national skill shortages in statistics through a review of factors that could contribute to middle school students' interest in statistical literacy.

Statistically literate adults should be able to interpret and critically evaluate messages that contain statistical elements (Gal, 2003). For example, they should be able to recognize bias as a possible source of error in media reports of survey data. Models have been conceptualized that describe the development of statistical literacy in learners (Gal, 2002; Watson, 2006). In his model of statistical literacy, Gal (2002) identified several key knowledge bases that were essential for the development of statistical literacy. He concluded, however, that such knowledge was of minor consequence if a person was unwilling to apply this knowledge. Gal's model of statistical literacy, therefore, included a dispositional component: A statistically literate adult should possess a readiness to criticize messages that contain statistical elements. Such a disposition, he argued, emerges when a person has a belief in their capabilities and an interest and willingness to engage in statistical thinking. In her model of statistical literacy development, Watson (2006) also included dispositional elements, under the broader category of task motivation. These dispositions included skepticism, imagination, and curiosity. It is argued that the dispositions identified by Gal and Watson are themselves developed as a result of positive emotional learning experiences with data. Such experiences formally commence in school and reach necessary levels of sophistication towards the end of middle school (grade 9), where it is argued dispositions are less likely to change. The development of positive affect in the middle school years is therefore a prerequisite to the acquisition of dispositions necessary for statistical literacy. It is also argued that the development of positive affect in the middle school has a considerable bearing on the subject choices that students make in senior secondary and tertiary contexts, thus influencing later skill shortages.

Students in a middle school context (grades 6 to 9) are typically in early adolescence, which appears to be a critical stage in their affective development. In the mathematics education context, for example, evidence points to a decline in levels of positive affect as a student progresses through school (Fredricks & Eccles, 2002) with such levels reaching a minimum in year 10 (Watt, 2004). The correlation between student attitudes towards mathematics and their achievement in mathematics, however, appears to be strongest for students in grades 8 to 10 (Ma & Kishor, 1997; Ma & Xu, 2004). The influence of affect on learning appears to be more pronounced for this group of students. Such findings are supported by reported physiological changes to the brain that occur during adolescence (Wigfield, Byrnes, & Eccles, 2006), changes that result in the greater likelihood of affective activity.

The preceding discussion highlights the need for research into the influence of affect on learning in a middle school statistics context. The next section argues that interest, a commonly used term, is a positive affect that is essential for human psychological development. Interest is therefore of particular relevance to the current context.

1.1. INTEREST AS A POSITIVE AFFECT

The Macquarie Dictionary (Delbridge et al., 1987, p. 910) defines interest as "the feeling of one whose attention or curiosity is particularly engaged by something." Therefore interest can be regarded as a positive affect that is specifically directed towards some object, termed the "object of interest." Deci (1992) argued that interest is fundamental in the development of a person's concept of self. Moreover recent research suggests that interest is necessary for psychological growth, with absence of interest in adolescents being linked with psychological disorders such as depression (Hunter & Csikszentmihalyi, 2003).

In the psychological literature the term "affect" is assumed to be "a broad rubric that refers to all things emotional" (Rosenberg, 1998, p. 247). Affective elements vary on a hierarchical continuum from emotional states, which are typically short in duration but characterized by high levels of arousal, to traits which are stable predispositions to respond in certain ways. As an affect, interest is regarded as having both trait and state characteristics (Schiefele, 1991). At the trait level "individual interest" is described as a "person's relatively enduring predisposition to reengage particular content over time" (Hidi & Renninger, 2006, p. 113). Interest at the state level is more transitory and is typified by a positive emotion akin to excitement. In a state of interest a learner may become so absorbed in the object of interest that they lose sense of time: They experience "flow" (Csikszentmihalyi, 1991). The state of interest can be induced by aspects of the environment and in such instances is termed "situational interest." Alternatively the state of interest can be induced from the individual's predisposition to engage and in such instances is termed "actualized interest." It is believed that individual interest can emerge from situational interest. Thus the requisite dispositions for statistical literacy may emerge from students experiencing the state of interest during their learning.

In a learning context, students' interest can explain some of their motivation to engage in learning activities. Such interest-driven motivation is termed "intrinsic motivation" and is the doing of an activity for its inherent value. The concept of intrinsic motivation features prominently in Self Determination Theory (Deci & Ryan, 1985) which posits that individuals are motivated to behave in seemingly unrewarded ways in order to meet basic psychological needs, including the need to be self-determined. Students who are motivated intrinsically, that is out of an interest in the subject, are known to produce qualitatively superior learning outcomes to their extrinsically motivated peers. For example, Schiefele (1991) reported that student interest was positively associated with deeper levels of cognitive processing, the use of self-regulatory learning strategies and students' ratings on the quality of their learning experience. Further, there is significant correlation between student interest and both academic achievement (Schiefele, Krapp, & Winteler, 1992), and choice of subjects (Köller, Baumert, & Schnabel, 2001). Given the importance of interest development in adolescence and its association with learning, a study of the development of affect in students should include the development of their interest.

1.2. THE DEVELOPMENT OF INTEREST

Current theories of interest development suggest that students' interest in statistical literacy will emerge as they gain expertise in the area. As an example, the Model of Domain Learning (Alexander, 2003) posits that students' interest in a given domain will increase as they gain knowledge in that domain. Further, the model suggests that in the early stages of knowledge acquisition, levels of individual interest will be quite low and learners will rely on situational interest for motivation. As learners move through the domain towards expertise, they will increasingly rely on their individual interest for motivation, with situational interest becoming of less importance. This relationship between levels of situational and individual interest during the development of domain expertise implies that the latter will emerge from the former. Indeed, Hidi and Renninger

(2006) argued that individual interest will emerge from situational interest. In particular they argued that if situational interest can be maintained then it will develop into an emerging individual interest and then finally into an enduring individual interest. The mechanism by which this transformation occurs is explained by Silvia (2001). He proposed that interest occurs when an individual resolves the cognitive conflict that is created when he or she interacts with the object of interest. More specifically, Silvia (2001) argued that during the person-object interaction, incoming stimuli are assembled with current personal information on the basis of a number of "collative" variables that are associated with the learner's response to the stimuli. These collative variables include novelty, uncertainty, and complexity. During this interaction, the learner will fail to engage in any significant way with stimuli that are considered routine (low levels of novelty). Similarly the learner will fail to engage when the stimuli are too unknown or frightening (high levels of novelty). For optimal levels of these variables a state of curiosity will be evoked that is characterised by high levels of arousal and positive emotions, including interest. In this state the learner will be motivated to resolve the conflict created by the particular collating variable. If this conflict cannot be resolved quickly, the learner will be motivated to persist with the object, even returning to it at later times. Such persistence with the object may uncover further stimuli that in turn create a conflict in need of resolution. In such a way it is hypothesized that both knowledge and interest in the object will develop, with people losing interest in simple objects and pursuing those with more complex associated knowledge.

The emergence of interest in statistical literacy may also occur as the result of the individual's unique set of "interests," those interest objects with which he or she is particularly interested. Krapp (2002) identified a number of ways that interests could develop. For example, he argued that a new interest could emerge as a result of the increased differentiation of one aspect of an existing interest. In the statistical literacy context, students with a particular interest in sports may gain an interest in statistical literacy through the analysis of sports related statistical reports. The highly idiosyncratic patterns of interests that students bring to the classroom, however, pose a number of challenges to the educator. Foremost among these is being able to identify and cater for such interests. It is argued that educators have the most influence in the creation of situational interest and the development of knowledge and skills that relate to statistical literacy.

1.3. RESEARCH QUESTION

In the preceding section, interest is identified as a positive affect that is necessary for psychological development and intrinsically motivated behaviors. In the statistics education area, research into interest per se is limited; related attitudinal research, however, may provide information on factors that lead to the development of interest in statistics. A review of the theoretical interest-based literature reveals that both knowledge and prior interests should contribute to the development of students' interest in statistical literacy. The preceding discussion, however, has not specifically examined interest development in the statistical literacy context. Accordingly, this paper seeks to answer the following question: what are the factors documented in the literature that influence the positive development of middle school students' interest in statistical literacy?

2. METHODOLOGY

The literature review was conducted in three phases, commencing with a search on the specific question and then generalising the search to encompass interest development in secondary mathematics contexts, and then to the development of positive affect in the tertiary statistics education context. Searches in all phases commenced with databases of academic journals and abstracts including: *A+Education*, *Emerald*, *ERIC*, *Expanded Academic*, *JSTOR Education*, *Proquest*, *PsycINFO*, *SAGE*, *SpringerLink* and *Wiley Interscience*. In addition *Google Scholar* was found to be a particularly useful search engine. Secondary searches of others' bibliographies and searches using citation indexes were also conducted in each phase.

The initial search specifically addressed the research question using the keywords interest and statistics (or statistical) in the article title. In addition to the databases discussed above, a search was conducted on specific statistics education journal archives including: *Statistics Education Research Journal*, *Journal of Statistics Education*, and *Teaching Statistics*. Further, an archive of statistics education dissertations retained by the International Association for Statistical Education was also searched. Only one study located in the search specifically examined the concept of interest as it relates to the learning of statistics in a school context. In the second phase the search was expanded to include interest development in both secondary school mathematics and tertiary statistics contexts. In the third phase the search was expanded to include attitudinal-related research in both school and tertiary contexts. The search keywords in this phase included attitude and statistics (or statistical) in the article title.

After retrieving relevant research articles, a content analysis (Krippendorff, 1980) identified common and conflicting outcomes that were related to the research question. The results of this review and analysis are discussed in the next section.

3. RESULTS

The results from the first two phases of the search are presented in Section 3.1. This section presents research, which is in the main part situated in a secondary mathematics context. The results from the last phase of the search, which is tertiary statistics related, are presented in Section 3.2. The relevance of both mathematics education and tertiary attitudinal research to the current context is then addressed in the discussion section of this paper.

3.1. MATHEMATICS EDUCATION RELATED RESEARCH FINDINGS

The final search from the first two phases resulted in 38 hits with publication dates that ranged from 1976 to 2008 Thirteen of these could not be readily accessed and in most cases were published prior to 1995. Of the remaining 25 articles, six were discarded as the term interest was used generically to describe a feeling of well-being that was neither defined nor measured. Three articles were also discarded as they had included interest items in larger mathematics attitudinal scales, but had not reported specific interest outcomes. A further two validation studies were not used in this review as they reported technical aspects of interest scales rather than empirical evidence that could contribute to the research question. The majority of the remaining studies (9 of 14) were in a secondary-school mathematics context and specific details of these studies are shown in Appendix A.

The content analysis of this literature revealed a number of common outcomes but also differences in the way the interest construct was operationalised. These commonalities and differences are reported below.

Common outcomes from the mathematics education literature A number of common factors were identified from the mathematics education literature as having a positive influence on interest development. These can be broadly classified into those that are situational and those that are individual (see Table 1). Situational factors include pedagogical strategies and aspects of the learning environment. Individual factors include the prior experiences and beliefs of the learners.

Pedagogical practices, including the types of learning experiences that students encounter and the classroom management strategies used by teachers, have been shown to promote interest. Several studies provide supporting evidence: Heinze, Reiss, and Rudolph (2005), Kunter, Baumert, and Köller (2007), Mitchell (1997), Mitchell and Gilson (1997), and Sciutto (1995). As an example, Mitchell (1997) noted that learning activities that involve puzzles, computers, and group work will catch students' interest. Similarly, teaching strategies that promote student involvement and which students find meaningful will hold students' interest. Mitchell was able to provide some evidence to suggest that the individual interest of students in environments high in situational interest will increase in both a mathematics (Mitchell & Gilson, 1997) and statistics (Mitchell, 1997) secondary-school context. It is arguable whether changes in interest reported after a period of only 14 weeks, the period used in these studies, reflect changes in individual interest. Nevertheless, pedagogical practices undoubtedly influence the situational interest in the classroom, which it is argued will ultimately develop into individual interest.

Table 1. Common study outcomes from the mathematics education literature

Situational factors	• Pedagogical practices can promote interest in mathematics (Bikner-Ahsbahs, 2004; Heinze et al., 2005; Mitchell, 1997, Mitchell & Gilson, 1997; Sciutto, 1995; Trautwein, Ludtke, Köller, Marsh, & Baumert, 2006). • The social climate of the classroom can promote interest development (Bikner-Ahsbahs, 2004). • The classroom management strategies used by teachers (Kunter et al., 2007) and the views of significant others (Fox, 1982) can promote interest in mathematics.
Individual factors	• Interest in mathematics is associated with students' prior knowledge and their competency-based beliefs (Fox, 1982; Köller et al., 2001; Lawless & Kulikowich, 2006; Lopez, Brown, Lent, & Gore, 1997; Marsh, Trautwein, Ludtke, Köller, & Baumert, 2005; Preckel, Goetz, Pekrun, & Kleine, 2008; Trautwein et al., 2006); and also their age (Köller et al., 2001).

The social climate of the learning environment also plays an important role in developing interest. Bikner-Ahsbahs (2004) argued that a type of interest, termed "situated collective interest," will emerge in a group situation where one-by-one students become involved in an activity and come to value the activity. Through observations of children she was able to provide some evidence to support this theory. In relation to the social environment, Fox (1982) found the views of significant others, including parents and teachers, influence student ratings of "career interest" (the type of career they would be interested in pursuing), but indirectly through their ratings of confidence and the utility of mathematics. Also in a secondary mathematics context, Kunter et al. (2007) were able to demonstrate that students' interest was influenced by their evaluation of their teacher's classroom management strategies. In particular, interest was predicted by student perceptions on the extent to which teachers clearly outlined class rules and the extent to which teachers monitored student progress.

At an individual level, several studies demonstrated an association between prior knowledge and interest (see Table 1). The direction of this relationship has also been explored. Köller et al. (2001) reported that achievement in early adolescence (grade 7) predicts interest in mid adolescence (grade 10). However, achievement in mid adolescence does not predict subsequent levels of interest. They concluded that age is a factor in interest development and argued that younger adolescents are more sensitive to achievement feedback than older adolescents who presumably have more stable interests. In addition, Köller et al. reported that interest in grade 7 does not predict achievement in grade 10 but interest in grade 10 does predict achievement in grade 12. The strength of the association between interest and prior knowledge is known to be influenced by the structure of the knowledge domain in question. Lawless and Kulikowich (2006), for example, reported a stronger association for statistics than for psychology, and argued that the former is a more structured knowledge domain.

Several studies also demonstrated a link between students' conceptions of their competency and their level of interest. Lopez et al. (1997) provided evidence to suggest that students' self-efficacy beliefs predict their interest in mathematics. Marsh et al. (2005) and Trautwein et al. (2006) both demonstrated the link between students' academic self-concept and interest in mathematics, with Trautwein et al. asserting that self-concept is a strong predictor of interest, which almost entirely mediates the influence of achievement and tracking (the assigned level of class). Moreover, Trautwein et al. argued that this relationship is influenced by the frame of reference used by students to judge their competency: High achievement students who are in a group of even higher achieving students report low levels of interest in mathematics while low achieving students in a group of even lower achieving students report high levels of interest.

Differences in the operationalisation of the interest construct In the mathematics education context, differences were evident regarding the operationalisation of the interest construct. The German studies (Köller et al., 2001; Kunter et al., 2007; Marsh et al., 2005; Trautwein et al., 2006) regarded interest as having both a value and an emotion component, with the former including the importance of the task and the latter the enjoyment of the task. The concept of importance may assess the usefulness or utility of the task, an extrinsic motivator. Students, who report mathematics as important, may do so because they perceive it to be necessary for future job prospects. Such importance may not reflect interest, although some evidence suggests that it may predict interest (Fox, 1982). Other studies operationalised interest through asking students to indicate their level of interest in a given task (Lawless & Kulikowich, 2006; Lopez et al., 1997; Sciutto, 1995). Of concern, is whether students' assessment of interest is similar to their assessment of enjoyment? Some authors suggest the two are quite distinct emotions (Izard, 1984; Reeve, 1989; Silvia, 2001).

3.2. TERTIARY STATISTICS RELATED RESEARCH FINDINGS

Expanding the search to include attitudinal research in a statistics education context resulted in more hits. In this phase there were 51 hits with publication dates ranging from 1980 to 2008. Of these hits, nine could not be readily accessed. Two theoretical discussion studies that highlighted the shortfalls of using attitudinal instruments were not used in the analysis. In addition, a further five validation studies were also discarded from the current review. These studies reported technical aspects of proposed statistics attitude scales rather than empirical evidence that could support the research question. The resulting 35 papers are described in Appendix B, which reports that the majority of these papers (31 of 35) were empirical studies and a majority (32 of 35) were based in a tertiary education context. This prevalence of tertiary based studies is of concern to the current review. As discussed earlier, the influence of affect on learning during adolescence, which is the context for this paper, appears to be more pronounced than for other stages of life. Nevertheless, the findings from tertiary based research may inform the current study.

The content analysis of this literature revealed a number of common outcomes (see Table 2), some conflicting outcomes (see Table 3) and differences in the way studies operationalised attitudes toward statistics. These are reported below.

Common outcomes from the tertiary statistics education literature At a situational level, the social climate of the classroom was shown to influence the value that students place on statistics (Cobb & Hodge, 2002). Moreover, Mvududu (2003) found that aspects of a constructivist classroom, in particular personal relevance and student negotiation, are associated with positive attitudes towards the field of statistics.

At an individual level, competency-based beliefs are known to be associated with attitudes towards statistics (Finney & Schraw, 2003; Sorge & Schau, 2002). The nature of this relationship was explored by Tempelaar, Schim Van Der Loeff, and Gijselaers (2007) who reported a strong correlation (r = 0.8) between the cognitive competence and affect subscales of the "Survey of Attitudes toward Statistics (SATS)" instrument (Schau, Stevens, Dauphinee, & Ann, 1995). This result suggests that a strong relationship exists between competency based beliefs and positive affect in the statistics education context: Students enjoy doing those tasks that they believe can be undertaken successfully.

Table 2. Common study outcomes from the tertiary statistics education literature

Situational factors	The social climate of the classroom can promote positive attitudes towards statistics (Cobb & Hodge, 2002; Mvududu, 2003).
Individual factors	Attitudes towards statistics are associated with students' prior knowledge of both mathematics and statistics (Carmona, 2004; Estrada, Batanero, Fortuny, & Diaz, 2005; Perney & Ravid, 1990) and their attributional (Budé et al., 2007) and competency based beliefs (Finney & Schraw, 2003; Sorge & Schau, 2002).

Conflicting outcomes from the tertiary statistics education literature Many studies sought to establish that innovative pedagogical strategies and the use of technology enhanced learning environments would be associated with positive student attitudes. As is summarised in Table 3, not all were successful.

Innovative pedagogical strategies were shown to promote positive attitudes towards statistics and presumably interest in statistics. These included the use of video clips that demonstrate real-life applications of statistics (Allredge, Johnson, & Sanchez, 2006), the embedding of statistical activities in stories (D'Andrea & Waters, 2002), and the use of real-life and person based scenarios (Leong, 2006). There is some evidence, however, to suggest that pedagogical practices that aim to improve attitudes towards statistics in fact promote attitudes to the particular class (or teacher) where the learning occurs. D'Andrea & Waters (2002), for example, found that attitude improvements in their study were directed towards the statistics course and not towards the field of statistics.

Table 3. Conflicting study outcomes from the tertiary statistics education literature

- Innovative pedagogical strategies were associated with positive student attitudes in some studies (Allredge et al., 2006; D'Andrea & Waters, 2002; Leong, 2006) but not in others (Carnell, 2008; Faghihi & Rakow, 1995).
- Technology enhanced learning environments were associated with positive student attitudes in some studies (Meletiou-Mavrotheris, Lee, & Fouladi, 2007; Schou, 2007; Suanpang, Petocz, & Kalceff, 2004) but not in others (Alajaaski, 2006; Cybinski & Selvanathan, 2005; Elmore, Lewis, & Bay, 1993; Gratz, Volpe, & Kind, 1993).

Not all studies were successful in establishing a positive link between innovative pedagogical strategies and changes in student attitudes. For example, in a recent quasi-experiment, Carnell (2008) prescribed one class of undergraduate statistics students with a data-collection project that they themselves designed and implemented. A second class, used as a control, did not have this option. Surprisingly, in a pre-test/post-test situation, there were no positive changes in attitudes for either group. In fact both groups showed a significant loss in interest during the course. It is likely that other, unreported, course conditions had a far greater impact on student attitudes than the teaching innovation in question.

Several studies demonstrated that the use of technology enhanced learning environments such as those employing integrated technology (Meletiou-Mavrotheris et al., 2007) and those using an online environment (Schou, 2007; Suanpang et al., 2004), could enhance attitudes towards statistics. Not all studies, however, were able to demonstrate this association (Alajaaski, 2006; Cybinski & Selvanathan, 2005; Elmore et al., 1993; Gratz et al., 1993). Unfortunately these latter studies provided scant details on how the technology was used to promote learning, concentrating instead on the statistical analysis of results. For example, Cybinski and Selvanathan (2005) described a quasi-experiment involving two groups. The first group attended a lecture, a tutorial, and a computer laboratory session. The second opted to receive their materials from an online environment but also attended the computer laboratory session. Students in the second group reported significantly lower levels of affect towards statistics (p-value = 0.09). Most students in this group (77%), however, had not completed high school mathematics. Further, the authors failed to report whether students using the online environment were able to interact with each other and with their teachers, that is, whether they received social support. It is likely, therefore, that these different outcomes are the result of how the technology was used rather than whether it was used.

Differences in the operationalisation of attitudes toward statistics In the statistics education context the focus of attitudinal instruments varied considerably. In his study relating attitudes towards assessment in statistics, Onwuegbuzie (2000) used an instrument that focussed entirely on statistical anxiety. On the other hand, many of the studies (10 of the 30 empirical studies) used SATS, which assesses affect and a broad range of beliefs. The choice of scales also varied considerably. Some evaluative studies developed their own small attitudinal instruments (Alajaaski, 2006; Garfield & Ahlgren, 1994) whereas many others used previously validated scales. Several studies also used a mix of mathematics education and statistics education attitudinal instruments (Lalonde & Gardner, 1993; Perney & Ravid, 1990; Tremblay, Gardner, & Heipel, 2000), not with the intention of exploring differences between the two disciplines but rather under the assumption that student attitudes to the two are equivalent.

4. DISCUSSION

The results of this research review raise two broad issues for statistics educators. The first concerns the possible divergence of mathematics education research and statistics research. The second concerns the relevance of tertiary statistics attitudinal research to the current context in high schools. In addition to these broad issues, the noted differences in the operationalisation of the interest construct suggest the need for further exploration of the relationship between the learning emotions of interest and enjoyment, and student beliefs regarding the value of statistics. In this section an overview of those factors that influence students' interest, as detailed in the research reviewed, is provided. The discussion then examines the relevance of the secondary mathematics and tertiary statistics contexts to the research question. It then explores the relationship between interest, enjoyment, and student beliefs; and, concludes with a brief account of how interest development may occur in the current context.

4.1. FACTORS THAT INFLUENCE MIDDLE SCHOOL STUDENTS' INTEREST IN STATISTICAL LITERACY

Self-determination Theory (Deci & Ryan, 1985) provides a unifying framework for interest (or attitudinal) studies such as those described in this paper. Deci (1992) argued that a person will experience interest when he or she encounters novel activities in a context that allows for the satisfaction of his or her basic psychological needs; namely,, competence, autonomy, and social-relatedness. In a middle school context, a student's need for autonomy (being able to choose what he or she does) and social-relatedness can be met if aspects of the classroom environment are conducive. The content analysis identified the social climate as a factor that positively influenced students' attitudes. Mvududu (2003), for example, reported a statistically significant association between student negotiation and positive attitudes towards statistics (r = 0.25). A student's need for competence in statistical literacy, however, will be met if he or she possesses the necessary individual factors; namely, a sufficient knowledge of statistical literacy and positive competency-based beliefs regarding his or her ability to acquire statistical literacy. The content analysis identified prior knowledge, competency-based beliefs, and prior interests as individual factors that contributed to interest and/or positive attitudes.

Overarching the meeting of basic psychological needs is the requirement that students encounter novel activities. The content analysis identified pedagogical factors that contributed to both interest and attitudes. The extent to which these strategies utilized novel activities, however, is unclear. In his study of interest development, Mitchell (1997) utilized learning activities that were meaningful to students and which encouraged student involvement. Arguably true involvement comes from collative sources that include novelty. In the statistics-education context, Allredge et al. (2006), D'Andrea & Waters (2002), and Leong (2006) provided students with familiar contexts and reported positive changes in attitudes. The use or otherwise of novel activities is perhaps the point at which interest development as opposed to attitude development, differ.

This review has established a significant gap in the literature. Of the studies cited in this review, several examined interest but in a mathematics education context, and a large proportion examined positive affect, but in tertiary statistics education context. Only one study (Mitchell, 1997) examined the concept of interest in a secondary school context, although an evaluative attitudinal study in that context by Garfield and Ahlgren (1994) used items that could be regarded as interest-based. As is discussed in the next sections, it is unclear whether mathematics education findings are relevant to the statistics education context. Similarly, it is also unclear whether findings associated with the affective development of adults are relevant to a middle school context. The evidence, as is discussed, suggests that adults have more stable attitudes than adolescents.

4.2. THE POSSIBLE DIVERGENCE OF MATHEMATICS AND STATISTICS RELATED RESEARCH

Statistics educators, such as Moore (1988) strenuously argue that statistics is a distinct methodological discipline from mathematics. Yet in most cases statistics is taught in mathematics classrooms by mathematics teachers, so that students themselves may see few differences between statistics and mathematics. Indeed many studies in this review reported prior mathematics experiences as having a significant influence on students' attitudes towards statistics. Despite the distinctive nature of statistics as a field of study it is likely that mathematics-related research into affect and/or interest will apply equally to statistics education contexts. There is a trend, however, to move the teaching of statistics away from a computational formula-driven approach to a practical data-oriented approach. Most current Australian secondary-school syllabi encourage a hands-on approach to the teaching of statistical concepts. Further, professional development of mathematics teachers in the teaching of statistics is occurring. The *StatSmart* project, described in Callingham and Watson (2007), is an innovative Australian-based professional development program that aims to develop teaching skills in statistical concepts. It is in this atmosphere, where the learning of statistical concepts is becoming less computationally driven, and where students are able to play with data, that possibilities exist for a divergence of mathematics and statistics related affective research.

4.3. THE RELEVANCE OF TERTIARY BASED ATTITUDINAL RESEARCH TO THE CURRENT CONTEXT

A substantial proportion of the statistics education studies in this review is based in a tertiary context. Of concern is the extent to which these studies are applicable to students in a middle school context. In their study of secondary school students, Köller et al. (2001) concluded that junior secondary students may be more sensitive to achievement feedback than their older peers. Such a conclusion is supported by related research into the emotional development of children. In a longitudinal study of 220 students across grades 5 to 12, Larson, Moneta, Richards, and Wilson (2002) reported that junior students show a greater variability in their emotions than older students: Emotional stability increases during adolescence. It is argued that the adults studying statistics in a tertiary context will have more stable emotions than adolescents in a middle school context; indeed they should have developed mechanisms for controlling changes in emotions. The situational and individual factors that the literature identified as contributing to positive attitudes in adults should apply to young adolescents. Young adolescents, however, will be prone to greater variation in emotions than adults: They will become more excited with interesting activities, but increasingly bored with mundane activities. It is therefore beholden on educators to harness the extreme positive emotions that younger adolescents may experience in their learning. That is, to utilise and develop their interest.

4.4. INTEREST, ENJOYMENT, AND STUDENT BELIEFS

The reported differences in the operationalisation of the interest construct suggest the need to distinguish between the emotions of enjoyment and interest. Reeve (1989) provided evidence to demonstrate that interest is derived from collative sources and enjoyment from the feelings of satisfaction that accompany task competency. He argued that both emotions were necessary for intrinsically motivated learning. Students enjoy success and are likely to reengage with tasks with which they perceive likely success. With no interest, however, they are likely to tire of the task. In a learning context, an item such as ‘I enjoy statistics’ is a poor operationalisation of the interest construct with students reflecting upon either their success in statistics, their interest in statistics, or both.

Given that interest has a value dimension it is necessary to differentiate between a student’s belief regarding the personal value of an interest object and his or her belief regarding its utility. As noted by Tempelaar et al. (2007), the value sub-scale of SATS is broad in that it seeks to assess both the utility of statistics and its personal value. In a middle-school context students may be extrinsically motivated to learn statistics because it is perceived as being useful rather than because of a personal valuing of the subject. The learning of statistics for its intrinsic value and hence the development of interest, may come later. Such an argument is supported by Ryan and Deci (2000) who proposed that extrinsically motivated behaviour may develop, over time and through stages, to intrinsically motivated behaviour and therefore to the emergence of interest.

4.5. THE DEVELOPMENT OF INTEREST IN STATISTICAL LITERACY

Watson and Callingham (2003) argued that the development of statistical literacy will occur in stages that reflect an increasing interaction with an increasingly unfamiliar context. The Model of Domain Learning (Alexander, 2003) predicts that during this development, students’ individual interest in statistical literacy will also increase, although Alexander cautions that students leaving school will likely reach competence rather than expertise in any domain of knowledge. It is argued that the development of statistical literacy and interest in statistical literacy will be interwoven and will occur in a time-frame that extends beyond the secondary school years. During this development, individual interest in statistical literacy will provide the motivation for re-engagement, while features of the context and/or task will provide the collative motivation for interest development.

The content analysis in this review highlighted a number of factors that may contribute to interest growth in the short term. But it is the short-term with which this review is concerned. Creating and nurturing interest in statistical literacy during the important developmental years of adolescence will rely on pedagogical practices that trigger situational interest and support adolescents’ needs for competence, autonomy, and social-relatedness. Long term development of such interest, however, will be a complex interplay of students’ knowledge of statistical literacy, their beliefs regarding their competency in this field, and their enjoyment of learning.

5. IMPLICATIONS

The literature review reported in this paper identifies a significant gap in the literature as it relates to interest in statistics and indeed statistical literacy. Related research in the mathematics education context indicates that interest in mathematics is predictive of later achievement for mid adolescent (grade 10) students but not for younger middle school students (Köller et al., 2001). This result suggests that interests stabilise towards the end of middle school and reinforces the need for further research into interest development during adolescence. Further, there appears to be a difference in the strength of this relationship according to the knowledge domain in question (Lawless & Kulikowich, 2006). Given that a difference exists between students' perceptions of mathematics and statistics, this result reinforces the need for study into students' interest in statistics as opposed to mathematics.

Given the need for research in the middle school statistics context and the differences in the way that interest has been operationalised, there arises a need for the development of a suitable instrument for undertaking large scale quantitative analyses. In the tertiary statistics context, the multi-faceted instrument SATS is extensively used, however it was developed specifically for undergraduate and graduate students who complete specific statistics courses. Consequently many items within SATS are not appropriate in a middle school context.

The research review also identifies factors that have been shown to promote positive affect and interest in students. Some studies identified situational factors, such as teaching strategies, others identified individual factors such as prior knowledge, but only one explored the relative influence of both of these types of factors. Kunter et al. (2007) reported that only 10% of the variation in students' interest in mathematics is explained by class membership, which suggests that individual factors may account for a much larger proportion of the variance. Future research into the development of students' interest should consider both situational and individual factors and the relationship between them.

ACKNOWLEDGEMENTS

The research was funded by Australian Research Council Linkage grant number LP0669106, with support from the Australian Bureau of Statistics. An early version of this paper was presented at the 31st annual conference of the Mathematics Education Group of Australasia (Brisbane, Australia, July, 2008).

REFERENCES

*References marked with an asterisk indicate studies included in the appendices.

Alajaaski, J. (2006). How does web technology affect students' attitudes towards the discipline and study of mathematics/statistics? *International Journal of Mathematical Education in Science and Technology*, *37*(1), 71-79.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, *32*(8), 10-14.

Allredge, J. R., Johnson, H. D., & Sanchez, J. J. (2006). Does viewing video of statistics in action affect student attitudes? In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/C335.pdf]

Australian Academy of Science. (2006). Mathematics and statistics: Critical skills for Australia's future. *The National Strategic Review of Mathematical Sciences Research in Australia.*
[Online: http://www.review.ms.unimelb.edu.au/Report.html]

*Biajone, J. (2006). Promoting positive attitudes towards statistics in pedagogy students through project work. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.

[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/C408.pdf]

Bikner-Ahsbahs, A. (2004, July). *Interest-dense situations and their mathematical valences*. Paper presented at the Tenth International Congress in Mathematics Education, Topic Study Group, Copenhagen.
[Online: http://www.icme-organisers.dk/tsg24/Documents/BiknerAshbahs.doc]

Budé, L., van De Wiel, M. W. J., Imbos, T., Candel, M. J. J. M., Broers, N. J., & Berger, M. P. F. (2007). Students' achievements in a statistics course in relation to motivational aspects and study behavior. *Statistics Education Research Journal*, *6*(1), 5-21.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(1)_Bude.pdf]

Callingham, R., & Watson, J. M. (2007). Overcoming research design issues using Rasch measurement: The *StatSmart* project. In P. Jeffery (Ed.), *Proceedings of the Australian Association for Research in Education Annual Conference*, Freemantle, Melbourne: The Australian Association for Research in Education.
[Online: http://www.aare.edu.au/07pap/cal07042.pdf]

Carmona, J. (2004, July). *Mathematical background and attitudes toward statistics in a sample of undergraduate students*. Paper presented at the 10th International Conference on Mathematics Education, Copenhagen.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/11/Carmona.doc]

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes towards statistics. *Journal of Statistics Education, 16*(1).
[Online: www.amstat.org/publications/jse/v16n1/carnell.html]

Cobb, P., & Hodge, L. (2002). Learning, identity, and statistical data analysis. In B. Phillips (Ed.), *Developing a statistically literate society: Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/2e1_cobb.pdf]

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York: Harper Collins.

Cybinski, P., & Selvanathan, S. (2005). Learning experience and learning effectiveness in undergraduate statistics: Modeling performance in traditional and flexible learning environments. *Decision Sciences Journal of Innovative Education*, *3*(2), 251-271.

D'Andrea, L., & Waters, C. (2002). Teaching statistics using short stories: Reducing anxiety and changing attitudes. In B. Phillips (Ed.), *Developing a statistically literate society: Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/8a2_dand.pdf]

Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43-70). Mahwah, NJ: Lawrence Erlbaum Associates.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

Delbridge, A., Bernard, J. R. L., Blair, D., & Ramson, W. S. (Eds.). (1987). *The Macquarie dictionary*. Sydney: The Macquarie Library.

Elmore, P. B., Lewis, E. L., & Bay, M. L. G. (1993). *Statistics achievement: A function of attitudes and related experiences*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED 360324)

Estrada, A., Batanero, C., Fortuny, J. M., & Diaz, C. (2005, February). *A structural study of future teachers' attitudes towards statistics*. Paper presented at the Research in Mathematics Education Conference, Sant Feliu de Guixols, Spain. [Online: http://cerme4.crm.es/Papers%20definitius/5/EstradaBatanero.pdf]

Faghihi, F., & Rakow, E. A. (1995, November). *The relationship of instructional methods with student responses to the survey of attitudes towards statistics*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MI. (ERIC Document Reproduction Service No. ED 392358)

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics course. *Contemporary Educational Psychology, 28*(2), 161-186.

Fox, L. H. (1982). *The study of social processes that inhibit or enhance the development of competence and interest in mathematics among highly able young women*. John Hopkins Institute of Education, Washington, DC. (ERIC Document Reproduction Service No. ED 222037)

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 516-533.

*Fullerton, J. A., & Umphrey, D. (2001, August). *An analysis of attitudes toward statistics: Gender differences among advertising majors*. Paper presented at the Annual Meeting of the Association for Education in Journalism and Mass Communication, Washington, DC. (ERIC Document Reproduction Service No. ED 456479)

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1-51.

Gal, I. (2003). Teaching for statistical literacy and services of statistics agencies. *The American Statistician*, *57*(2), 80-84.

Garfield, J. B., & Ahlgren, A. (1994). Student reactions to learning about probability and statistics: Evaluating the Quantitative Literacy Project. *School Science and Mathematics*, *94*(2), 89-95.

*Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal, 3*(1), 4-59. [Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(1)_gordon.pdf]

Gratz, Z. S., Volpe, G. D., & Kind, B. M. (1993, March). *Attitudes and achievement in introductory psychological statistics classes: Traditional versus computer-supported instruction*. Paper presented at the 7th annual conference on Undergraduate Teaching of Psychology, Ellenville, NY. (ERIC Document Reproduction Service No. ED 365405)

*Gunnarsson, C. L. (2001). Development and assessment of students: Attitudes and achievement in a business statistics course taught online. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 3*(2).

Heinze, A., Reiss, K., & Rudolph, F. (2005). Mathematics achievement and interest in mathematics from a differential perspective. *ZDM, 37*(3), 212-219.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.

Hunter, J. P., & Csikszentmihalyi, M. (2003). The positive psychology of interested adolescents. *Journal of Youth and Adolescence, 32*(1), 27-35.

Izard, C. E. (1984). Emotion-cognition relationships and human development. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 17-37). New York: Cambridge University Press.

*Kaplan, J. J. (2006). *Factors in statistics learning: Developing a dispositional attribution model to describe differences in the development of statistical proficiency*. Unpublished doctoral dissertation, University of Texas, Austin.
[Online: www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php]

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education, 32*(5), 448-470.

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383-409.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: SAGE Publications.

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, *17*(5), 494-509.

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology. *Canadian Journal of Behavioural Science*, *25*(1), 108-125.

Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, *73*(4), 1151-1165.

Lawless, K. A., & Kulikowich, J. M. (2006). Domain knowledge and individual interest: The effects of academic level and specialization in statistics and psychology. *Contemporary Educational Psychology, 31*(1), 30-43.

Leong, J. (2006). *High school students' attitudes and beliefs regarding statistics in a service-learning-based statistics course*. Unpublished PhD Thesis, Georgia State University, Georgia.

Lopez, F. G., Brown, S. D., Lent, R. W., & Gore, P. A. (1997). Role of social-cognitive expectations in high school students' mathematics-related interest and performance. *Journal of Counseling Psychology*, *44*(1), 44-52.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude towards mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*(1), 26-47.

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education, 110*(3), 256-280.

Marsh, H. W., Trautwein, U., Ludtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397-416.

McPhan, G., Morony, W., Pegg, J., Cooksey, R., & Lynch, T. (2008). *Maths? Why Not?* (Report): Australian Department of Education, Employment and Workplace Relations (DEEWR).
[Online: http://www.dest.gov.au/sectors/school_education/publications_resources/]

Meletiou-Mavrotheris, M., Lee, C., & Fouladi, R. T. (2007). Introductory statistics, college student attitudes and knowledge - A qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology*, *38*(1), 65-83.

*Mills, J. D. (2004). Students' attitudes towards statistics: Implications for the future. *College Student Journal, 38*(3), 349-361.

Mitchell, M. (1997, March). *Situational interest in the statistics classroom*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 413345)

Mitchell, M., & Gilson, J. (1997, March). *Interest and anxiety in mathematics*. Paper presented at the annual meeting of the American Research Association, Chicago. (ERIC Document Reproduction Service No. ED 412116)

Moore, D. S. (1988). Should mathematicians teach statistics? *The College Mathematics Journal*, *19*(1), 3-7.

Mvududu, N. (2003). A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course. *Journal of Statistics Education*, *11*(3).
[Online: http://www.amstat.org/publications/jse/v11n3/mvududu.html]

Onwuegbuzie, A. J. (2000). Attitudes toward statistics assessments. *Assessment and Evaluation in Higher Education*, *25*(4), 321-339.

Perney, J., & Ravid, R. (1990, April). *The relationship between attitudes towards statistics, math self-concept, test anxiety and graduate students' achievement in an introductory statistics course*. Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 318607)

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students: Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *The Gifted Child Quarterly*, *52*(2), 146-150.

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, *13*(2), 83-103.

*Renninger, K. A., Ewen, L., & Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. *Learning and Instruction, 12*(4), 467-491.

Rosenberg, E. L. (1998). Levels of analysis and organization of affect. *Review of General Psychology*, *2*(3), 247-270.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67.

Schau, C., Stevens, J., Dauphinee, T. L., & Ann, D. V. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, *55*(5), 868-875.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3 & 4), 299-323.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183-212). Mahwah, NJ: Lawrence Erlbaum Associates.

Schou, S. B. (2007). A study of student attitudes and performance in an online introductory business statistics class. *Electronic Journal for the Integration of Technology in Education*, *6*.
[Online: http://ejite.isu.edu/Volume6/Schou.pdf]

*Schultz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences*, *10*(4), 291-308.

Sciutto, M. (1995). Student-centered methods for decreasing anxiety and increasing interest level in undergraduate statistics courses. *Journal of Instructional Psychology*, *22*(3), 277-280.

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, *5*(3), 270-290.

Sorge, C., & Schau, C. (2002, April). *Impact of engineering students' attitudes on achievement in statistics: A structural model*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
[Online: http://evaluationandstatistics.com/AERA2002.pdf]

Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Educational Technology and Society*, *7*(3), 9-20.

Tempelaar, D. T., Schim Van Der Loeff, S., & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities, and course performance. *Statistics Education Research Journal*, *6*(2), 78-102.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_Tempelaar.pdf]

Trautwein, U., Ludtke, O., Köller, O., Marsh, H. W., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788-806.

Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science*, *32*(1), 40-48.

Trewin, D. (2005). Improving statistical literacy: The respective roles of schools and the National Statistical Offices. In M. Coupland, J. Anderson & T. Spencer (Eds.), *Twentieth biennial conference of the Australian Association of Mathematics Teachers* (pp. 11-19). Adelaide: Australian Association of Mathematics Teachers.

*Vaisanen, P., Rautopuro, J., & Sakari, Y. (2004, September). *Modelling the impact of teacher education students' real and perceived maths ability and motivational-affective factors on their success in elementary statistics*. Paper presented at the European Conference on Education Research, University of Crete.
[Online: http://www.leeds.ac.uk/educol/documents/00003646.htm]

*Vanhoof, S. (2006). Attitudes toward statistics and their relationship with short and long term exam results. *Journal of Statistics Education*, *14*(3).
[Online: http://www.amstat.org/publications/jse/v14n3/vanhoof.html]

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

Watson, J. M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3-46.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ2(2)_Watson_Callingham.pdf]

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th through 11th grade Australian students. *Child Development*, *75*(5), 1556-1574.

Wigfield, A., Byrnes, J. P., & Eccles, J. S. (2006). Development during early and middle adolescence. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 87-113). Mahwah, NJ: Lawrence Erlbaum Associates.

Zanakis, S. H., & Valenzi, E. R. (1997). Student anxiety and attitudes in business statistics. *Journal of Education for Business, 73*(1), 10-17.

COLIN CARMICHAEL
2273 Gore Highway
Southbrook, Q. 4363
Australia

APPENDIX A: INTEREST BASED STUDIES IN A MATHEMATICS OR STATISTICS EDUCATION, SCHOOL BASED CONTEXT

Article	Description	Context
Bikner-Ahsbahs (2004)	Observational	Sixth grade mathematics class (Germany)
Fox (1982)	Empirical (*n* = 125)	Junior-secondary (year 7) mathematics (US)
Heinze et al. (2005)	Empirical (*n* = 500)	Junior-secondary (years 7 and 8) mathematics (Germany)
Köller et al. (2001)	Empirical (*n* = 602)	Secondary (years 7 to 12) mathematics (Germany)
Kunter et al. (2007)	Empirical (*n* = 1900)	Secondary (years 7 & 8) mathematics (Germany)
Lawless and Kulikowich (2006)	Empirical (*n* = 267)	Tertiary statistics (US)
Lopez et al. (1997)	Empirical (*n* = 296)	Secondary (age 15 -16) mathematics (US)
Marsh et al. (2005)	Empirical (*n* = 7913)	Secondary (years 7 to 12) mathematics (Germany)
Mitchell (1997)	Empirical (*n* = 51)	High-school statistics (US)
Mitchell and Gilson (1997)	Empirical (*n* = 598)	School and tertiary mathematics (US)
Preckel, Goetz, Pekrun, and Kleine (2008)	Empirical (*n* = 362)	Secondary (years 7 to 12) mathematics (Germany)
Renninger, Ewen, and Lasher (2002)	Observational	Primary school mathematics (age 11)
Sciutto (1995)	Evaluative (*n* = 17)	Tertiary statistics (US)
Trautwein et al. (2006)	Empirical (*n* = 14341)	Secondary (grade 9) mathematics students (Germany)

APPENDIX B: ATTITUDINAL BASED STUDIES IN A STATISTICS EDUCATION CONTEXT

Article	Description	Context
Alajaaski (2006)	Empirical (*n* = 53)	Undergraduate
Allredge et al. (2006)	Empirical (*n* = 203)	Undergraduate
Biajone (2006)	Observational	Undergraduate
Budé et al. (2007)	Empirical (*n* = 200)	Undergraduate
Carmona (2004)	Empirical (*n* = 827)	Undergraduate
Carnell (2008)	Empirical (*n* = 42)	Undergraduate
Cobb and Hodge (2002)	Observational	High-school
D’Andrea and Waters (2002)	Empirical (*n* = 17)	Graduate
Cybinski and Selvanathan (2005)	Empirical (*n* = 99)	Undergraduate
Elmore et al. (1993)	Empirical (*n = 289*)	Undergraduate
Estrada et al. (2005)	Empirical (*n* = 367)	Undergraduate
Faghihi and Rakow (1995)	Empirical (*n* = 75)	Undergraduate
Finney and Schraw (2003)	Empirical (*n* = 140)	Undergraduate
Fullerton and Umphrey (2001)	Empirical (*n* = 275)	Undergraduate
Garfield and Ahlgren (1994)	Empirical (*n* = 917)	High-school
Gordon (2004)	Empirical (*n* = 259)	Undergraduate
Gratz et al. (1993)	Empirical (*n* = 55)	Undergraduate
Gunnarsson (2001)	Empirical (*n* = 42)	Graduate
Kaplan (2006)	Empirical (*n* = 434)	Undergraduate
Lalonde and Gardner (1993)	Empirical (*n* = 91)	Undergraduate
Leong (2006)	Observational	High-school
Meletiou-Mavrotheris et al. (2007)	Observational	Undergraduate
Mills (2004)	Empirical (*n* = 203)	Undergraduate
Mvududu (2003)	Empirical (*n* = 229)	Undergraduate
Onwuegbuzie (2000)	Empirical (*n* = 225)	Graduate
Perney and Ravid (1990)	Empirical (*n* = 68)	Graduate
Schou (2007)	Empirical (*n* = 31)	Undergraduate
Schultz, Drogosz, White, and Distefano (1998)	Empirical (*n* = 94)	Graduate
Sorge and Schau (2002)	Empirical (*n* = 264)	Undergraduate
Suanpang et al. (2004)	Empirical (*n* = 230)	Undergraduate
Tempelaar et al. (2007)	Empirical (*n* = 1618)	Undergraduate
Tremblay et al. (2000)	Empirical (*n* = 166)	Undergraduate
Vaisanen, Rautopuro, and Sakari (2004)	Empirical (*n* = 123)	Undergraduate
Vanhoof (2006)	Empirical (*n* = 264)	Undergraduate
Zanakis and Valenzi (1997)	Empirical (*n* = 102)	Undergraduate

A FRAMEWORK FOR THINKING ABOUT INFORMAL STATISTICAL INFERENCE

KATIE MAKAR
The University of Queensland
k.makar@uq.edu.au

ANDEE RUBIN
TERC
andee_rubin@terc.edu

ABSTRACT

Informal inferential reasoning has shown some promise in developing students' deeper understanding of statistical processes. This paper presents a framework to think about three key principles of informal inference – generalizations 'beyond the data,' probabilistic language, and data as evidence. The authors use primary school classroom episodes and excerpts of interviews with the teachers to illustrate the framework and reiterate the importance of embedding statistical learning within the context of statistical inquiry. Implications for the teaching of more powerful statistical concepts at the primary school level are discussed.

Keywords: *Statistics education research; Informal inferential reasoning; Statistical inquiry; Ill-structured problems; Teacher professional development*

To be uncertain is to be uncomfortable, but to be certain is to be ridiculous.
- Chinese proverb

1. INTRODUCTION

Today, schools are increasingly being asked to prepare students to be flexible thinkers, lifelong learners, and to manage complexities of an uncertain world. Together with a dramatic rise in access to information and availability of technological tools, the increased focus on incorporating data into curriculum and on learning statistics in school has been an obvious welcome outcome. Recommendations in the U.S. (National Council of Teachers of Mathematics, 1989, 2000) and Australia (Curriculum Corporation, 2006; Queensland Studies Authority, 2006) have included a much greater emphasis on the teaching of statistics in school. Unfortunately, these recommendations are translated into lessons and assessments that often consist of little more than computation of averages and basic interpretation of graphs (Sorto, 2006). Specifically in primary school, "statistics is frequently portrayed in a very narrow and limited way, which can be encapsulated: *every phenomenon can be captured by a bar chart*" (Ben-Zvi & Sharett-Amir, 2005, p. 1).

New data visualization tools aimed at middle and high school students (e.g., Finzer, 2001; Konold & Miller, 2005) have provided opportunities and the impetus to include foundational statistical concepts not previously taught to this age, such as inferential reasoning (Ben-Zvi, 2006) and the broader process of statistical investigations (Wild & Pfannkuch, 1999). Konold & Pollatsek (2002) articulated this as "a new level of

Statistics Education Research Journal, 8(1), 82-105, http://www.stat.auckland.ac.nz/serj

commitment to involve students in the analysis of real data to answer practical questions" (p. 259). This commitment goes beyond simple interpretations of graphs and calculations of averages as is commonly taught in schools. The foundational difference in newer approaches to working with data is the shift from learning statistical tools and artefacts (measures, graphs, and procedures) as the focus of instruction, towards more holistic, process-oriented approaches to learning statistics. This move was initiated by the foundational work of Tukey's (1977) Exploratory Data Analysis (EDA) movement, but is now expanding to concepts that go beyond data analysis techniques. Exploratory Data Analysis is "about looking at data to see what it seems to say" (Tukey, 1977, p. v). It focuses on visual impressions of data as partial descriptions and supports attempts to "look beneath them for new insights" (ibid). More recently, research has focused on understanding attempts not just *beneath* the data, but also *beyond* the data, towards thinking and reasoning inferentially with data.

Although early research is showing promising results, little has been written specifically about what an inference-driven approach to learning statistics entails or what is meant by "informal inferential reasoning" in learning statistics. In broad terms, we consider informal inferential reasoning in statistics to be the process of making probabilistic generalizations from (evidenced with) data that extend beyond the data collected. Inferential reasoning will take on different levels of depth and technical detail at different levels of schooling and experience. In this paper, we build on previous research on inferential statistical reasoning to propose a theoretical framework for learning statistics using informal inference as a process of meaning-making and evidence-building. We then use classroom episodes to illustrate this approach in practical terms. Finally, we suggest avenues for teaching informal inferential reasoning that focus on statistics as a process in preference to teaching statistics as artefacts and objects. This discussion is situated within the issues and opportunities that arose in embedding teaching inferential reasoning with data (not always intentionally) into purposeful contexts with primary school students in Years 3 through 5 (ages 7 – 10) in Australia.

2. INFERENTIAL STATISTICAL REASONING

Research in statistics education has long suggested that students have difficulty using statistical processes appropriately in applied problems. "Inference is so hard that even professional researchers use it inappropriately" (Erickson, 2006, p. 1). For example, research on initial university statistics courses suggests that even students who could successfully implement procedures for hypothesis testing and parameter estimation were unable to use these procedures appropriately in applications (Gardner & Hudson, 1999; Reichardt & Gollob, 1997). Parallel findings in school statistics report difficulties students have appropriately using descriptive statistics and graphs to draw conclusions that make sense in problem contexts in which they are used (Pfannkuch, Budgett, & Parsonage, 2004); this is largely due to an overwhelming focus in schools on constructing graphs without knowing the reasons for doing so (Friel, Curcio, & Bright, 2001).

There has been a dramatic shift in statistics education research over the past few years from a focus on procedures—calculating a mean, interpreting box plots, comparing groups—towards a greater focus on statistical reasoning and thinking. One area of focus has been on reasoning about variation and distributions within the context of making meaning of the data (e.g., Cobb, 1999). Bakker and his colleagues (Bakker & Gravemeijer, 2004) have argued that by focusing on shape, students are able to shift their attention on holistic aspects of distributions. Konold and his colleagues (Konold et al., 2002) and others have argued for the need to focus learners on the modal clump (e.g., Bakker, 2004; Cobb, 1999; Makar & Confrey, 2005) and aggregates (Konold, Higgins, Russell, & Khalil, 2003). Despite this research, there is still an overwhelming emphasis in curriculum documents, national standards, and international assessments on statistics as interpreting graphs and finding averages (e.g., Sorto, 2006).

Although these research studies have often focused on statistical tools, it is important to note that they have situated these tools within statistical processes. It would be of concern if the intensive focus on graphs and calculations in school were simply replaced by a focus on clumps and distribution shapes as *tools* and *artefacts* of statistics rather than as embedded in the *processes* and *contexts* under investigation. Averages, distributions, variation, samples, modal clumps—these can be studied as objects in themselves, or as tools for understanding processes or group characteristics. It is vital that the focus in using statistical tools is embedded in the reason that we do statistics—to understand underlying phenomena. Certainly these authors are arguing for the latter.

This shift in research from statistical tools to statistical processes has been an important one and has raised new issues. Wild and Pfannkuch (1999) wrote extensively about the statistical investigation cycle and the dispositions and thinking that align with these processes. Research by Hancock, Kaput, and Goldsmith (1992) highlighted the challenges students encountered in connecting their statistical questions to the data needed as evidence, and then again linking their conclusions back to the questions under investigation. They argued that this part of the statistical process is largely ignored in school and needs greater attention. Focusing on investigating phenomena entails understanding the statistical investigation cycle *as a process of making inferences*. That is, it is not the data in front of us that is of greatest interest, but the more general characteristics and processes that created the data. This process is indeed inferential.

This recognition has sparked a great deal of interest in students' inferential reasoning in statistics and researchers over the last several years have "grappled with the conceptual building blocks for informal inferential reasoning" (Pfannkuch, 2006, p. 1). Like many of our colleagues in this area of research (see Pratt & Ainley, 2008), we would argue that inferential reasoning and statistical investigations cannot be separated. With this in mind, this paper discusses a framework for working to understand building blocks of informal statistical inference and inferential reasoning within a context of statistical investigations. By choosing to illustrate the framework within classrooms that are learning statistics through inquiry and investigation, our goal is not to focus on the distinction between product and process as it relates to inference, but to demonstrate that inferences are meant to be embedded within processes that create them. As such, we discuss both inference and inferential reasoning in this paper without trying to artificially separate these notions. In designing the framework, our aim was to investigate the concepts and processes of statistical inference and inferential reasoning more generally. We wanted to examine the potential of rebalancing the over-emphasis on procedures and calculations in school statistics, and capture the kind of informal inferential reasoning reported by Ben-Zvi and Sharett-Amir (2005) in their research with very young children exploring predictions of the number of baby (milk) teeth lost by their classmates.

3. A FRAMEWORK FOR INFORMAL STATISTICAL INFERENCE

A recent goal in statistics education has been to broaden the concept of inference from its immediate association with hypothesis testing at the tertiary level, to allow its application to work with children through their making of inferential statements. To separate these inferences which clearly do not involve formal procedures of hypothesis testing, we will adopt the term widely becoming utilized by statistics education researchers: informal statistical inference. We consider in broad terms *statistical inference* as both an outcome and a reasoned process of creating or testing probabilistic generalizations from data. By formal statistical inference, we refer to inference statements used to make point or interval estimates of population parameters, or formally test hypotheses (generalizations), using a method that is accepted by the statistics and research community. Informal statistical inference is a reasoned but informal process of creating or testing generalizations from data, that is, not necessarily through standard statistical procedures (see Zieffler, Garfield, delMas, & Reading, 2008 for an in-depth discussion of informal reasoning). The use of the word informal here is only meant to emphasize the broad application of inferential reasoning and open the possibility to consider statistical inference outside of formal procedures. Although the teaching of informal inference supports conceptual understanding of later formal statistical inferential processes, the goal is not necessarily to prepare students to do formal statistical inference. We see the potential for informal inference in deepening students' understanding of the purpose and utility of data more generally with direct applicability to making meaning of their world. Our goal here is not so much to *define* informal statistical inference as it is to *broaden accessibility to inferential reasoning with data*.

Initial concepts that we saw as critical included the following:

- Notion of uncertainty and variability articulated through language that broke from the mathematical convention of claims of certainty;
- Reliance on the concept of aggregate (as opposed to individual points) through the use of generalizations about the group;
- Acknowledgement of a mechanism or tendency that extended beyond the data at hand; and
- Evidence for reasoning based on purposeful use of data.

From these elements, three key principles (Figure 1) appeared to be essential to informal statistical inference: (1) *generalization*, including predictions, parameter estimates, and conclusions, that extend *beyond describing the given data*; (2) the use of data *as evidence* for those generalizations; and (3) employment of probabilistic language in describing the generalization, including informal reference to levels of certainty about the conclusions drawn. The first of these principles is particular to the process of inference, whereas the latter two are specific to statistics.

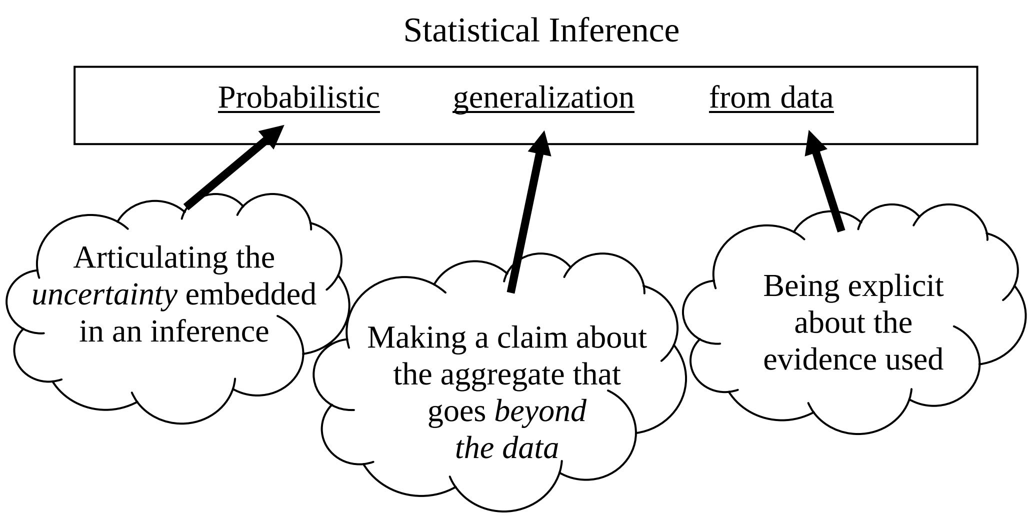


Figure 1: A framework for thinking about statistical inference

3.1. GENERALIZATIONS 'BEYOND THE DATA'

As in mathematics, statistical generalizations are abstractions from particular cases (data) to holistic statements that apply to a broader set of cases (population). In a traditional sense, generalizations (inferences) move from a sample to a specified population—for example, inferring from a sample of one class of 12-year-old children to a population of all 12-year-olds in a school. A fundamental difference between descriptive and inferential statistics is the act of looking *beyond the data* to cases outside of the sample at hand. For example, in some ways a mean is a generalization because it is an abstraction from individual cases to a general property of the aggregate. However, because the mean does not extend beyond the data at hand, we do not consider it to be inferential.

Consequently, EDA, which focuses on "looking at data to see what it seems to say" (Tukey, 1977, p. v) is not necessarily inferential if the focus remains on the data at hand. Dewey (1910) uses the term inference to talk about this move beyond the data: "The exercise of thought is, in the literal sense of that word, inference … it involves a jump, a leap, a going beyond what is surely known to something else accepted on its warrant" (p. 26). Being able to separate description of the data at hand from an inference to the population is not a trivial matter. Researchers like D. Pratt (personal communication, 7 July, 2005) and Pfannkuch (2006) have documented difficulties encountered by teachers and learners in slipping between talking about the sample (data at hand) and population (inference beyond the data). Our hope is that by making explicit the importance of moving beyond the data when making generalizations in statistics, these slippages can be reduced.

Alternatively, a population may be more conceptually defined by describing a mechanism to a process (as in the making of widgets in a factory, see Rubin, Hammerman, & Konold, 2006) or future population not yet existing (as in students who will be in the school in coming years, see Makar & Confrey, 2004). Generalizations may be used to either generate hypotheses or evaluate them. By "evaluate" here, we are not limited to the use of standard statistical tests, but refer to attempts at processes to formally or informally assess the viability of a claim against the given data. Generative hypotheses, on the other hand, are speculative statements which are created by a reasoned process but for which their likelihood has not necessarily been systematically assessed.

3.2. DATA AS EVIDENCE

By data, we refer to evidence that is accepted by the community in which the evidence is being presented. Data may be numerical, observational, descriptive, or even unrecorded. What is important is that their use as evidence is accepted within the context it is being used. It might be expected that the person making the inference would provide an explanation or argument (implicitly or explicitly) that draws on the data as evidence for the inference. At a very young age, for example, we may be more likely to accept observation as data to encourage initial development of inferential thinking (Ben-Zvi & Sharett-Amir, 2005) and develop a language for explaining and reasoning with data. Later on, we encourage students to critique this type of evidence in preference for more robust and reliable approaches. Hancock et al. (1992) have noted the difficulty that learners have in connecting data collected to the question under investigation and conclusions drawn. Attention to the need to make this connection more explicit may help teachers to better support these links.

One type of informal inference is a creative, inductive process in which a learner generates a tentative hypothesis by observing patterns in the data. A potential support for creating tentative hypotheses is through a process of *abduction*, an inference or theorizing to *explain* or account for the data in relation to the context. This process, as explained by Galileo (1638), “prepares the mind to understand and ascertain other facts without the need to recourse to experiment” (as cited in Magnani, 2001, p. 37). This focus on context and explanation is important, as noted by Dewey (1910) in saying that “the data at hand cannot supply the solution; they can only suggest it” (p. 12). It is up to the one making the inference to connect the evidence meaningfully to the claim and to explain it in terms of the context.

A powerful approach to improving students’ use of statistical reasoning and thinking is by embedding statistical concepts within a purposeful statistical investigation that brings the context to the forefront (Makar & Confrey, 2007). By focusing on trying to find out something of interest to students, they gain important insight into how statistical tools can be used to argue, investigate, and communicate foundational statistical ideas. Wild and Pfannkuch (1999) argue that a number of elements are central to statistical thinking, including the following: recognizing the need for data; transforming situations and representations into meaningful statistical tools that can provide insight into the problem; having opportunities to recognize, work with, and deepen understanding of variation; envisioning statistics within a framework of its utility to gain insight; and being able to shuttle between the context sphere and statistical sphere. Their inclusion of *recognition of the need for data* as a type of thinking foundational to statistical thinking is often overlooked, particularly in schools. Perhaps it’s considered trivial, too obvious. After all, without data, no statistics can be done.

3.3. PROBABILISTIC LANGUAGE

Because inferring to a population contains elements of uncertainty, statistical inferences must contain probabilistic language, implying *statistical tendency*, and/or *level of confidence* or *uncertainty* in a prediction. We are not implying here that an explicit or quantified level of confidence needs be indicated (as is done with confidence intervals), although the idea is related to concepts of chance and our confidence in inferential predictions based on the strength of evidence presented (Rossman, 2008). Because generalizations go beyond the given data, they cannot be stated in absolute terms. The problem of deterministic thinking in statistics has been well documented (e.g., Abelson, 1995). Probabilistic language can be any language appropriate to the situation and level of students to suggest uncertainty in a speculated hypothesis, that a prediction is only an estimate, or that a conclusion does not apply to all cases. For example, in using data to estimate the average height of an eight-year-old, students may suggest the typical height to be ‘around 130-138 cm’ rather than reporting more precisely that the typical height is 132 cm, which may be the mean of the data or value with the highest frequency (mode) from their class (Makar & McPhee, in press). Or a six-year-old child may suggest that the most common way for children at school to travel to school ‘may be’ by bus rather than stating it ‘is’ by bus. Probabilistic language can go beyond simply avoiding deterministic claims, however, as relationships involving overlapping distributions or multiple interpretations of a given distribution also necessitate avoidance of overly conclusive or excessively precise statements (Makar & Confrey, 2004).

Our focus on language in inferential reasoning emphasizes the importance of expressing uncertainty in making inferences. School statistics must work harder from an early age to break the black-and-white approach to making inferences from data. Within statistics education, this is an area that needs a great deal more attention and research.

4. CONTEXT OF EPISODES

Section 5 reports on episodes from the initial phase (18 months) of an ongoing four-year study investigating processes of teachers' learning to teach mathematics and statistics through inquiry in a problem-based environment. Although the focus of the larger study is on inquiry, not inference, the data collected provide a number of opportunities to gain insight into teachers' use of inference in teaching statistical inquiry. The larger study follows a model for design experiments (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), where the context is simultaneously studied and improved through iterative cycles. The cycles in the study served several purposes. For the researcher, it provided multiple opportunities to observe and influence the study context. For the teachers, it gave them ongoing experiences to reflect on and improve their ability to teach statistical inquiry. For the students, it allowed opportunities to build on previous learning and engage with statistical inquiry in increasingly sophisticated ways.

Four primary school teachers at a suburban state school in Australia participated in the first year of the study. The teachers volunteered for the study after being identified by their school's deputy principal (in some cases, with a bit of friendly coercion) and attending a briefing session by the first author. They were all considered to be effective and innovative teachers, in agreement with the *idea* of including inquiry in mathematics, and interested in learning how to implement statistical inquiry with their students. Although they were all reform-oriented teachers, they represented a broad range of experience – from a teacher in his first year in the classroom to a veteran teacher with over 30 years of experience. After the first year (2006), one teacher (Josh) in the study was transferred and replaced (Elise); another teacher went on leave after the first term in 2007 (Carla).

At the time the reported episodes were analyzed, the study had undergone six cycles over 18 months with a cycle being one school term (lasting about 10 weeks). In each cycle, teachers taught an inquiry-based unit on statistics that they designed themselves or adapted from published units (e.g., Gideon, 1996). These units were not explicitly designed to focus on inference and indeed many did not include any significant use of inferential reasoning as inference was not directly a part of the mathematics syllabus for this age group (Queensland Studies Authority, 2006). For units that did lend themselves to informal statistical inference, we draw attention to ways in which inference played a role in the teachers' learning to teach statistical inquiry.

In four of the six terms, teachers met with the first author for a full day workshop after which teachers designed and implemented an inquiry unit with their students lasting from one to three weeks (Table 1). Full-day learning seminars were audio or video recorded. Each seminar involved four distinct sessions, each lasting 30-90 minutes. The initial session consisted of a discussion of overarching issues that arose in teaching the units. Next, the teachers would engage in an activity, as learners, aimed at a particular aspect of statistical inquiry. The teachers were given some planning time with resources following this. The day would end with a sharing sessions of plans and expectations for the next unit.

Each term (except in the third cycle), the researcher videotaped several lessons in the units that teachers implemented (60%-75% of lessons), photographed or collected unit artefacts (e.g., lesson plans, student work), recorded researcher observations and reflections, and audio or videotaped follow-up interviews to debrief after the unit. The episodes below draw on relevant aspects of the data that provided insight into the framework described in Section 3.

Table 1. Inquiry units taught by the teachers in the study

TERM		Units
2006	1	*Can you roll your tongue?* - Exploring hereditary traits (Kaye & Carla, Year 4) *Are athletes getting faster?* - Investigating winning times at the Commonwealth Games (Natasha & Josh, Year 5)
	2	*What's in your lunchbox?* - Investigating healthy lunches *Kangaroos!* - Modelling and interpreting data from a predator-prey game (Natasha) *How fast is a blue-tongued lizard?* - Class negotiated investigation (Josh)
	3	*Tibia mystery* - Estimating height from a tibia bone found at an archaeological dig (Kaye & Carla) *Is anyone a "typical" Year 5 student?* - Developing a survey and exploring "typical" (Natasha)
	4	*How many commercials does a typical Year 4 student watch in a year?* (Kaye) *Comparing students' ages* – Contrasting student ages with family members (Carla) *Investigating paper airplane designs* (Natasha) *Designing a parachute for an egg* (Josh)
2007	1	*How many spritzigs do we have in our class?* - Collecting, organizing, displaying, and interpreting survey data (Kaye, Year 4) *How tall are Year 6 students compared to Year 1?* (Carla, Year 6) *Citizenship in Australia* (Natasha, Year 5) - Collecting, organizing, displaying, and interpreting opinion poll data *Comparing handspans* – Collecting and organizing data on students' handspans (Elise, Year 3)
	2	*Investigating healthy cereals* – Analysing nutritional information on cereals (Kaye) *The effect of pollution on plant growth* – Experimental design (Natasha) *Do we have healthy lunches?* – Organizing categorical data (Elise)

5. EPISODES

In studying teachers' teaching of inferential reasoning, we chose a context in which the focus is on understanding a particular question or situation, rather than examining decontextualized data. This gave us an opportunity to envision its use in a more purposeful way. In this section we will use three episodes 'to think with' to consider how the framework might suggest the potential for introducing informal inferential reasoning at the primary school level:

- Section 5.1 examines Natasha's survey with Year 5 (age 9) students, focusing on their challenges in (not) looking beyond the data;
- Section 5.2 focuses on the investigation by Carla and Kaye's Year 4 (age 8) students on healthy lunches and how they used generalizations to connect findings to evidence;
- Section 5.3 discusses Elise's Year 3 (age 7) students making predictions about handspans of children, focusing on the concept and language of uncertainty.

5.1. NATASHA: CHALLENGES IN (NOT) LOOKING BEYOND THE DATA

An important tenet of statistical inference is its power to utilize given data to make predictions, estimate parameters, or draw conclusions about a population or process *beyond the data*. In order for an inference to be valid, it must incorporate its target – that is, be an inference about a particular population or process for which the data are being used. Pfannkuch (2006) and Pratt (2005) showed that this is not a trivial concept. In this episode, we examine how this slippage between the data and population potentially created challenges in the unit.

Natasha is a teacher of students in Year 5 (age 9). Part of the Year 5 syllabus for social studies is an introduction to government and citizenship. Drawing on the success of the units she designed the previous year which integrated statistics with other content areas, Natasha decided to teach a statistical inquiry unit (her fifth in the study) in which students investigated the opinions of children and adults about their views of Australian rights of citizenship. A pre-constructed survey instrument from a local school resource on citizenship was used to collect data. After preliminary discussion on citizenship in Australia, she introduced the issue they would investigate.

> Natasha: Over the last few months, there have been a lot of things in the media about who has the right to be an Australian citizen and what's required to become an Australian citizen. So, I thought it might be interesting if we found out what people think. OK? The people being, who?
>
> Student: Us?
>
> Natasha: Yes, you people, your parents, people in our community. It might be interesting to find out what they think. (Class 5N, 8 March 2007)

Natasha introduces the purpose of the unit to students to find out "what people think" about citizenship issues in Australia. She alluded to the 'people,' or target of their investigation, as "you people, your parents, people in our community." By not posing a specific question, it was unclear when she set a goal of finding out "what people think," whether the intention was to develop hunches, draw conclusions, gather insights, make predictions, or just describe the opinions of those that they would be gathering data about.

Clearly there was no intention to survey an entire community to get this information, so if her intention was to draw conclusions beyond their survey data, Natasha thought it would be important to discuss particular issues that would allow them to use a sample of the data (their survey data) to draw conclusions about the views of the community (beyond the data they collected). Several issues were considered, such as sample size, representativeness, and comparisons.

> Natasha: Ok, how many people do you think we would need to survey to get a fairly good idea of what people in this local community, in this school community, think?
>
> Student: 5?
>
> Natasha: Five people! Would that be a really good indication of what people in this school community think?
>
> Students: 55 … 100 or so? … 250? … Maybe 3 or 4 classes?
>
> Natasha: Ok, let me clarify I'm not just interested in what kids think. I'm interested in what parents, and perhaps grandparents of our school community think. Or perhaps aunts and uncles, older people not just those under 18.
>
> Students: 500? … 80%?
>
> Natasha: So, 80% of the whole [River] School community?
>
> Student: Maybe, one of the teachers from every grade?

Natasha: Do you think if you ask all of the teachers what they think, you'd get a good indication of what people in [River] School community think? … Do you think teachers might be completely representative of the community? … Teachers do have a [more liberal] political bias. OK? So maybe just giving it to the teachers it won't work. …

Student: I was thinking maybe about surveying the adults about the same amount as kids.

Natasha: Ok, so you'd like to see if kids think differently than adults? Good! That would be very interesting, who would like to know that, if kids think differently than adults? [Students respond positively] … Ok, that sounds great. I have printed off two copies of the surveys for each person, one on each side of the form. Why don't you get an adult to fill out one side? Now it would be good if you didn't just ask a parent, if you have access to an aunt, an uncle, a grandparent, an elderly neighbour, but with mum and dad's permission. (Class 5N, 8 March 2007)

Natasha was trying to stress to students that not only was it important to think about *how many* people they would need to ask, but also to ensure that their sample was *representative* of the community (however vaguely defined) that they were describing. It was not clear at this point how the sample would be used to find out the views of the local community – whether it be to explore or predict general views, estimate proportion of people having particular views, or another purpose. Although somewhat vague about the population they would be investigating, Natasha worked to have students consider the validity of the data they were collecting in order to later draw conclusions.

After students collected their data, the unit struggled to make significant progress over the next couple of weeks and a formal 'conclusion' was never really made. In reflecting on the unit, Natasha made a number of observations that she felt may have explained this. What is interesting about her explanations is that in some ways they allow us to speculate about possible considerations for supporting children in thinking inferentially about data.

The primary issue, she felt, was that the topic of citizenship was not one that engaged the kids, saying, "They weren't really interested in it. So they didn't have their heart in getting to the bottom of something." This was an interesting comment and suggests that in working to look 'beyond the data,' it is not just making a conclusion *about data* that provides the conceptual muscle to draw inferences, but a conclusion *about the situation* which the data are meant to represent or signify. Perhaps a focus on an interesting problem and context may influence students' engagement with being inclined to look beyond the data they have. This further suggests that students need a particular level of complexity to engage with in order to consider possible avenues to connect the data with the context. Natasha spoke at some length about the structure of the data as a factor that limited the students' ability to connect the data to something beyond graphs.

Natasha: On top of that, most of the data was dichotomous, so there was really not much that we could do with it. After they had done the 'yes, so many people thought this and so many people thought that,' there wasn't an awful lot that I could find that I could do with the data. … It was hard for me to keep the enthusiasm going to keep them enthused. So, I got to the point where I just thought I ought to throw this out the window. [laughs] So all I actually really ended up discovering from that was the extent to which children could graph. And that was interesting in itself, their understanding of data and how to represent data, but it really didn't allow them to do any of the higher-order thinking or explanation, uh, exploration for themselves. (Interview, Natasha, 26 May 2007)

Natasha pondered the lack of complexity of the data as a factor that inhibited the unit from successfully transitioning from describing data to generalizing from the data in order to draw conclusions or make predictions about the beliefs in the community. Another issue she felt was critical was the driver behind the work, the purpose for which they were using the data:

> Natasha: I also think that part of that was posing the question. In the studies [teaching units] that have been successful, [they] started off with quite a clearly posed question and I lost the plot on the government unit because I didn't pose a clear question, so when I got kind of bogged down with it all, I didn't have direction. So I didn't really have direction for the children. And I think that's what really made that unit hard, too.
>
> Researcher (R): … I know that one of the goals that you had said before, was that you wanted to see how they do with the graphing.
>
> Natasha: … It gave me data to use for graphing, but it would have been no different if I had just done, the old, you know, 'how many of you play soccer,' 'how many of you play netball.' Whatever. And put that up on the board and had them graph it. … It was just a straightforward graphing unit, it wasn't really an investigation. (Interview, Natasha, 26 May 2007)

Natasha's point is that she was not sufficiently focused on answering any particular question, rather than on getting the students to graph. The spotlight on the tool rather than the purpose makes drawing inferences particularly challenging if students do not know the purpose of the inference within a meaningful context. Although Natasha's initial intent of the unit was likely to use the data inferentially, the focus on the data at hand provided little opportunity to do so. This may provide some evidence of the importance of being explicit in articulating the population and particular question under investigation. Table 2 provides a summary of the alignment issues between Natasha's episode and our framework for thinking about informal statistical inference (Figure 1).

Table 2: Natasha's episode aligned with principles of informal statistical inference

Framework	Comments
Generalization beyond the data	Although the initial intent was likely to move students towards drawing conclusions about their survey data as inferences to the population, this did not materialize during the unit. The teacher named a number of issues that may have contributed to the difficulties – the lack of a driving question, a context that was likely not engaging to students, and overly simplistic (dichotomous) data that lacked complexity needed for interesting interpretations. Another possibility may be that too much energy was focused on graphing skills and there was not explicit attention to differentiating between describing the data and using the data to draw inferences about the larger population.
Data as evidence	Students struggled to connect conclusions to the data collected. Without the unit being completed to the point of drawing conclusions, this element of the framework was difficult to assess.
Probabilistic language	No significant attention to language of uncertainty or level of confidence appeared to be expressed.

In Natasha's unit, the focus on the data in a descriptive lens likely prevented opportunities to consider the other two aspects of the framework – data as evidence and probabilistic language. This suggests that the focus on generalization beyond the data is foundational to the other aspects of inferential reasoning.

5.2. CARLA AND KAYE: USING GENERALIZATIONS TO CONNECT CONCLUSIONS TO EVIDENCE

Two Year 4 (ages 8-9) teachers, Carla and Kaye, developed an inquiry unit in Term 2 (their second data inquiry unit in the study) to investigate whether students had healthy lunches. The unit was designed to tie in with a set of lessons on healthy eating as part of the Health curriculum. Because their previous (first) unit had been more structured (with students being given the investigation questions and guided through the inquiry cycle), Carla and Kaye decided to make this unit more student-directed, where students (in collaborative groups of 3-4 students) would develop questions to investigate, decide on appropriate data to collect, do their own analysis, and present their findings to the class. As students were preparing their findings to present to the class, the teachers found that groups were struggling to connect their conclusions to the data, as well as link the data back to the question under investigation. In a resource book they were using (Gideon, 1996), Carla and Kaye found mention of generalizations as part of the process of communicating conclusions that provided them with an idea to help students make these connections.

In an interview at the end of the lunchbox unit, Carla and Kaye spoke about the role that 'making generalizations' had in helping students draw connections and for using their investigation as a launching point for generating new hypotheses.

Kaye: I found some still needed support [making generalizations]. They didn't have any trouble collecting the data. They're quite happy to go out and do that. But then when the data comes back, actually looking at what they need to specifically [answer the question] … they collected all this lovely data, and they might have collected boys and girls and things like that, but the original question was 'Do students in Year 6 eat healthier than students in Year 2?' … I said [to students], 'This is extra. If you've collected that data, when you present the data, these are extra things, so you might be able to, once you've answered your question as such with a generalization, you can then go back and use that extra data [boys and girls] to give you extra information, but really, the first thing that, the first task that you've got to get around is that this is your question, and this is what you're setting out to collect that data on.'

Carla: And my guys, we actually, they quite liked at the end when they were doing their presenting, when we asked them, when I asked them, or the kids asked the questions, um like 'if you had collected this extra data, what other generalizations might you … have made?' and I think that was a good link. … I was happy with that, 'what if' [questions]. That's what I was hoping for. (Interview, Carla and Kaye, 17 July 2006)

For these teachers as well as their students, generalizations played an important role in supporting understanding of two important processes. They supported students in making conclusions for the question at hand (connecting investigation questions to data collected and subsequent findings), and in seeing how conclusions and data (actual and potential) can also generate novel hypotheses and questions to investigate, particularly when students work to explain their findings (abduction). Although Carla's use of generalizations often didn't extend beyond the data, this term helped students move from a focus on individual points towards a more aggregate view of the data. Carla further discussed how the class discussions of their generalizations helped students situate their data investigation within the larger context of peer pressure and body image.

> Carla: It led to a discussion, really, on stereotypes and body image. Because my kids came to the realization that … boys ate more junk food, because, in the upper grades, because girls were watching their figure, but boys could afford to because they did more sport or they used more energy. So, and that led onto our [health] unit discussions which was perfect. [Laughs] Yeah, that was really good. (Interview, Carla and Kaye, 17 July 2006)

Overall, Kaye and Carla recognized that though students tended to struggle making clear connections between conclusions and the questions under investigation, their use of generalizations to make sense of these connections was a productive avenue for supporting student understanding. This connection not only helped students understand this connection, but helped the teachers in both learning and teaching this concept.

> R: Any surprises or unexpected outcomes [in the unit]?
>
> Carla: Just that we figured out what generalizations are! There was one day that, last-, the first time we did it [make generalizations, in the unit in Term 1], it was very, I found it very difficult to, kind of, teach the genre, I suppose, or what is a generalization. I tried, I don't know. But this time, it was just a matter of saying [to students], 'Well, what does that graph *mean*? Or what does it *tell me*? Ok, I haven't time to read all those dots, so *what does it tell me*?' And the kids would say "Oh, there, there were more girls than boys [who had a healthy lunch]." And it just made it so much easier. Yeah, that was my 'whew!' moment.
>
> R: … There was just one day that it just seemed to really click for you. Talking about generalizations, and then how you link it to the data and try to find out, well, 'How do you know that that's true?' What made you decide to-
>
> Carla: Well, I've just been battling the whole time, thinking, how can I make it clearer, really, what generalizations are? I don't know what it was. The kids were not getting it. I think I was standing with a group [in one lesson], they were still all looking at me. I didn't know how to ask and they didn't know how to answer me. And then it was just, 'Ok, well, imagine I didn't have time,' like I said, 'what would that tell me?' And then I realized, 'oh, that's what a generalization is!' It's just that more, that simple idea or notion of that made it easier. Yeah. … And then turning it around, 'If this is your generalization, where did you get that information from?' That made the big difference. Because then, the kids had to figure out what graph it was. So asking both ways, that was good, that worked well. (Interview, Carla and Kaye, 17 July 2006)

Kaye and Carla further elaborated how using generalizations helped students go beyond thinking about individual data points towards considering the data as an entity.

> Kaye: I guess trying to get them, rather than just saying, 'there were two people who liked ___,' we had to come with something, but, um, to answer the question. [Students would say] 'There were two who did this, and two who did that, and who did that, and four who did this and four who did that.'
>
> Carla: And that's not a generalization.
>
> Kaye: And it's not really, it does *interpret* the results, but it's not an *overall* interpretation. (Interview, Carla and Kaye, 17 July 2006)

It was unlikely that Carla and Kaye were thinking about generalizations as being 'beyond the data' in the sense of making predictions or theorizing about populations or processes. However, their use of generalizations in this unit served as a step in better understanding the process of a statistical investigation themselves. Table 3 provides a summary of the alignment issues between Carla and Kaye's episode and our framework for thinking about informal statistical inference.

Table 3: Carla and Kaye's episode aligned with informal statistical inference framework

Framework	Comments
Generalization beyond the data	The teachers used the concept of generalization to support students in moving beyond a focus on individual points towards a more aggregate view of the data. Although they did not distinguish between generalizations that interpreted the data at hand (descriptive) and those that stated a generalization beyond the data (inference), their utilization of generalizations may better prepare them to make this distinction later on.
Data as evidence	As with the previous example (Natasha's Year 5 class), students struggled to connect conclusions to the data they had collected. However, the teachers found that by having students create generalizations, it was easier to then ask them to connect that generalization back to its evidence and to the question that drove the data collection.
Probabilistic language	No significant attention to language of uncertainty or level of confidence was expressed.

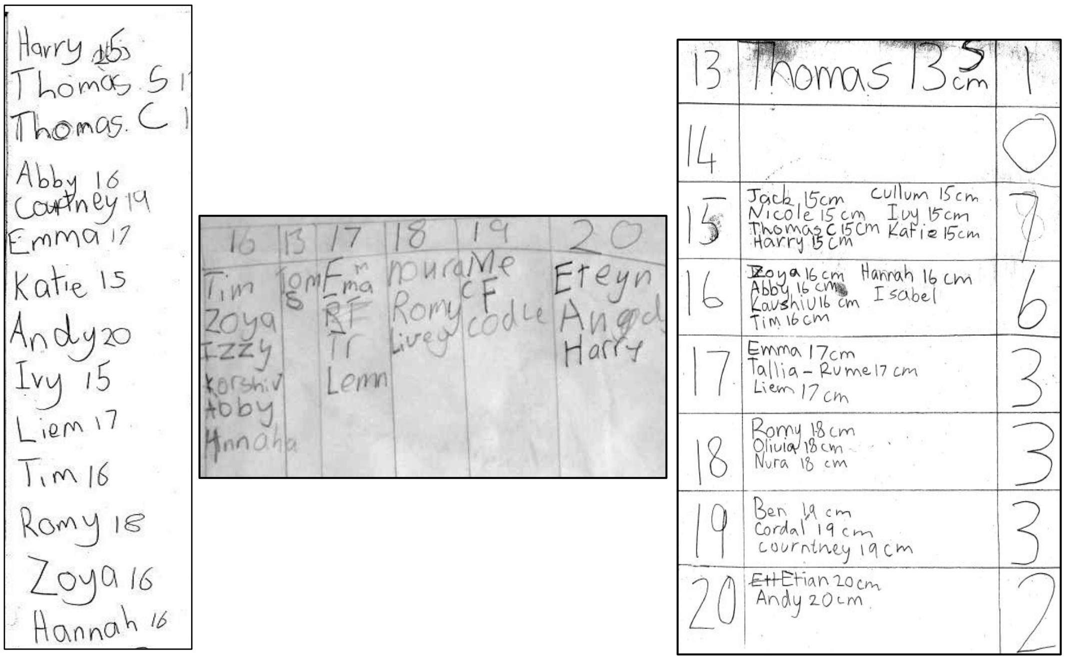
5.3. ELISE: USING INFORMAL INFERENCE TO MAKE PREDICTIONS

In a Year 3 (age 7) class early in the year, Elise developed a unit lasting about seven days that had students complete multiple cycles of collecting and organizing data on the handspans (distance from thumb to smallest finger on an outstretched hand) of students in their class. To find their handspan, students each traced their hand on paper, drew a line to connect the end of their thumb with the farthest tip of their smallest finger, and then measured this handspan with a centimeter ruler. They created their own methods to collect and record the data of their classmates. This process was not straightforward, but through sharing ideas and challenges encountered, students came up with a process of recording each student's handspan next to their name (not in any order). In one lesson, after students had collected and recorded their data, Elise began probing students to consider how the data might help them find out more about their class's handspans.

> Elise: What I'd like for us to think about this morning is finding out how many children have the smallest handspan, how many children have the biggest handspan, how many people have different measurements for their handspans. So I want you in some way, to go and find that information, so that you can share it with somebody. (Class 3E, 8 March 2007)

Elise worked hard not to *tell* students how to organize their data, but rather used skillful questioning to encourage students to seek more efficient and purposeful ways to organize their data. In the first iteration of collecting and organizing their data, students chose to organize the names and handspans of the students in their class as a list (Figure 2a). To encourage the students to organize the data further, Elise probed the class to find a more efficient way to find out the shortest handspan, the longest handspan, as well as the most common handspan measurement than to search through the data each time.

From the list, students began sorting the data into groups, sometimes with a partial ordering of the groups and with frequencies listed (Figures 2b and 2c). Clearly in asking students to consider the extreme handspans (shortest and longest) and most common group, the questions did not prompt students to think beyond their data. However, the cues Elise used to prompt the students to seek a purpose for organizing the data enabled greater cognitive access to consider an underlying structure of the data. It was this process of moving from thinking about individual handspans towards considering a collection of handspans that may have provided a foundation for later inferential reasoning.



(a) listed (b) grouped into columns (c) ordered with frequencies

Figure 2: Students' organization of their class's handspan data. Students collected their data (a) listed, (b) grouped into columns, or (c) ordered with frequencies

At the end of the lesson, Elise asked the researcher to suggest what she might do next with the students to deepen the investigation. Following a suggestion from the researcher, Elise decided to present an opportunity in the next lesson for students to predict handspan data from the class next door. The next day, after students had shared their strategies for organizing their class's data more efficiently, Elise asked them whether they thought the class next door would have similar handspans to their class.

Elise:	So which group [in our class] did you find had the most number of children in it?
Student:	15
Elise:	15 cm. So if we went next door and asked [Miss Miller's class, 3M], do you think that 15 would be the biggest group as well?
Student:	I think yes.
Elise:	Because?

Student: Because … [our class is] about average height and taller people will probably have bigger hands and smaller will have smaller hands.
Elise: So you think that children in 3E [our class] are average size children? And you're going to suggest that children in 3M [the class next door] are probably average sized children. Do you think that anyone in 3M is likely to have a handspan shorter than 13 cm? … Do you think that some children in 3M would have handspans longer than ours? (Class 3E, 9 March 2007)

The class ran out of time at that point, but the discussion gave Elise an idea to extend the investigation further by asking Miss Miller, the teacher next door, to have her students trace their hands onto paper, just as Elise's class had done to collect their own data. The next week, Elise presented her class with an opportunity to organize 3M's data and compare them to their own.

Elise: If we went next door, would that be true of the class next door? Would we also find that 15cm was the largest handspan group? Would we also find that 20 was the longest? … Some of you were saying yes, and some of you were saying no. So I did a bit of a tricky thing. I actually went next door and on Friday I got Miss Miller to draw, for the kids next door to actually [trace] their handspans. (Class 3E, 13 March 2007)

After measuring the neighboring class's handspans, Elise asked students to work in their groups to decide how they might plan to collect, record, and organize this new data, then to share their ideas with the class. One student described her group's idea to set up rows for each measurement and list the names of students with that measurement in the row (as in Figure 2c).

Student: So we're sort of mixing all the ideas. We're mixing the columns idea, the number idea, and the how many …
Elise: Now, how did you know to start with thirteen there at the top?
Student: Well because we already know, we've already answered one of your questions [about our class] which is which was shortest which was longest, which was Fletcher and Greg was the longest and Eddie is the shortest, so we knew that nothing was under Eddie's, which is Eddie's is 13. That must mean we must start with 13. (Class 3E, 30 March 2007)

Elise noted that in listing their measurement categories, Beth had not included a column for 14.

Elise: Why did you think there would be no one with 14?
Beth: Because I didn't really think anyone in 3M would have a big enough hand or small enough hand to make it 14.
Elise: Nobody in [our class] had 14 did they?
Beth: No. So I was guessing about kind of the same amount of numbers. (Class 3E, 30 March 2007)

The responses in the excerpts above and later in the lesson suggested to the researcher that for many of the students, their predictions about the distribution of handspans in another class would be the same as, and perhaps even identical to, the distribution of handspans in their own class. Later, as students were beginning to organize the data that they had collected from the class next door, one of the students noted with surprise that the data from 3M differed from what she expected, "I just got a bit surprised when I found that someone had [a handspan of] 12" (Class 3E, 30 March 2007).

Elise decided to work with the students further to make these predictions more explicit by asking the class to consider whether they thought that the data they had (and already discussed) about their own class would be similar to the handspans one would find in other classes (data they didn't have). She had written on the board: "Are our handspans the same as another class?"

Elise: This [points to the question] was one of the questions that we were working towards last time. And we were going to see, well, if our class would be the same as the other. There was a word that we were using, I don't know if you remember this word, but it was 'typical.' [Writes the word 'typical' on the board.] Would we be saying that our work is typical? Like, our [class's] handspans, are they typical? So would every Year 3 class at [River] State School have the same highest handspan measurement? [Class: No] Or the same lowest handspan measurement? [Class: No] Or would we have the same middle handspan measurement? [Class: No] Or, which was our most common group? Remember what – 15 cms? Well, would every Year 3 class have 15cm as their most common measurement?
Class: No.
Student: No, you never know. (Class 3E, 30 Mar 2007)

The use of the word "typical" here is quite different than is generally encountered in primary schools, where "typical" often refers to an average (Makar & McPhee, in press; Mokros & Russell, 1995). In this case, the teacher was asking (perhaps without realizing it) whether the distribution of their handspan data was representative of the distribution of handspans for all Year 3 students at their school. Here, Elise was working with students on building their ability to think "beyond the data" to build their informal inferential reasoning (possibly due to encouragement by the researcher to consider more inferential thinking). It was interesting to note, however, that initially the students did not see the data they had as useful evidence for predicting what they might find in handspans more generally. However, the comparison activity did appear to be important for two reasons. For one, it supported students' thinking about the data as an aggregate by having to compare characteristics of the two classes of handspans. Secondly, it appeared to expose students informally to the notion of variability between distributions; this between-group variability that arises from comparing distributions may have provided foundation for thinking inferentially.

Although Elise had been able to get the students thinking about whether the data next door, in Class 3M, would be the same as their own class, it didn't appear that students were seeing the data they had as useful evidence for predictions. In order to investigate this further, the first author asked to teach the class to see whether their experiences in comparing data from the two classes could be used as evidence to quantify predictions about data they didn't have from a third class, 3K.

R: I want to know something about 3K's handspans. Who can tell me something about 3K's handspans?
Student: That we don't know anything about their handspans.
R: Oh, you don't know *anything* about their handspans?
Student: Because we haven't done it yet.
R: Because you haven't got the data?
Class: Yes.
R: You don't know anything *at all*?
Class: No
Student: Cause we haven't even started on 3K's [data]! (Class 3E, 2 April 2007)

It seemed clear from their responses that the children did not see the data they had already worked with from their own class and another class as potential information about data from a new class. The researcher pushed them further to try to see their data as evidence for making an inference about unknown data.

R:	Can any of you make a *prediction* then, about what you might find in 3K's data?
Student:	They might have bigger handspans?
R:	They might have a bigger handspan? Now I noticed you used the words 'they *might* have.' You mean you're not sure? … Is it totally guessing?
Class:	No.
R:	Does someone want to make a prediction about what we might find in the data for 3K?
Student:	I think the largest handspan might be 19 or 21, maybe? And the shortest one might be 13 or 11? (Class 3E, 2 April 2007)

Other students offered similar predictions, including predictions for the most common group. This was the first time that students began to articulate their predictions probabilistically, using the data they had as evidence for their predictions. The researcher probed the students further about the source of their predictions more explicitly.

R:	So how could you make those predictions when we haven't collected the data? How can you make those predictions?
Student:	… I think it's because everyone's actually using the data that we already collected. (Class 3E, 2 April 2007)

The researcher continued, this time asking them to predict the handspan of a new student who might join the class.

R:	Now I want you to make another prediction. Let's say that you get a new student in 3E [your class]. I want you to write down a prediction about what you think their handspan might be, if you get a new student in 3E. …
Students:	I think they might be 15, 16, or 17. … Maybe 16 or 15?
R:	I love how I'm hearing that word 'maybe.' Maybe 15 or 16? Where are you coming up with that, those numbers 15 or 16?
Student:	Because …. I don't think it would be 13 or 20. (Class 3E, 2 April 2007)

It appeared to take several iterations of predicting and probing students to enable them to begin to see the data they had already collected was useful as evidence for making predictions. In addition, students appeared to have some intuition already that their predictions contained an element of uncertainty. They expressed this through more uncertain (probabilistic rather than deterministic) language and estimating a range of values rather than a single value. Finally, the work they had done earlier in organizing their own class's data and comparing it to another class appeared to support the move to more inferential reasoning by helping them to improve their thinking of the data as an aggregate and to move beyond a deterministic view of the distribution to incorporate potential variability between distributions. Table 4 below provides a summary of the alignment issues between Elise's episode and our framework for thinking about informal statistical inference.

Table 4: Elise's episode aligned with principles of informal statistical inference

Framework	Comments
Generalization beyond the data	In probing students to consider characteristics of their data, Elise was able to get students to think about the data as more than a list of individual students. The categories students formed to organize their data into columns may have supported them to think of the data as a distribution. However, initially, they seemed to assume that data collected from another class would have very similar, or even identical, properties (smallest, largest, most common) to the data from their own class. The comparison with a second class's data helped students begin to perceive a possible distribution for each data set, but still not generalize beyond data they had in hand. However, it may have moved them to think about between-group variability that supported later inferential thinking in making predictions.
Data as evidence	It took several iterations of discussing and organizing data before students began to systematically use the data they had as evidence for making predictions.
Probabilistic language	Students only began to articulate uncertainty when they were asked to quantify predictions about data they didn't have.

6. DISCUSSION

This paper investigated a framework for considering the way that students and teachers might employ inferential reasoning when working with data to solve problems. The three aspects of informal inferential reasoning—generalizations, data as evidence, and probabilistic language—provided insight into the teaching and learning of statistical reasoning in an inquiry-rich problem-based environment.

6.1. GENERALIZATION BEYOND THE DATA

The first aspect of the framework, *generalization beyond the data*, provides the foundational inferential lens to move from describing the data at hand and shift towards the target of the inference. It considers not just the data that are being reasoned about, but the context in which the larger data set is likely situated. By looking beyond describing the data (through graphs or descriptive statistics) to consider the larger population or mechanism that the data represent, a shift in thinking can occur. This shift potentially moves the target of learning from statistical tools towards the problems for which statistical processes can provide powerful insights. Without considering generalization beyond the data, there is really no possibility for inference.

Natasha (Section 5.1) discussed three elements missing in her Year 5 unit that she considered important for providing opportunities for students to tap into inferential reasoning.

- *Pose a driving question*. An important element that Natasha named as missing from the investigation was a driving question. A driving question would have provided a clear direction when the investigation encountered obstacles or students became sidetracked. Without the driving question, Natasha said, "it wasn't really an investigation."
- *Include an engaging context*. In her unit on citizenship, Natasha felt that the topic of citizenship was too far situated from nine-year-olds' experiences and interests to engage their thinking beyond the data at hand. This reminds us of the underlying purpose of statistics to provide insight into phenomena under

investigation. She noted the importance of focusing attention on investigating the context, not the data in isolation.

- *Ensure sufficient complexity in the data*. The data that the students collected were primarily dichotomous and this lack of complexity did not trigger potentially innovative and insightful conclusions from the data collected. Without the opportunity to develop perceptive interpretations, the focus turned to the more mundane tasks of drawing graphs and reciting outcomes of data compilations. Natasha summarized it well by saying, “there was not really much that we could do with it … [and] it didn’t really allow them to do any higher-order thinking.”

Generalizations also played a key role in supporting students’ thinking in Carla and Kaye’s Year 4 unit on healthy eating. Because students often wanted to consider only individual data points or report on discrete values, the focus on creating generalizations (although not always beyond the data) helped students develop a more aggregate perspective in interpreting their data. Carla was able to use the idea of generalizations to support her students in this move to aggregate thinking. Elise’s unit on handspans likewise provided interesting insight into her students’ thinking about making generalizations beyond their data. When the class focused on using their data to begin making predictions about other classes (moving beyond their own data), this gave the teacher insightful evidence of her students’ thinking about data.

6.2. DATA AS EVIDENCE

Like the principle of generalization beyond the data, this principle supports the idea of focusing on the context under investigation rather than on investigating the data itself. For example, in Elise’s Year 3 class, they initially seemed to assume that the distribution of students’ handspans in a neighboring class would be just like (or even identical to) their own class, including the same smallest handspan, largest handspan, most common handspan, and even gaps (concluding no one in the neighboring class would have a handspan of 14 cm because no one in their own class did). In collecting and organizing the data from the neighboring class, students began to appreciate ways in which the variability in the distribution both differed from, and showed similar patterns to, their own data. Finally, in attempting to draw conclusions about the handspan distribution from a third class, students initially concluded that they could not describe that distribution at all because they did not have the data. In pushing them to consider how they might make and quantify predictions about the third class, students began to see the data they had as evidence for making these predictions.

In many ways, Carla’s focus with her class on making generalizations from the data supported them in seeing the data as evidence for their conclusions, a connection that students often find difficult (Hancock et al, 1992; Marshall, Makar, & Kazak, 2002). Natasha’s class, in turning their attention to graphing skills, never got back to the problem they were investigating that would have allowed them to make the connection between the data they collected and its potential as evidence for drawing inferences. The use of data as evidence is a key principle of informal inference that reminds learners of (1) the purpose of collecting and analyzing data; and (2) the importance of focusing on the problem and process of statistics in inquiry rather than just a data set as an isolated artifact. This concept of data may also help to curb students’ and teachers’ tendencies to focus on unproductive aspects of data (Pfannkuch et al, 2004).

6.3. PROBABILISTIC LANGUAGE

Finally, the third principle of informal statistical inference is the use of probabilistic language to articulate uncertainty and level of confidence in making predictions. The use of probabilistic language as a critical aspect of informal inference was most apparent in Elise's Year 3 class (age 7) when they were using the data they had collected on handspans to make predictions about the distribution of handspans in a neighboring classroom for which they had not collected data. Once they made the connection between using their own data as evidence to make predictions, their language changed to include notions of uncertainty and level of confidence. For example, students incorporated phrases like 'might be 13 or 11 [cm]' for the smallest handspan or 'around 15 to 16 [cm]' for the most common handspan. Additionally, students broadened their predictions from a single point to a range of values to articulate their uncertainty and also to improve the level of confidence in their prediction. The language of uncertainty may have also allowed students to take the risk in making their predictions without worrying about possibly being 'wrong.' When you consider the difficulty that even university students have in moving away from making absolutist-type conclusions that communicate a more deterministic perspective of inferences from data (Abelson, 1995), the ability of these young students to articulate some uncertainty in making their predictions is very encouraging.

7. CONCLUSION

Informal inferential reasoning (IIR) has been highlighted in a number of studies as a potential pathway for deepening learners' understanding of statistical processes and outcomes (see Ben-Zvi, 2006; Pfannkuch, 2006; Rubin et al., 2006; and research reported in Ainley & Pratt, 2007 and Pratt & Ainley, 2008). In addition, IIR may provide new opportunities to infuse powerful statistical concepts very early in the school curriculum (Ben-Zvi & Sharett-Amir, 2005) and return the focus of statistics to a tool for insight into understanding problems rather than only a collection of graphs, calculations, and procedures (Sorto, 2006). This paper presents a potential framework for better understanding key principles of informal inferential reasoning. By focusing on inference as the process of making probabilistic generalizations from data, the framework can be used to support teachers in understanding the importance of working with students to think beyond the data at hand, towards using that data as evidence for making predictions about a larger process or population. Also important is to articulate predictions with probabilistic rather than deterministic language in order to communicate both the uncertainty and the level of confidence of a prediction. The principles that underlie the framework further have the potential to both help students make better connections between the data collected and the problem under investigation, and to help deter the overly rigid stance that often accompanies statistical conclusions. The framework is also potentially useful to support the research community in "grappling with the conceptual building blocks for informal inferential reasoning" (Pfannkuch, 2006, p. 1) and to provide directions for further research on specific elements of informal statistical inference.

ACKNOWLEDGEMENTS

The research and writing of this paper was funded by The University of Queensland, Education Queensland, and the Australian Research Council (LP0776703). The authors wish to thank Jim Hammerman, Cliff Konold, and Arthur Bakker for discussions that were critical to early thinking on this topic, the teachers in the project who were so generous in contributing their time and insights, SRTL colleagues who encouraged, critiqued and discussed the use of the model with us at SRTL-5 and ICOTS-6, and the anonymous reviewers who suggested many improvements to the paper.

REFERENCES

Abelson, R. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ainley, J., & Pratt, D. (Eds.) (2007). *Reasoning about Informal Inferential Statistical Reasoning: A collection of current research studies*. *Proceedings of the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5)*, University of Warwick, UK.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Utrecht, the Netherlands: CD Beta Press.

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*. Dordrecht, the Netherlands: Kluwer Academic Publisher.

Ben-Zvi, D. (2006). Using *Tinkerplots* to scaffold students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/2D1_BENZ.pdf]

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distributions: A collection of recent research studies. Proceedings of the Fourth International Research Forum for Statistical Reasoning, Thinking and Literacy (SRTL-4), University of Auckland (New Zealand)*. Brisbane: University of Queensland.

Cobb, P. (1999). Individual and collective mathematical 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.

Curriculum Corporation (2006). *National Statements of Learning in Mathematics*. Melbourne: Author.

Dewey, J. (1910/1997). *How we think*. Mineola, NY: Dover.

Erickson, T. (2006). Using simulation to learn about inference. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/7G2_ERIC.pdf]

Finzer, W. (2001). *Fathom!* (Version 1.16) [Computer Software]. Emeryville, CA: KCP Technologies.

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

Gardner, H., & Hudson, I. (1999). University students' ability to apply statistical procedures. *Journal of Statistics Education, 7*(1).
[Online: www.amstat.org/publications/jse/secure/v7n1/gardner.cfm]

Gideon, J. (Ed.) (1996). *Teach-Stat activities: Statistics investigations for grades 3-6*. Parsippany, NJ: Dale Seymour Publications.

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

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2003). *Data seen through different lenses*. Unpublished manuscript.

Konold, C., & Miller, C. (2005). *Tinkerplots* (v. 1.0). Emeryville, CA: Key Curriculum Press.

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

Magnani, L. (2001). *Abduction, reason, and science: Processes of discovery and explanation*. New York: Kluwer Academic Publisher.

Makar, K., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 353-373). Dordrecht, the Netherlands: Kluwer Academic Publisher.

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

Makar, K., & Confrey, J. (2007). Moving the context of modeling to the forefront. In W. Blum, P. Galbraith, H-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI Study* (pp. 485-490). New York: Springer.

Makar, K., & McPhee, D. (in press). *Young children's explorations of average in an inquiry classroom*. Paper to be presented at the 32nd annual conference of the Mathematics Education Research Group of Australasia, Massey University, Wellington NZ, 5-9 July 2009.

Marshall, J., Makar, K., & Kazak, S. (2002). Young urban students' conceptions of data uses, representation, and analysis. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the Twenty-fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA24), Athens GA (USA), Vol. 3*, pp. 1293-1304. Columbus, OH: ERIC Clearinghouse, Mathematics and Environment Education.

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

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

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

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf]

Pfannkuch, M., Budgett, S., & Parsonage, R. (2004, July). *Comparison of data plots: Building a pedagogical framework*. Paper presented at the Tenth Meeting of the International Congress on Mathematics Education, Copenhagen, Denmark.

Pratt, D., & Ainley, J. (Eds.) (2008). Introducing the special issue on informal inference. *Statistics Education Research Journal, 7*(2), 3-4.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Pratt_Ainley.pdf]

Queensland Studies Authority (2006). *Essential learnings and standards (Year 3, 5, 7, and 9*. Spring Hill QLD: Author.

Reichardt, C. S., & Gollob, H. F. (1997). When confidence intervals should be used instead of statistical significance tests, and vice versa. In L. L. Harlow. S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (Vol. I, pp. 259-284). Mahwah, NJ: Lawrence Erlbaum Associates.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Rossman.pdf]

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/2D3_RUBI.pdf]

Sorto, M. A. (2006). Identifying content knowledge for teaching statistics. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/C130.pdf]

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

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

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf]

KATIE MAKAR
School of Education
The University of Queensland, QLD 4072
Australia

AN EMPIRICAL CONSIDERATION OF A BALANCED AMALGAMATION OF LEARNING STRATEGIES IN GRADUATE INTRODUCTORY STATISTICS CLASSES

BRANDON K. VAUGHN
The University of Texas at Austin
brandon.vaughn@mail.utexas.edu

ABSTRACT

This study considers the effectiveness of a "balanced amalgamated" approach to teaching graduate level introductory statistics. Although some research stresses replacing traditional lectures with more active learning methods, the approach of this study is to combine effective lecturing with active learning and team projects. The results of this study indicate that such a balanced amalgamated approach to learning not only improves student cognition of course material, but student morale as well. An instructional approach that combines mini-lectures with in-class active-learning activities appears to be a better approach than traditional lecturing alone for teaching graduate-level students.

Keywords: *Statistics education research; Graduate-level education; Active learning; Collaborative learning; Lecture-based learning; Team projects; Classroom instruction*

1. PURPOSE

Theresa is typical of most graduate students I meet in the first statistical course of their program. Sitting uneasy as we go around the class making introductions, Theresa starts out with basic information about herself, and then exclaims "To be honest, I have tried avoiding this class for as long I could! I'm scared to death of statistics!" She isn't alone.

Students often consider statistics as the "worst" course they take while in college (Hogg, 1991). For instructors, there is often a struggle with how best to reach students, to help them learn statistics, and to help them become practical consumers of the knowledge – especially when students enter statistics courses with negative self-images. As some of this negative imagery comes from the massive amounts of formulas students can face while in the course, one solution is to structure an introductory statistics course (possibly all statistical courses) around data analysis versus mathematical technique. Another solution is found in innovative instructional paradigms in which the traditional lecture, with students passively listening, is replaced with more hands-on activities.

Yet in graduate statistical education, the actual implementation of these different approaches into a classroom setting can be quite challenging and confusing. Many of these approaches involve unique learning opportunities which have not customarily been incorporated in traditional graduate-level statistics classes. Moreover, because most research has been conducted on undergraduate statistics classes (see next section), one might ask "Would the same techniques of active or cooperative learning *actually* work in a graduate introductory statistics class? Or possibly in more advanced classes?"

Statistics Education Research Journal, 8(1), 106-130, http://www.stat.auckland.ac.nz/serj

The purpose of this research is to consider alternative instructional methods in the teaching of introductory statistics to graduate students. Based upon personal informal surveying of graduate instructors, some feel uneasy about totally doing away with lecturing, especially when teaching graduate level courses. Indeed, for many graduate instructors I have interviewed, incorporation of innovative strategies is still foreign and they feel uneasy about abandoning a lecture-format class. Thus, I decided to embrace these teaching strategies and see whether, indeed, innovative strategies work in higher level statistics education. This particular study considers the possible linkage of research in undergraduate introductory statistics to a graduate-level introductory statistics class. Future research will consider the use of such teaching strategies for more advanced statistical classes.

The “balanced amalgamated” approach to teaching (with both traditional lecture and active learning) of interest in this study would allow many graduate instructors the opportunity to explore the advantages of active learning without the concern of losing the benefits of lecturing. Thus, the purpose of this research is to compare various amalgamated instructional models that involve lecturing and active/cooperative learning in graduate statistics classes. Comparisons were made with previous semesters in which these approaches were not undertaken. Specifically, this study attempted to answer the following research questions:

1. Can active or cooperative learning be successfully implemented and accepted in graduate introductory statistics classes? Can these strategies be combined with lectures to create a balanced amalgamated instructional approach?
2. Does more active student involvement help graduate students learn introductory statistics?
3. What benefit in affective and cognitive measures is seen by introducing active or cooperative learning along with lectures in graduate introductory statistics education? As males and females may gravitate toward different teaching approaches, do these benefits differ by gender?
4. Does a particular amalgamation work better than others with graduate students?

2. THEORETICAL FRAMEWORK

The question of how a student best learns statistics has been much considered in articles on statistics teaching (e.g., Chance, 2005; Gal & Garfield, 1997; Garfield, 1995; Lovett & Greenhouse, 2000), and has mainly focused on instructional content or methods. In terms of instructional content, many statisticians, including Bradstreet (1996) and Cobb (1991), are convinced that an introductory statistics course should emphasize data analysis over mathematical technique and concepts over formulas. Hogg (1991) stressed that statistics should not be presented as a mathematics course at all. Rather, the andragogy should emphasize statistical reasoning and thinking rather than algebraic precision.

Hogg (1991) further describes the problem of traditional instruction of statistics in terms of instructional design: Students are passive learners and do not directly come into contact with the many issues that occur in data collection and analysis. He suggests students would be better off generating their own data rather than utilizing a data set from a textbook or instructor. By working with projects involving their own data, students have opportunities to define problems, formulate hypotheses, design experiments, and have genuine data to analyze and summarize. In support of Hogg, Snee (1993) emphasized that because collecting data is the nucleus of statistical analysis, learning that centers on the analysis of real data that students collect connects them to the practicality of statistical thinking. Singer and Willett (1990) argued that real data should be the nucleus of all statistical education, although their emphasis was more on using available datasets and not on students collecting their own data.

Emphasizing this statistical content also leads to a more active involvement of students in the course, and the traditional lecture approach in teaching statistics has had much criticism in the last two decades of research (e.g., Delucchi, 2006; Garfield, 1993; Giraud, 1997; Moore, 1997). Garfield (1995) suggests that students learn best by constructing knowledge and becoming active participants in the learning process. Smith (1998) indicated that

> One way to help students develop their statistical reasoning is to incorporate active-learning strategies that allow students to supplement what they have heard and read about statistics by actually doing statistics -- designing studies, collecting data, analyzing their results, preparing written reports, and giving oral presentations. (¶ 3.7)

Steinhorst and Keeler (1995) is another great resource in this matter. In support of an active-learning approach, Bradstreet (1996) writes that "Learning is situated in activity. Students who use the tools of their education actively rather than just acquire them build an increasingly rich implicit understanding of the world in which they use the tools and of the tools themselves" (pp. 73-74). Thus in this study, "active learning" refers to any activities in which the student participates and learns in a non-passive way (e.g., simply answering questions from the teacher would not be considered "active learning" in this study).

There are a variety of ways in which to incorporate active learning and projects into instruction, in particular, personal collection of data. These might include some of the following: computer simulations (Garfield & delMas, 1991); laboratory-based courses (Bradstreet, 1996); in-class activities (Dietz, 1993; Gnanadesikan, Scheaffer, Watkins, & Witmer, 1997); a single three-week project (Hunter, 1977); or a course-long project (Chance, 1997; Fillebrown, 1994; Ledolter, 1995; Mackisack, 1994).

In regard to cooperative learning, many researchers have reported significant accomplishments from introducing cooperative learning experiences in introductory statistics classes (Dietz, 1993; Jones, 1991; Keeler & Steinhorst, 1995; Shaughnessy, 1977). Teams help encourage cooperative learning, develop team-working skills, and usually build substantial friendships (Smith, 1998). However, some of this research tends to limit such activities to external learning situations such as homework or studying. Although much research exists indicating the effectiveness of alternative teaching techniques, how a teacher should implement such teaching strategies is not always clear and should be a point of more research (Garfield, 1995). Johnson and Dasgupta (2006) found that undergraduate students predominately prefer non-traditional instructional styles. Yet, exactly how a class should be structured and which techniques work best with each other was not considered in this research and would seem to be of importance to instructors wishing to incorporate such styles. Jordan (2007) stressed that the implementation of such instructional styles is open to much interpretation. In fact, some research has indicated that the implementation of these techniques does not always happen. Cobb (1993) investigated the results of various NSF grants whose purpose was to significantly improve statistical instruction, and discovered that none of the grants involved team projects, nor cooperative learning situations. Bryce (2005) indicated that few textbooks have embraced these new ideas in statistical education.

Furthermore it is not clear which methods work better for different students. Students have differing learning styles. Ford and Chen (2002) showed that males and females performed differently under various teaching styles. Kolb (1984) provided a model of learning styles: concrete–abstract and reflective–active. The combinations of these styles indicate learners that are labeled as accommodators (concrete, active), divergers (concrete, reflective), assimilators (abstract, reflective), and convergers (abstract, active). Grasha (1996) suggested bipolar characteristics of learning styles: competitive vs. collaborative, dependent vs. independent, and participant vs. avoidant. Regardless of the label, research has shown that students vary in how they best learn. Thus a class structure that only emphasizes one learning style (e.g., lecture or active learning) might in fact disadvantage some students in the attempt to reach others. In terms of classroom dynamics and pedagogical styles, Grasha identified five teacher styles: expert, formal authority, personal model, delegator, and facilitator. One can see in these labels the full spectrum of classroom dynamics from pure lecture (expert) to pure active/collaborative learning (facilitator). Instruction that flows in and out of these different dynamics would in essence touch on the diversity of learning styles of students.

Lovett (2001) said that "a successful route to improving students transfer of statistical reasoning skill may rely heavily on integrating instructional and cognitive theory, while maintaining a link to the realities of the classroom" (p. 348). Some research has considered the effect of combining different instructional techniques to create an amalgamated approach in teaching statistics. Ward (2004) created an amalgamated class consisting of online and face-to-face classes and found little difference in student performance. Keeler and Steinhorst (1995) created an amalgamated class consisting of collaborative groups and mini-lectures. They showed an improvement in students' attitudes and grades when incorporating more active student involvement with lectures. Their research focused on undergraduate, introductory-level statistics classes. No research has been found that applies such techniques to graduate-level statistics classes.

In fact lecture-based approaches appear to still dominate graduate-level statistics classes. A preliminary study by this researcher which interviewed 14 graduate instructors from colleges and departments of Education, Business, and Statistics at four major universities in the United States found this tendency, and when asked about the possibility of utilizing different learning strategies in their higher-level statistics classes, the response is typically expressed somewhat like "Yes, that may work for an undergraduate statistics class, but it would never work in this class. This class is too high level and demands a lecture format be predominant or even exclusive." Although these instructors are familiar with the vast research on innovative learning strategies in statistical education, there appears to be a huge gap between knowledge and practice. For these instructors, whether the same results shown in undergraduate statistics education would apply to graduate students has been minimally investigated.

Bligh (2000) suggests that lectures do have their place in education, yet the problem lies with instructional strategies that have unrealistic expectations. For example, Bligh indicates that whereas lectures are good at imparting ideas, they are not as good at motivating students: "Use lectures to teach information. Do not rely on them to promote thought, change attitudes, or develop behavioral skills if you can help it" (p. 20). For Bligh, many critics of lecture-based approaches in instruction almost de-emphasize the role of lectures completely. Yet, the complexity of graduate-level statistics classes suggests that some form of lecture might be beneficial. It has been this researcher's own observation that a statistics class that revolves around total active learning does not provide students with enough security in statistical methodology, especially for more advanced statistics classes. Often students need to see things demonstrated before they can apply those techniques to real life. However, this does not exclude the possibility of incorporating many of the proven techniques that have been shown to make a difference in undergraduate education.

3. RESEARCH DESIGN

Due to various limitations in the study (detailed below), a quasi-experimental design was implemented for this study. Specifically, an "untreated control group design with dependent pretest and posttest samples" (Shadish, Cook, & Campbell, 2002, p.137) was used for this study. More details on the sampling, treatment, instruments, and statistical analysis used in this design are detailed in this section.

3.1. PARTICIPANTS

The data were collected from three graduate educational statistics classes conducted at a major university over a period of three consecutive semesters. The classes consisted of both master's and doctoral students (approximately equal ratio). Specially prepared notes by the author served as the textbook in all classes. All classes were exposed to the statistical package *R*. Although the class is offered by an Educational Psychology department, students from many different departments and colleges at the university take the course. These other colleges/departments included Social Work, Communications, Journalism, Psychology, Business, and so on. The first semester students received the traditional lecture, the second semester class added the active learning element, and the third semester added the team projects. The designs of these instructional approaches are discussed in the next section. The first semester class (lecture only) consisted of 59 students, the second semester class (lecture + active learning) consisted of 44 students, and the third semester class (lecture + active learning + team projects) consisted of 47 students. Some demographic characteristics of the classes over varying semesters were considered (specifically major, gender, race, and number of previous statistics courses), and similar frequency patterns were found between all classes. Age was obviously a bit higher on average for doctoral students than master's. The average age for all classes was 25, with each class having ages ranging from 21 to 35.

3.2. INSTRUMENTS

Various affective measures were taken from a series of items developed and validated by the university at which the research took place. A bundle of items dealing with course content and instructional assessment were used in this study. This instrument consisted of eight items measured on a 5-point scale ranging from "poor" to "excellent." Only four of the eight items were used in this study as they dealt directly with the subject of this study. The items included such evaluations as "communication of ideas and information" and "stimulation of interest in the course." These affective measures were given before the final cognitive measure which is described next. Because these items were not used in a scalar sense, the psychometric properties of the scale of items are not discussed.

The cognitive measure was taken from a master list of questions which was developed by the researcher over a 10-year period. The master list contains 100 multiple-choice questions, each with three distracters and one correct answer, covering topics in regression and hypothesis testing. A second set of items, covering more elementary concepts such as graphs and descriptive statistics, was used in this study as a covariate. The common cognitive measure for all three classes (items dealing with regression and hypothesis tests) was used as a measure of instructional effectiveness in this study.

Psychometric issues, such as validity and reliability, have been considered for these items. Face validity was conducted with other statistics instructors deemed exemplary. Construct validity has been carried out in various analyses over the course of development, as well as reliability assessment and item analysis. As a result, items have been removed, adjusted, or improved upon. This final bundle of items has been shown to have sound psychometric properties of consistency (coefficient alphas at least 0.80) and factor loadings consistent with the construct that the item is measuring. Each assessment per class consisted of 33 randomly selected questions from this bank of questions. Students are not allowed to keep their exam thus providing some security of the questions from semester to semester. The grade from this assessment was part of other grades used to determine the overall course grade for the class. The grading scheme for the "lecture only" and "lecture + active learning" group was identical. The grading scheme for the "lecture + active learning + team projects" class included a component on student presentation and papers as part of their grade.

A final instrument concerning team dynamics was created to measure students' perceptions of team projects. Students in the team-project group were asked to evaluate their experience with team projects during that semester. Forty-eight items were developed, and forty of these items reflected an attitudinal measure on a 5-point Likert scale. The other questions dealt with opinions on such matters as preferred group size and on locus of control in terms of project assignment. This instrument was not developed as a means of assessment but rather as a beginning explanatory consideration of the statistical results found in this study. Thus, the psychometric properties of validity (beyond face validity) were not as stringently considered as the prior measures. The coefficient alpha for the forty Likert items was 0.98.

3.3. STATISTICAL ANALYSIS

The dependent variables in this study focused on both affective and cognitive measures. The researcher felt that student attitudes toward instruction would be just as vital as cognitive measures in evaluating instructional design effectiveness. Thus, student attitudes towards the different aspects of the class and instructional approach were considered. Due to limited student identification in the baseline (lecture only) group, a multivariate analysis with both measures was not possible. Chi-square analyses were conducted to detect significant patterns of response for the affective measures. A two-factor ANCOVA design was implemented for the cognitive data. Independent variables consisted of type of instruction (lecture only, lecture + active learning, lecture + active learning + team projects) and gender (control variable). The covariate was the prior cognitive assessment of knowledge over basic statistical knowledge (graphs, descriptive statistics, and other such introductory statistics topics).

4. INSTRUCTIONAL DESIGN METHODS

Using the labels of Grasha (1996), instruction can be designed in many different ways as shown in Figure 1. This visualizes each approach as a river. The relationship of each approach can be seen by either independent, non-intersecting rivers (part a), or connected tributaries which feed into a common collection (part b). If we think of instructional approaches as such, we can consider instruction as consisting of independent approaches (e.g., lecture only (expert) or only constructivist (facilitator)), or as dependent approaches which draw upon each other for a combined approach. Whereas most might consider the first picture as unrealistically extreme of actual educational practices, the reality is that

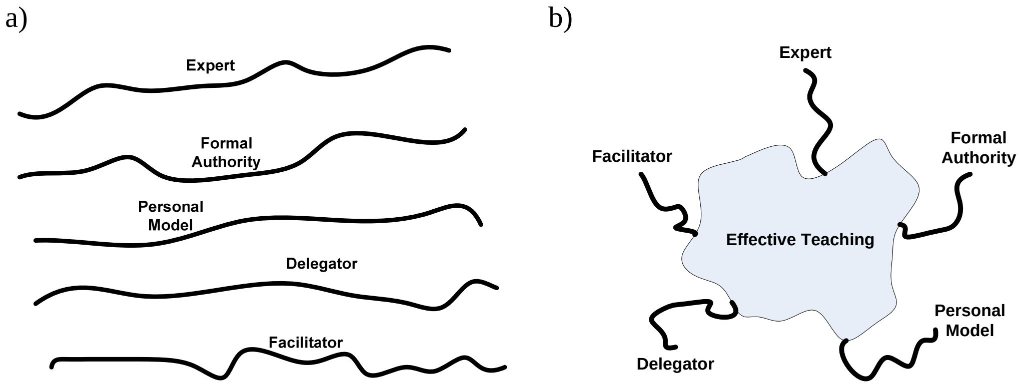


Figure 1. Independent (a) versus dependent (b) instructional approaches

some tributaries in part b might be trivial or completely forsaken (Figure 2a). Or, some may even design a particular class as consisting of only one approach (e.g., Tuesday’s class is lecture, Thursday’s class is hands-on application) as shown in Figure 2b. This study designed instruction that sought to favor each tributary equally within the same class. The approach as shown in Figure 1b was adopted as the instructional design of this study to test the effectiveness of balanced amalgamated approaches in both the understanding and attitudes of students. The details for each approach are presented, followed by a discussion of the statistical analysis of the relative effectiveness of the instructional approaches.

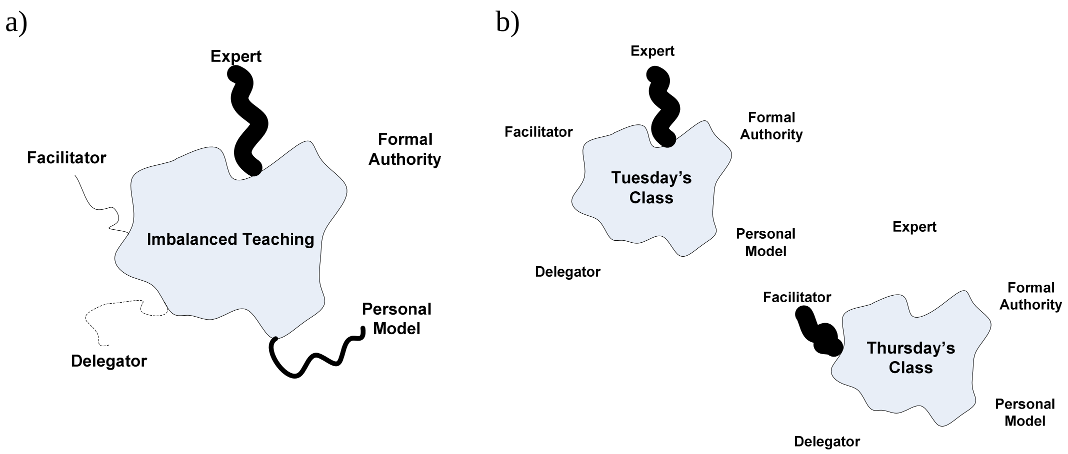


Figure 2. Examples of imbalance in instructional approaches

For this study, three instructional methods are compared: traditional lecture, mini-lectures with in-class activities, and mini-lectures with both in-class activities and team projects. The classes were all taught by the same instructor in different semesters. The instructor had ten years of experience teaching graduate level statistics courses.

4.1. "LECTURE ONLY" CLASS

The class that comprised only lecture-based instruction contained no active component or teamwork. Each class consisted of 45 minutes of lecture, along with the time for students to ask questions. Sample problems were demonstrated in class, yet the students were never involved in either the process or data collection (i.e., they were passive learners).

4.2. "LECTURE + ACTIVE LEARNING" CLASS

For the classes with in-class activities, the instructor still provided instruction, but only in short segments. A general setup for each class meeting involved small lectures with short instructional elements (5 to 10 minutes) followed by active application of the knowledge (5 to 15 minutes). Each activity was discussed afterward in class, before the next small lecture began. Thus, for a 50 minute class, a typical meeting consisted of three mini-lectures, along with three direct applications of the knowledge learned (typically in teams). This is illustrated in Figure 3.

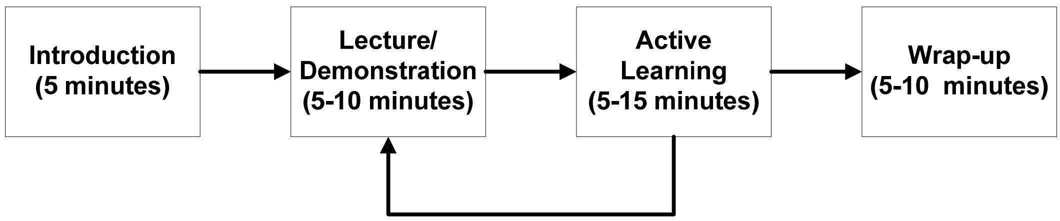


Figure 3. Class structure with active learning emphasis

This approach was modeled after one of the few textbooks that incorporates active learning within the text (Aliaga & Gunderson, 2003). From this perspective, a mixture of traditional lecture and active learning is incorporated. An example of this approach in a worksheet is presented in Appendix A. Immediately after learning new concepts or ideas, students are presented with problems (often real situations) to apply that knowledge in individual activities or as teams. For example, after learning about stratified sampling, students then split up into teams and take a stratified sample of their particular team. This allows for more immediate feedback as to how a student is assimilating the information. In addition, this approach allows a student to see real-life examples on their own. Because of the nature of this approach, however, students do not have the time to devote to more advanced statistical methodology such as sampling designs. Thus, this research also considers the addition of team projects to this paradigm.

Worksheets were incorporated and made available to students via a course webpage. The worksheets contained material that the instructor first introduced in the "mini" lecture. Students completed the worksheets after each mini lecture either individually or in groups (depending on the activity of the worksheet) during class. The activities were designed so that the amount of individual versus group-related activities approximated a 50/50 proportion over the entire semester. However, on a given day of instruction, this proportion was not always 50/50. For half of the group activities, students were allowed to work with their neighboring students (which was helpful in larger classes). Predominately, these groupings were fixed for the entire semester. The other half of group activities were done so that students congregated in new ways around the classroom. This was sometimes accomplished by instructor assignment or students randomly assigning themselves to a group (using random number techniques and areas assigned number labels), or in just pseudo-randomly walking around the class to find a group (which can be too time consuming).

4.3. "LECTURE + ACTIVE LEARNING + TEAM PROJECTS" CLASS

The last mode of instruction involved a class that combined the active-learning element previously mentioned with team projects. Students were split up into groups of four to five students as chosen by the students, and assigned a project from a list of projects the instructor had assembled. Teams were allowed to modify projects, with instructor permission. However, only one team modified a project. The modification only changed the population of the study for the particular project. In order to encourage student involvement, many of the projects were devised from more "sensational" ideas that made the presentation of the results interesting to students. For example, one of the projects in the "regression" unit focused on surveying students to see whether there was any correlation between GPA and the number of alcoholic drinks consumed per week. Although such a question is highly dramatic, the results provide a memorable discussion into the ideas of correlation, and the problem of causation. From there, a short discussion of response bias is often appropriate. Students were required to stay in the same team the entire semester. The in-class group activities utilized teams whose compositions differed from those of the team projects. Inevitably, students from the same team might be members of the same in-class groups, yet no team group was contained within the groups for in-class activities.

The initial focus of this research centered on projects that involved educational settings. This, however, proved more difficult than earlier imagined. In particular, studies on young children or teenagers often require special permission and take much time to setup. Projects needed to be devised that did not demand too much of a student's time. Although educational data sets are available, the purpose of this research was to involve the students in all aspects of research – from data collection to analysis. The solution in this research was to create projects that tended to be behavioral or sociological in nature. Although some projects did involve student attributes (and thus were educational in nature), most projects were forced to be generally behavioral. Four projects were given during the semester which focused on application of various topics discussed in class:

1. Data (sampling, bias, etc.)
2. Descriptive statistics (central tendency, variability, boxplots, etc.)
3. Hypothesis testing (*t* tests, ANOVA, etc.)
4. Regression (correlation, simple linear regression, multiple regression, etc.).

A listing of sample projects used is given in Appendix B. Some of these projects were modified from Smith (1998).

On a set date, students were asked to turn in a written summary of results (in APA style) and to give a presentation. As mentioned previously, research has shown that summarizing and presenting statistical results is a more demanding learning taxonomy level than simply analyzing the numbers. Because of the size of the class, presentations were randomly selected and only those teams randomly selected had to present their findings in class. This allowed for half of the class period to still involve learning units, then the last half of class was used for project presentations. The teams were asked to limit their presentations to ten minutes.

5. RESULTS

5.1. AFFECTIVE RESULTS

Comparisons were made for the attitudes of students in the various instructional treatments. First, students were asked to rate the effectiveness of the instruction. A summary of the results is presented in Figure 4. Because many of the frequencies for the "poor" and "fair" responses were zero for the classes with active learning, the responses for "poor," "fair," and "good" were combined into an "at most good" category so that the assumptions of the chi-square test would be valid. Thus, the chi-square tested the independence of instructional approach to ratings of excellence.

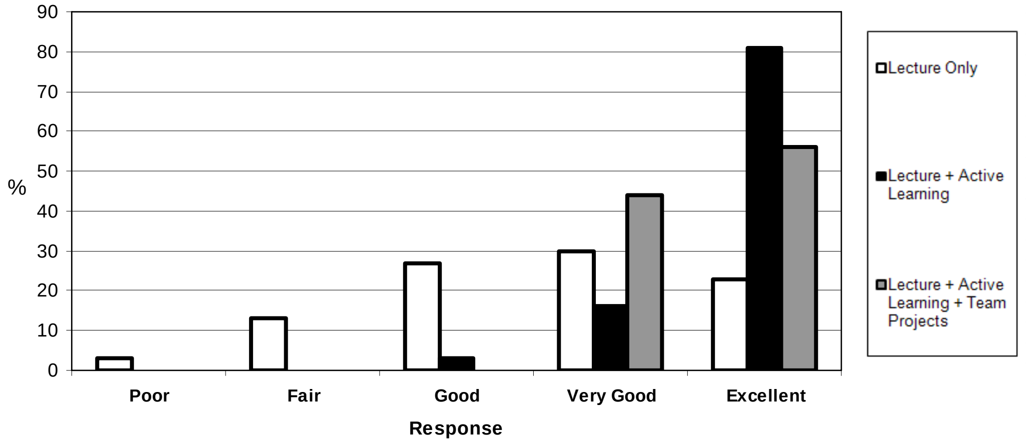


Figure 4. Communication of ideas and information

There is a sizeable difference that the active-learning component made for students' assessment of instructional effectiveness ($\chi^2(4) = 35.2$, p-value < 0.01). Whereas only 23% of students in the lecture-only class viewed the communication of ideas and information as being excellent, 81% of students in the amalgamated lecture/active learning class rated this communication as excellent. Interestingly, this high ranking did not hold when team projects were introduced into the course. For this class, the percentage of "Excellent" responses dropped to 56%. Although, the extreme satisfaction with the course ideas did not remain when team projects were introduced, it is important to note that there is a drastic positive shift in student attitude with either approach as compared to the traditional lecture approach.

Next, students were asked to rate the stimulation of interest in the course they experienced. This is an important factor because motivation in statistics classes has often been noted as a major problem by instructors. These results are summarized in Figure 5.

The results of a chi-square test were significant ($\chi^2(4) = 18.9$, p-value < 0.01). Although more students felt the communication of ideas was excellent for either the activity-based class or team-based, there was a slight drop in percentage of students who felt their interest in the course was stimulated. However, it is important to realize the great increase in student interest by incorporating active-learning components or active learning with team-based projects. For students in the activity-based class, 74% of them gave an "Excellent" rating for stimulation, as compared to just 23% in the traditional lecture class. A slightly lower, but still positive, rating (63%) was recorded for students in the team-based instructional classes.

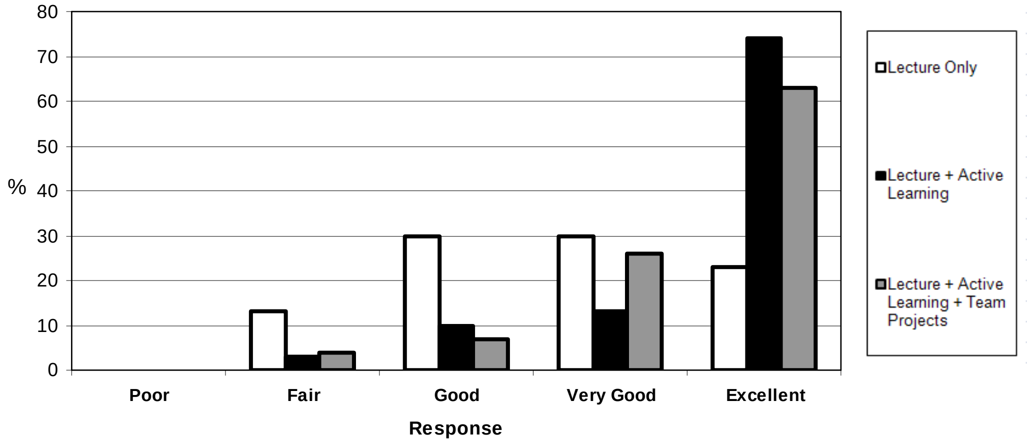


Figure 5. Stimulation of interest in the course

Students were also asked to rate how well the class structure facilitated their learning. This question was more specific than the one presented in Figure 4. The question represented in Figure 4 considered an overall communication of the class ideas. For the question on facilitation, students were asked to rate how well the instructional approach helped in learning. The results are presented in Figure 6.

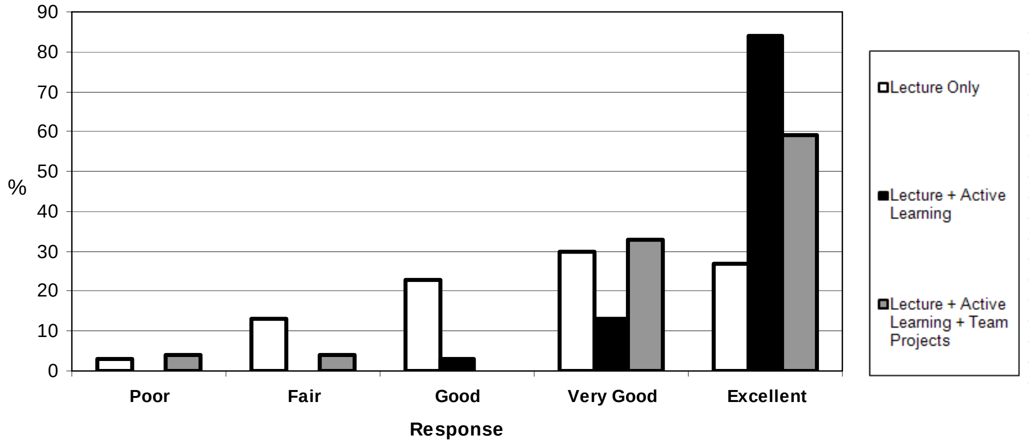


Figure 6. Facilitation of learning

As before, a much higher percentage of students in the activity-based class (84%) and team-based class (59%) rated the instructional approach as "Excellent" as compared to the traditional lecture-only approach (27%) ($\chi^2(4) = 26.1$, p-value < 0.01). Students seemed to feel much more confident with the activity-based learning (alone) than either of the other methods, although the activity + team project-based approach was favored by students over the lecture-only as well.

Finally, the overall rating of the instructor was considered. These results are presented in Figure 7. The percentage of students rating the instructor as "Excellent" rose for each teaching approach: 37% (lecture only), 90% (lecture + active learning), and 70% (lecture + active learning + team projects) ($\chi^2(4) = 24.2$, p-value < 0.01). This is expected because the question addresses more of rating of a person than course materials or instruction. Thus other confounding variables (such as friendliness of teacher) may account for the percentage increase across all types of instruction. Yet the same pattern remains in which the “Lecture + Active Learning” class is rated best by students, followed by the “Lecture + Active Learning + Team Projects” class, and lastly the “Lecture Only” class.

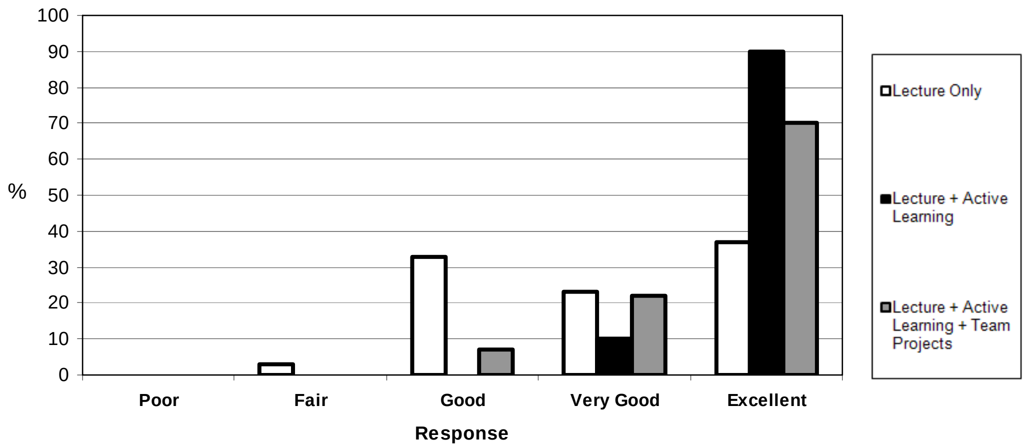


Figure 7. Overall assessment of instructor

5.2. COGNITIVE RESULTS

Assessment of student knowledge of regression and hypothesis tests was considered as the cognitive measure of the effectiveness of each instructional method. This final measure consisted of 33 questions randomly chosen from a validated and reliable bank of questions about regression and hypothesis testing ideas. A prior cognitive measure was also used: another 33 questions randomly selected from a validated and reliable bank of questions about basic statistical knowledge (graphs, descriptive statistics, and so on). Table 1 summarizes the cognitive measures for each class.

Table 1. Descriptive statistics for prior and final cognitive measures

		Prior Cognitive Measure		Final Cognitive Measure	
	n	Mean	SD	Mean	SD
Lecture Only	59	85.62	10.52	84.84	8.92
Lecture + Active Learning	44	88.64	8.81	91.55	6.03
Lecture + Active Learning + Project	47	83.51	9.10	89.20	9.90
Total	150	85.84	9.75	88.17	8.93

An initial check of assumptions revealed possible problems with outliers and normality (Figure 8). Because the cognitive outcomes were measured in such a way as to be considered a proportion, the cognitive measures were transformed using an arcsin transformation as suggested by Neter, Kutner, Nachtsheim, and Wasserman (1996). No outliers remained after transformation. Using these transformed variables, the assumptions of normality of sampling distributions, linearity, homogeneity of variance, homogeneity of regression, and reliability of the covariate were found to be satisfactory.

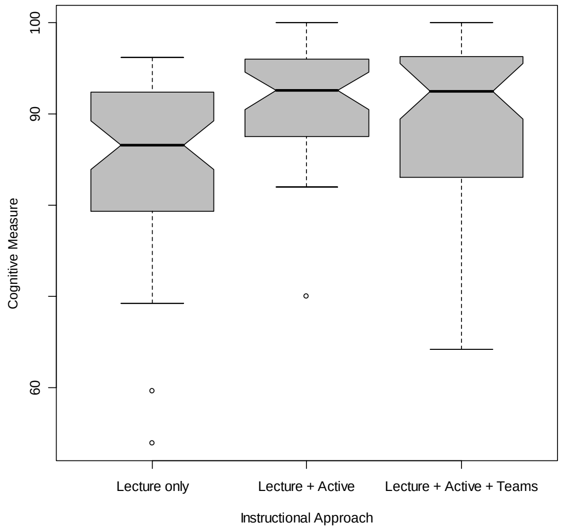


Figure 8. Boxplots of final cognitive measure by instructional methods

After adjusting for prior cognitive abilities, final cognitive assessment varied significantly with type of instruction ($F(2, 143) = 18.925$, p-value < 0.001). The ANCOVA results are summarized in Table 2. The strength of the relationship between adjusted cognitive final measure and type of instruction was moderate with partial $\hat{\eta}^2 = 0.209$. The adjusted non-transformed marginal means, as shown in Table 3, indicate that the highest final cognitive measure was for the class that had lecture, active learning, and team projects. The lowest final cognitive measure was for the class with only the traditional lecture. A post-hoc test revealed significant differences between the "lecture + active learning" approach and the "traditional lecture," as well as between the "lecture + active learning + team project" approach and the "traditional lecture." There were no significant differences between the "lecture + active learning" and "lecture + active learning + team project" approaches. Based on these results, graduate students who are more actively involved in their learning (whether through in-class activities or in-class activities and team projects) have significant gains in cognitive understanding, on average.

No statistically significant differences between males and females were found, nor was there a significant interaction effect between instruction type and gender. For gender, partial $\hat{\eta}^2 < 0.001$ indicated an almost non-existent relationship between gender and final cognitive measure. For the interaction, partial $\hat{\eta}^2 = 0.01$ indicated a weak relationship between adjusted final cognitive measure and combinations of gender/instructional method.

Table 2. Analysis of covariance of final cognitive measure

Source of Variance	Adjusted SS	df	MS	*F*
Type of instruction	1.542	2	0.771	18.925*
Gender	0.001	1	0.001	0.018
Interaction	0.059	2	0.029	0.721
Covariate (adjusted for all effects)				
Prior cognitive measure	4.589	1	4.589	112.634*
Error	5.826	143	0.041	

*p-value < 0.01

Table 3. Adjusted and unadjusted mean cognitive measure for three designs of instruction (untransformed)

Type of Instruction	Adjusted Mean	Unadjusted Mean
Lecture Only	84.92	84.84
Lecture + Active Learning	89.58	91.55
Lecture + Active Learning + Team Projects	90.91	89.20

5.3. POST-HOC QUESTIONNAIRE OF TEAM DYNAMICS

The addition of team projects did not significantly improve student attitudes or the cognitive measure as compared to the class that only had in-class activities. This is surprising as students were often heard discussing the results with other students, and the researcher felt that peer instruction was occurring more often than in previous semesters without team projects. To understand this phenomenon better, students who were exposed to both active learning and team projects were asked to describe their learning experiences with the team projects. These results differ from earlier questions in that the goal of these questions was to isolate student attitudes specifically on team projects and not on the active-learning component of the course. The results for the overall experience with team projects are presented in Table 4. Overall, 87.3% of the class reported having a positive experience by inclusion of team projects. A small minority (4.2%) had negative experiences with team projects. In reviewing the reasons why, it was noted that one of these individuals did not feel respected by the other team members, and another student felt frustrated by the lack of work from some team members.

Table 4. "What was your experience of team projects in this class?"

	Frequency	Percent
Very Negative	1	2.1
Negative	1	2.1
Undecided	4	8.5
Positive	28	59.6
Very Positive	13	27.7

Students were also asked how the team projects helped in the learning of the material. The summary of these results is presented in Table 5. The inclusion of team projects was meant to make the statistical content more meaningful to students. In this case, 82.9% of students felt that the projects helped them understand the material better. Interestingly, this percentage was less than the percentage for the overall experience. Slightly more students had a positive experience with team projects, with more students not feeling that the projects helped them. Yet, overall, inclusion of projects did seem to help.

Table 5. "Did team projects help you understand course materials?"

	Frequency	Percent
Strongly Disagree	3	6.4
Disagree	5	10.6
Undecided	0	0.0
Agree	27	57.4
Strongly Agree	12	25.5

Teams were set at four to five people at the beginning of the semester. Students were asked what group size they preferred working in. These results are presented in Table 6. Nearly half the subjects indicated that a three-member group would have been preferred over a larger group. The optimal group size for most students was three to four people, with most students feeling better about a 3 member group.

Table 6. "What team size do you prefer to work in?"

	Frequency	Percent
2 person team	7	14.9
3 person team	23	48.9
4 person team	16	34.0
5 person team	1	2.1

One potential problem with working in teams is that students do not always participate. Students in this sample indicated that this was sometimes a problem in their team projects (Table 7). Over one-third of all responses (36.1%) indicated that some teammates consistently failed to produce the results required of them in a timely manner. Obviously, this appears to be a problem with any team activities and should be addressed.

Table 7. "Did any team members fail to produce results in a timely manner?"

	Frequency	Percent
Don't Know	1	2.1
Never	19	40.4
Rarely	10	21.3
Sometimes	9	19.1
Often	8	17.0

As a final measure of the "lecture + active learning + team project" treatment, students were asked if more emphasis should be given to the projects in class (e.g., write-up of discussions, examples from the literature, more computer emphasis). This question was needed to ascertain whether the balance of lecture, active learning, and cooperative teams was good. These results are presented in Table 8. The results were mixed. More than one-third of the class (42.5%) felt that more emphasis should be placed on the cooperative learning element.

Table 8. "Should team project be emphasized more in class?"

	Frequency	Percent
Strongly Disagree	2	4.3
Disagree	9	19.1
Undecided	16	34.0
Agree	16	34.0
Strongly Agree	4	8.5

6. DISCUSSION

6.1. CONCLUSIONS

The inclusion of student activity (whether in-class activities or in-class activities combined with team projects) was seen to have a major impact upon graduate students learning of statistics. This effect was not only seen in specific cognitive measures, but also in the students' perceptions and attitudes toward the course. By doing hands-on activities, whether in class or outside of class in team projects, students appeared better able to make real connections to the knowledge they were learning. This, in turn, should provide more motivation to learn.

In fact, by getting away from a pure lecture approach, significant gains in cognition of course content and attitudes appear to be possible. Similar to Keeler and Steinhorst (1995), this study included a typical lecture in addition to innovative instructional models. Instead of abandoning a method that has worked somewhat in the past (especially in graduate-level statistics education), this research has attempted to incorporate other proven instructional techniques to create a "balanced amalgamated" approach to teaching statistics to graduate students. In addition to Keeler and Steinhorst's approach, this study also considered varying methods of active learning in teaching along with a team-based cooperative learning model. And though the results of this study are not conclusive about the use of team projects, more active involvement by students does appear to be the better way of approaching statistical instruction for graduate students.

The optimal structure of a class that emphasizes more student activity is unclear. In this study, the biggest impact came from combining mini lectures with in-class activities. Adding team projects diminished this impact somewhat. Yet this may be due to many things other than the non-effectiveness of team projects. The addition of team projects (along with active in-class activities) did provide significant increases in cognitive measures over a traditional lecture approach. The diminishing effect of teams versus a pure in-class activity approach seems to be influenced heavily by group dynamics. Some of the groups in the study had personality clashes. One student actually asked to be moved to another team because of personality issues. In addition, almost one-third of the class mentioned the timeliness of other team members as being a detrimental factor. Finally, team projects do require more work. This may have an impact on team effectiveness.

6.2. LIMITATIONS

One of the major limitations of this study is the presence of confounding effects and potential sampling bias. The same teacher taught all three classes, and students were not randomly selected to participate nor randomly assigned among the three treatments. Though prior knowledge was included as a covariate, higher scores on the pretest could indicate a more positive disposition of students entering the course affecting their attitudes and learning gains. Replication of this study with other teachers would be beneficial, as well as with departments other than Education. This research always assumed that a lecture approach was part of any class. No teaching paradigms were considered where a class setting whould be purely active or cooperative. The use of team projects along with mini lectures with no active-learning component, the use of only active learning, or the use of team projects as the sole instructional approach was not considered. A course that spends more time on team projects may see a bigger team project effect. Such a class would be a "practitioner" course in that the total emphasis is on application of statistics and not so much theoretical knowledge (as provided in a lecture). This type of course was not considered in this study. However, to know the impact of these instructional elements fully, studies should be conducted that look at an approach using these methods in teaching statistics.

6.3. SUMMARY

This research attempted to answer a central question: Does a structure incorporating more active student involvement help graduate students learn statistics? The answer based on this study does appear to be “yes.” Still, the question of what combinations of instructional design to use remains unanswered. Though incorporating both active learning and team projects with lectures showed more positive effects than a lecture only approach, there was no statistical difference between that approach and the “lecture + active learning” approach. Furthermore, from this instructor’s experience, student attitudes are often better when the instructor was more in control of the active component. Perhaps the inclusion of team projects shifts the locus of control too far for graduate students? Such consideration is worthy of future study. Regardless, this study suggests that any active component in a graduate-level statistics class makes it better, both affectively and cognitively.

6.4. IMPLEMENTATION STRATEGIES

My own experience with incorporating active learning in teaching has always been successful. When I incorporated the active component into the class, I could see an immediate effect upon students. In previous semesters I had tried a more constructivist approach to teaching and felt that the class was not successful. Based on that experience, the inclusion of some form of lecture component seemed necessary. This amalgamated approach took away many of the fears I had in incorporating the ideas. By having small active-learning components in the middle of lecture components, I could see whether students really understood what I had just shown them. Also, this builds confidence in students. Theresa, the student mentioned earlier, was one such example as she was in the “lecture + active learning” group. By the end of the second week, she expressed to me how she was really enjoying the class and how it was not nearly as bad as she had originally thought. I cannot imagine teaching without some form of active learning now. I have applied this same teaching paradigm to more advanced statistical classes with similar results of improved morale, attention, and assimilation of information.

The inclusion of team projects posed mixed results, however. First, there was more of a struggle with how to allocate class time for such matters. Although including active learning would seem to present the same challenge, team projects always presented the major challenge in time. For me, team dynamics was a factor that grew tiring to oversee. I still see the benefits in team projects, yet it seems that I see little difference in performance by leaving it out and using active components in class (which still involve working with others). This is a matter of further investigation for me. Of particular interest would be studying the different outlooks and motivations for undergraduate versus graduate students in regard to team projects and team learning.

To successfully implement such a change in a classroom requires a radical shift in perspective for most instructors. Thus, I offer the following suggestions that I have learned over the past 10 years of trying to incorporate such designs in my own classes:

1. *Review learning and cognition materials* – Although direct research (like this study) is obviously important to consider, I have found that there is a wealth of information about such matters in the learning and cognition domain. One would also greatly benefit by finding someone who specializes in this field and discussing matters with them. Though I have not always incorporated their ideas or techniques, I have often come away from such conversations with clarity of thought and purpose. The

statistical education domain would greatly benefit if there were more collaborations in research with learning/cognition researchers, not to mention how beneficial textbooks might be if such collaborations extended into authorship. As mentioned previously, most current graduate statistical textbooks ignore these new teaching methods altogether.

2. *Study effective teaching models* – There is a wealth of ideas often right on our own campuses about effective teaching models. I occasionally ask other faculty members to let me sit in on their class for my own personal benefit. In my younger years, I would too often want to emulate the habits and mannerisms of those I admired or respected. I remember the first time I saw Jaime Escalante in the *Stand and Deliver* famed movie. After seeing him in action, I wondered if I should dress up in costumes and use props in my classes, or have my classes do chants in the classroom. Yet, in time, I have seen that I need to become a unique expression of these effective techniques I have learned. I believe this is important because at times I have rejected whole approaches in teaching because I could not see myself implementing them in the same manner. For example, I have seen instructors using a "game" for an in-class activity, and had inner struggles with incorporating this particular expression: "That may work for an undergraduate class, but this is a graduate class on Bayesian analysis! Won't this seem 'childish' to the students? Even if I were to try it, how can I come up with a game in this class?" What I have learned is that I may not always prefer or use a particular expression, but I can still strive to incorporate the spirit of the expression (e.g., in this case by creating active components that are interesting and fun to students).
3. *Start small* – If you are new to active-based learning (whether in-class activities or team projects), it might be beneficial to incorporate this new learning style in small ways until you are comfortable with the approach. Consider the worksheet in Appendix A. One suggested approach is to design such a worksheet for a particular class lecture, and have students work on it for a few minutes in the middle or end of class. The key to incorporating such active-based learning is to keep the activities short (average of 5-10 minutes). Some activities may require longer periods of time to complete, possibly up to 30 minutes. But those activities should be the exception, not the rule. Take one lecture and add one active component to it. Do this over a period of time and increase the amount that you use until you feel you have reached a balance that works for you and your teaching style.
4. *Plan and delegate* – One concern in using active learning along with lecture is anxiety about covering the breadth of material that is covered in a lecture-only format. Though this concern might have some degree of merit, I have found in my own experience that this is not the case. However, successful implementation does require some degree of thought and preparation. I have personally found that I save time while lecturing by not spending as much time working through an example in class. Before, I would spend a lot of time in order to make sure students understood every facet. Now, I realize that any part they don't understand will be magnified in the active component which follows my example. Also, there are often concepts that can be learned in the active component rather than taught in lecture. This provides a unique form of constructivist learning and is easy to implement within such a structure. Whereas there is much discussion of the fatigue that faces students in a long lecture, I think a neglected area of research is the fatigue that occurs with an instructor. I find by using mini-lectures that I stay focused on my teaching and can cover the same material in less time.

5. *Be active yourself* – While students are working on an activity, consider walking among them and observe their progress. This can be beneficial for many reasons including making sure students are working on the task you have asked them to work on. Also, some students may not volunteer problems they are having in front of the entire class and this provides you with an opportunity to see particular problems they may be having. Obviously with moderate (as in the case of this study) to large classes, it might be impossible to go to each person or group in the allotted time for the activity. In this situation, if you have more than one activity for that day, consider randomly moving around during each activity. In my case, I will mentally cluster students in various parts of the room and randomly choose clusters to visit during an activity. I strive to visit each cluster at least once during a class.
6. *Don't confuse noise for control (or lack of)* – Students in my classes are often quiet during the first few weeks of active components, perhaps because it seems such a foreign approach for a mathematically-based class. Yet over time, they usually embrace it heartedly to the point that the classroom is filled with talk and laughter. For someone coming from a lecture-only format, this can seem threatening, as if you have lost control of the class. Always keep in mind that you are actively engaging your students and helping them master the material. Some of my classes have been so engaged that it has taken me quite a few seconds to regain their attention. At first this would bother me greatly to the point of questioning my new approach. Now, I can usually regain control of the class easier by simply talking to them while moving around among them and simply saying, "OK, let's talk about something that I see you all are having a problem with." Others may wish to utilize visible clues that are discussed earlier with the class (e.g., turning the lights on and off a few times). If you are new to this teaching style, do not let such matters deter you from exploring this "brave new world."

6.5. FUTURE RESEARCH

One possible extension of this study is to adjust the way in which team projects are done. Some students wished for more freedom in the choice of their projects. A future study could consider this effect. Smaller group sizes should also be considered as this may reduce some of the tensions in groups found in this study. Also, an approach in which the instructor supervises the teams more vigorously might prove valuable and eliminate the slightly diminishing results from team projects. The approach taken in this study was not as rigorous because the instructor assumed that graduate students would not need as much oversight. This study considered only the basic-level statistics course for graduate students in order to serve as a bridge with similar research in undergraduate studies. Further studies applied to advanced courses (e.g., regression, multi-level models, structural equation modeling) should be investigated. Tying together affective and cognitive measures in a unified statistical analysis, as well as tracking changes or growth over time in longitudinal studies, would also be beneficial.

REFERENCES

Aliaga, M., & Gunderson, B. (2003). *Interactive statistics*. Upper Saddle River, NJ: Prentice Hall.

Bligh, D. A. (2000). *What's the use of lectures?* San Francisco: Jossey-Bass.

Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, *50*, 69-78.

Bryce, G. R. (2005). Developing tomorrow's statisticians. *Journal of Statistics Education*, *13*(1).
[Online: http://www.amstat.org/publications/jse/v13n1/bryce.html]

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education*, 5(3).
[Online: http://www.amstat.org/publications/jse/v5n3/chance.html]

Chance, B. L. (2005). Integrating pedagogies to teach statistics. In J. Garfield (Ed.), *Innovations in Teaching Statistics*. *MAA Notes, Vol*. 65 (pp. 101-112). Washington, DC: Mathematical Association of America .

Cobb, G. W. (1991). Teaching statistics: More data, less lecturing. *Amstat News*, *December*, *No. 182*, 1 and 4.

Cobb, G. W. (1993). Reconsidering statistics education: A National Science Foundation conference. *Journal of Statistics Education*, *1*(1).
[Online: http://www.amstat.org/publications/jse/v1n1/cobb.html]

Delucchi, M. (2006). The efficacy of collaborative learning groups in an undergraduate statistics course. *College Teaching*, *54*(2), 244-248.

Dietz, E. J. (1993). A cooperative learning activity on methods of selecting a sample. *The American Statistician*, *47*, 104-108.

Fillebrown, S. (1994). Using projects in an elementary statistics course for non-science majors. *Journal of Statistics Education*, *2*(2).
[Online: http://www.amstat.org/publications/jse/v2n2/fillebrown.html]

Ford, N., & Chen, S. Y. (2002). Matching/mismatching revisited: An empirical study of learning and teaching styles. *British Journal of Educational Technology*, *32*(1), 5-22.

Gal, I., & Garfield, J. B. (Eds.) (1997). *The assessment challenge in statistics education.* Amsterdam: IOS Press/International Statistical Institute.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*(1).
[Online: http://www.amstat.org/publications/jse/v1n1/garfield.html]

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, *63*(1), 35-48.

Garfield, J. & delMas, R. (1991). Students' conceptions of probability. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics, Volume 1* (pp. 340-349). Voorburg, The Netherlands: International Statistical Institute.

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, *5*(3). [Online: http://www.amstat.org/publications/jse/v5n3/giraud.html]

Gnanadesikan, M., Scheaffer, R. L., Watkins, A. E., & Witmer, J. A. (1997). An activity-based statistics course. *Journal of Statistics Education*, 5(2).
[Online: http://www.amstat.org/publications/jse/v5n2/gnanadesikan.html]

Grasha, A. F. (1996). *Teaching with style*. Pittsburgh, PA: Alliance Press.

Hogg, R. V. (1991). Statistical education: Improvements are badly deeded. *The American Statistician*, *45*, 342-343.

Hunter, W. G. (1977). Some ideas about teaching design of experiments, with 2^5 examples of experiments conducted by students. *The American Statistician*, *31*, 12-17.

Johnson, H. D., & Dasgupta, N. (2005). Traditional versus non-traditional teaching: Perspectives of students in introductory statistics classes. *Journal of Statistics Education*, *13*(2).[Online: www.amstat.org/publications/jse/v13n2/johnson.html]

Jones, L. (1991). Using cooperative learning to teach statistics. *Research Report Number 91-2,* The L. L. Thurstone Psychometric Laboratory, University of North Carolina.

Jordan, J. (2007). The application of statistics education research in my classroom. *Journal of Statistics Education*, *15*(2).
[Online: http://www.amstat.org/publications/jse/v15n2/jordan.html]

Keeler, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, *3*(2).
[Online: http://www.amstat.org/publications/jse/v3n2/keeler.html]

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Upper Saddle River, NJ: Prentice Hall.

Ledolter, J. (1995). Projects in introductory statistics courses. *The American Statistician*, *49*, 364-367.

Lovett, M. C. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 347-384). Mahwah, NJ: Erlbaum.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, *54*(3), 196-206.

Mackisack, M. (1994). What is the use of experiments conducted by statistics students? *Journal of Statistics Education*, *2*(1).
[Online: http://www.amstat.org/publications/jse/v2n1/mackisack.html]

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2), 123-137.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston, MA: McGraw-Hill.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group activity-based model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*, 285-315.

Singer, J. D., & Willett, J. B. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician*, *44*, 223-230.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, *6*(3).
[Online: http://www.amstat.org/publications/jse/v6n3/smith.html]

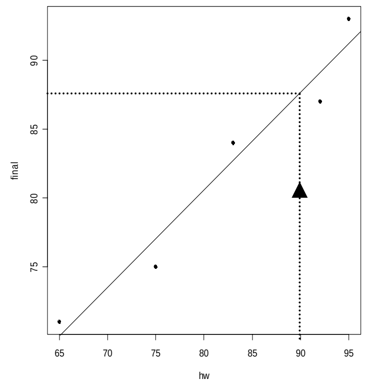
Snee, R. D. (1993). What's missing in statistical education? *The American Statistician*, *47*, 149-154.

Steinhorst, R., & Keeler, C. (1995). Developing material for introductory statistics courses from a conceptual, active learning viewpoint. *Journal of Statistics Education*, *3*(3).
[Online: http://www.amstat.org/publications/jse/v3n3/steinhorst.html]

Ward, B. (2004). The best of both worlds: A hybrid statistics class. *Journal of Statistics Education*, *12*(3).
[Online: http://www.amstat.org/publications/jse/v12n3/ward.html]

BRANDON VAUGHN
Department of Educational Psychology
1 University Station, D5800
The University of Texas at Austin
Austin, TX 78712-1294

APPENDIX A: AN EXAMPLE OF ACTIVITY-BASED IN-CLASS LEARNING



The least square regression line for homework score (x) and final exam score (y) is $\hat{y} = 23.87 + 0.71x$. The slope of b = 0.71 indicates that if homework grade were to increase by 1-point, we would expect on average for the final exam grade to increase by 0.71 points. The estimated y-intercept (a = 23.87) indicates that a student who made a 0 on their homework grade would have a predicted final exam grade of 23.87.

Let's pretend we have a student who has a 90 homework grade, and we wish to predict their final exam grade. One way to do this is to use the scatterplot and our least squares regression line. If you find 90 on the x-axis and trace up to the regression line, you can look across and get a predicted value for y (final grade) of around 87 or 88. Another way to do this would be to plug the value x = 90 into the regression equation: $\hat{y} = 23.87 + 0.71(90) = 87.77$.

Your Turn!

Assume that the hours a student watches television and their test grade are linearly related. The least squares regression line was calculated as $\hat{y} = 95.3 - 1.54x$, where y is the test grade and x is the number of hours a student watches television during a week.

(a) Interpret the value of the estimated slope b = –1.54.

(b) How would you interpret the y-intercept of a = 95.3? Does the interpretation make practical sense in this instance?

(c) What would you predict the test grade to be for a student who watches 5 hours of television on a given week?

(d) What would you predict the test grade to be for a student who watches 70 hours of television on a given week?

(e) Why do you think your answer to part (d) was so inaccurate?

APPENDIX B: SAMPLE OF TEAM PROJECTS

Data

1. It has been reported that college students are pessimistic about the future of the social security system. How would you survey students at our school to learn more about how confident they are that they will receive substantial social security checks when they retire? Develop a short questionnaire, and explain exactly how you would select the students to be questioned.
2. Ask 25 randomly selected people this question: "Karl Marx said, 'Whenever a form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it.' Do you agree?" Now ask 25 randomly selected people this question: 'The U.S. Declaration of Independence says: 'Whenever a form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it.' Do you agree?" Compare your responses.
3. Working either in pairs, or alone, go to a well populated area (like a shopping center). Randomly select people and ask them if you can do a quick survey for your statistics class. If they say yes, ask them the following two questions: "Do you believe in cloning? Do you believe in life-saving organ replication?" Record the answers for at least 50 people. Summarize your findings.
4. Working either in pairs, or alone, go to a well populated area (like a shopping center). Randomly select people and ask them if you can do a quick survey for your statistics class. If they say yes, ask them the following two questions: "Disposable diapers account for 6% of landfill waste, whereas yard wastes and plastic bottles account for 40% of yard wastes. In light of this fact, do you think it is fair to ban disposable diapers?" Record the answers for at least 30 people. Summarize your findings. Then rephrase the question to be more neutral. Redo another random sample of 30 people. Compare the response to the original question. Comment.
5. Should kids between 12 and 18 have cell phones? Conduct a survey of local residents to find their opinions. Be sure to ask both adults who have wireless service and those who don't. Report any general findings, and also see what general differences in opinion exist for adults who use cell phones and those who do not. Be sure to report on your sampling method (whether good or bad), possible bias in the survey, and ways the study could be improved. Try to get about 50 subjects if possible.
6. Conduct a survey of students (other than your statistics class) on the subject of the internet and cheating. In particular, design and administer a survey that asks several questions about whether students have used the internet and committed plagiarism. Make the survey anonymous, and stress that to students. What effect do you think such a loaded question might have on the response? Report your findings in a general sense? How often do students tend to cheat if they do at all?

Descriptive statistics

7. Sample a group of students in regard to a fictitious question (for example, "Do you support the rebel efforts in Alfa-Centuri?" … Alfa-Centuri does not exist.). Feel free to use my question or come up with one of your own. Develop a survey with a few factual questions and stick this fictitious question among them. Survey at least 30 students or people. Try to get a mixture of males and females.

Compare males and females using some of the graphs and summaries discussed in class.

8. Select teachers (you cannot select me ☺), and record the number of times that this teacher says “Uh” or “Um” during the class period. Try to obtain at least 10 teachers, and if possible at least 5 male and female. Summarize the results numerically by teacher. Do a side-by-side boxplot comparing the gender of the teacher. What differences did you find? Which gender had more variability? [Note: Creating boxplots with so few data values becomes a class discussion point.]
9. Post a sign on the main entrance to a campus building requesting the use of a less convenient entrance; for example, “Please use the door on the north side of building.” From an inconspicuous location, observe how many people ignore the sign and use the main entrance and how many people do not use the main entrance. Compare the behavior of students and professors or males and females. Try to pick a building and time when traffic is light, so that large numbers do not try to enter simultaneously. Try to get at least 50 in your sample.

Hypothesis testing

10. Do more expensive cookies taste better than less expensive cookies? Choose two brands of cookies that appear to be similar but cost quite different amounts. Ask at least 40 people to taste an unlabeled cookie from each brand and to rate each cookie on a scale of 1 to 10; use a matched-pair test to assess the statistical significance of your results.
11. Do males over-exaggerate their heights? Take a random sample of at least 50 males. Ask them their heights. Then measure their heights. Do a matched pairs *t* test to see if males tend to over-exaggerate their heights.
12. Do males and females differ in terms of the number of traffic tickets they get? Or in the number of accidents they’ve been involved in? Do a survey of at least 25 males and 25 females. Compare each gender on both issues using a 2-sample *t* test at a 5% significance level. Are there any differences?
13. Does a difference exist between males and females in regard to the number of hours of television they watch per week? Does a difference exist in the number of hours of video games played? Conduct a survey and test the differences between males and females on both issues at a 5% significance level using a 2-sample *t* test. Try to get at least 25 males and 25 females.
14. Is there a statistical difference in the pulse rates of smokers and nonsmokers? In particular, test that the pulse rate of smokers is higher than that of nonsmokers. Sample at least 25 from both groups. Use a 2-sample *t* test and a 5% significance level.

Regression

15. Go to your campus bookstore, and select 30 new hardcover textbooks. (Be sure to explain how these books were selected.) For each book, record the number of pages and the price. Now use a scatterplot to see if there appears to be a positive or negative relationship between the number of pages and the price. Calculate the correlation coefficient, and determine whether there is a statistically significant relationship between number of pages and price.

16. Collect data for at least 10 years on the cost of attending your college. Using years as an independent variable, perform a regression analysis to see if there is a trend in the real cost of attending this college. [Note: Adjusting for inflation can become a class discussion point.]
17. Ask at least 50 people their height, and the height of their parent (same gender). Perform a regression analysis with the student's height as the dependent variable and the parent's height as the explanatory variable.
18. Conduct a survey to see if there is a relationship between GPA and the number of alcoholic drinks that a student consumes. What type of relationship do you think would exist? Did it? You need to consider what constitutes an alcoholic drink, and must also decide what to measure for the number of drinks (oz? Containers? etc). Let GPA be the dependent (response) variable.
19. A *CNN/USA TODAY* poll conducted by Gallup asked a sample of employed Americans the following question: "Which do you enjoy more, the hours when you are on your job, or the hours when you are not on your job?" Construct a 10-item Likert scale survey (of which 5-items address satisfaction with leisure, and 5-items measure satisfaction with job), and give the survey to a random selection of 30-40 adults who work. Sum each of the 5-items to get an overall satisfaction score. Perform a complete regression analysis to see what relationship exists between leisure and job satisfaction.
20. Conduct a study to see whether there is a relationship between a student's GPA and the number of hours the student watches television each week. Perform a thorough regression analysis, and try to get at least 30 students in your survey.
21. Conduct a survey of 30-40 students where you ask them to estimate the number of hours spent on the internet (including school work), number of hours working each week (job), number of classes, and GPA. Find correlations between each of these variables. Pick two variables of interest and perform a regression analysis.
22. Conduct a study to see whether there is a relationship between a student's grade point average and the number of hours the student studies each week. Perform a thorough regression analysis, and try to get at least 30 students in your survey.

FORTHCOMING IASE CONFERENCES

ISI-57
THE 2009 SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE
Durban, South Africa, August 16 – 22, 2009

IASE sponsored Invited Paper Meetings for 57th Session in Durban are being organised by Helen MacGillivray (Australia, h.macgillivray@qut.edu.au). The IASE Programme Committee for ISI-57 has chosen the theme "Statistics Education for the Future."

IASE has nine IPM (Invited Paper Meeting) sessions, two of which include issues raised by the local organisers, and has two joint sessions with IAOS.

Session Number	Section representation	Title of Invited Paper Meeting	Organiser(s)
IPM15	IAOS IASE	The challenge of building a supply of statisticians for the future	To be determined, c/o Nancy McBeth, Nancy.McBeth@stats.govt.nz
IPM36	IASE IAOS	The roles of statistical agencies in developing statistical literacy	Reija Helenius, Finland, Reija.Helenius@stat.fi
IPM37	IASE Local Hosts	Educating the public on how to use official statistics.	Peter Wingfield-Digby, pwdigby@loxinfo.co.th
IPM38	IASE Local Hosts	Challenges faced in Statistics Education in African countries	Delia North, South Africa, northd@ukzn.ac.za
IPM39	IASE	Balancing the training of future statisticians for workplace and research	Charles Rohde, USA, crohde@jhsph.edu
IPM40	IASE	Exploiting the Progress in Statistical Graphics and Statistical Computing for the benefit of Statistical Literacy	Juana Sanchez, USA, jsanchez@stat.ucla.edu
IPM41	IASE	Survey Research in Statistics Education	Irena Ograjensek, Slovenia, irena.ograjensek@ef.uni-lj.si
IPM42	IASE	Research on Informal Inferential Reasoning	Katie Makar, Australia, k.makar@uq.edu.au
IPM43	IASE	Teaching, Learning and Assessing Statistics Problem Solving in Higher Education	Neville Davies, UK, neville.davies@ntu.ac.uk
IPM44	IASE	Technologies for learning and teaching in developing countries	Gabriella Belli, USA, gbelli@vt.edu
IPM45	IASE	Virtual Learning Environments for Statistics Education	Adriana Backx Noronha Viana, Brazil, backx@usp.br and Pieternel Verhoeven, Netherlands, n.verhoeven@roac.nl

Statistics Education Research Journal, 8(1), 131-137, http://www.stat.auckland.ac.nz/serj

The website http://www.statssa.gov.za/isi2009/ has information on all matters relating to ISI 2009, including important dates, and will be regularly updated as new information develops.

More information: Helen MacGillivray, h.macgillivray@qut.edu.au

2009 IASE SATELLITE CONFERENCE TO THE 57TH SESSION OF THE ISI "NEXT STEPS IN STATISTICS EDUCATION" Durban , South Africa August 14 -15, 2009 (Immediately before ISI 57 in Durban)

All submissions addressing the theme "Next Steps in Statistics Education " will be welcome. This theme has been chosen to particularly attract papers under the following headings:

1. What constitutes best practice for the curriculum beyond the "Introductory Statistics" course? What courses should follow on for those wishing to major in Statistics and what additional training should we offer to those in other disciplines?
2. What elements of our undergraduate curriculum specifically prepare our students for their careers post-graduation, either in the workplace or as masters/doctoral students? How can we improve these elements?
3. Now that more countries have school curricula that include substantial emphasis on data and chance, how can we better prepare teachers for implementing those curricula? What curricular materials and tools can we develop to improve students' learning of statistics at school level?
4. Since the 1949 formation of its precursor, the ISI Statistical Education Committee, the IASE has matured as an organisation. As we move towards ICOTS 8, we note that great progress has already been made in the field of Statistics Education but the challenge we face now is to consider the next steps that we must take. How can we build on past progress to raise the profile of our field so that it becomes a more visible and vibrant pursuit?

More information can be found on conference webpage:
http://www.ucd.ie/statdept/2009_iase_satellite.html
Conference Email: IASE_Satellite@maths.ucd.ie

SRTL-6 THE SIXTH INTERNATIONAL RESEARCH FORUM ON STATISTICAL REASONING, THINKING, AND LITERACY The Role of Context and Evidence in Informal Inferential Reasoning Brisbane, Australia, July 10 - 16, 2009

The sixth in a series of International Research Forums on Statistical Reasoning, Thinking, and Literacy (SRTL-6) is to be held in Brisbane, Australia from July 10 to July 16, 2009. The School of Education at The University of Queensland, will host the Forum.

The Forum's focus will build on the work presented and discussed at SRTL-5 on informal ideas of statistical inference. Recent research suggests an important role for developing ideas of informal types of statistical inference even at early educational levels. Researchers have developed instructional activities that encourage students to infer beyond samples of data and use technological tools to support these informal inferences.

The findings of these studies reveal that the context of the data and the use of evidence may be important factors to study further. The role of context is of particular interest because in drawing (informal) inferences from data, “students must learn to walk two fine lines. First, they must maintain a view of data as ‘numbers with a context’ (Moore, 1992).” At the same time, “they must learn to see the data as separate in many ways from the real-world event they observed” (Konold & Higgins, 2003, p. 195). That is, they must abstract the data from that context. The role of evidence is also of particular interest because in learning how to make data-based claims (argumentation), students must consider the evidence used to support the claim, the quality and justification of the evidence, limitations of the evidence and finally, an indication of how convincing the argument is (Ben-Zvi, Gil, & Apel, 2007).

Based on SRTL-5, we characterize Informal Inferential Reasoning (IIR) as the cognitive activities involved in drawing conclusions with some degree of uncertainty that go beyond the data and having empirical evidence for them. Three principles appear to be essential to informal inference: (1) generalizations (including predictions, parameter estimates, and conclusions) that go beyond describing the given data; (2) the use of data as evidence for those generalizations; and (3) conclusions that express a degree of uncertainty, whether or not quantified, accounting for the variability or uncertainty that is unavoidable when generalizing beyond the immediate data to a population or a process (Makar & Rubin, 2007).

An interesting range of diverse research presentations and discussions have been planned and we look forward to a stimulating and enriching gathering. These papers will address the role of context and evidence when reasoning about informal inference at all levels of education including the professional development of elementary and secondary teachers.

The structure of the scientific program will be a mixture of formal and informal sessions, small group and whole group discussions, and the opportunity for extensive analysis of video-taped research data. There will also be a poster session for exhibiting current research of participants on additional topics related to statistics education. The Forum is co-chaired by Dani Ben-Zvi (University of Haifa, Israel) and Joan Garfield (University of Minnesota, USA), locally organized by Katie Makar and Michael Bulmer (The University of Queensland), and planned by a prestigious international advisory committee. Conference attendance is by invitation only.

For more information, visit the SRTL website at:
http://srtl.stat.auckland.ac.nz/ or email SRTL2009@gmail.com.

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

The 2010 International Conference on Teaching Statistics will be held in the city of Ljubljana, Slovenia, July 11-16. It is being organised by the IASE and the Slovenian Statistical Association. The venue will be the Ljubljana Cultural and Congress Centre.

Statistics educators, statisticians, teachers, and educators at large are invited to contribute to the scientific programme. Types of contribution include invited papers, contributed papers, and posters. No person may author more than one

Invited Paper at the conference, although the same person can be co-author of more than one paper, provided each paper is presented by a different person.

Voluntary refereeing procedures will be implemented for ICOTS-8. Details of how to prepare manuscripts, the refereeing process and final submission arrangements will be announced later.

INVITED PAPERS

Invited Paper Sessions are organized within 10 Conference Topics as follows.

Topics and Topic Convenors

1. Data and Context in Statistics Education: Towards an Evidence-based Society.
   Brian Phillips (Australia) bphillips@swin.edu.au
   Irena Ograjensek (Slovenia) irena.ograjensek@ef.uni-lj.si
2. Statistics Education at the School Level.
   Mike Shaughnessy (USA) mikesh@pdx.edu
   Doreen Connor (UK) doreen.connor@ntu.ac.uk
3. Learning to Teach Statistics.
   Katie Makar (Australia) k.makar@uq.edu.au
   Joachim Engel (Germany) engel@math.uni-hannover.de
4. Statistics Education at the Post Secondary Level.
   Elisabeth Svensson (Sweden) elisabeth.svensson@esi.oru.se
   Larry Weldon (Canada) weldon@sfu.ca
5. Assessment in Statistics Education.
   Beth Chance (USA) bchance@calpoly.edu
   Iddo Gal (Israel) iddo@research.haifa.ac.il
6. Statistics Education, Training and the Workplace.
   Gabriella Belli (USA) gbelli@vt.edu
   Peter Petocz (Australia) peter.petocz@mq.edu.au
7. Statistics Education and the Wider Society.
   Richard Gadsden (UK) R.J.Gadsden@lboro.ac.uk
   Oded Meyer (USA) meyer@stat.cmu.edu
8. Research in Statistics Education.
   Arthur Bakker (The Netherlands) a.bakker@fi.uu.nl
   Tim Burgess (New Zealand) t.a.burgess@massey.ac.nz
9. Technology in Statistics Education.
   Deborah Nolan (USA) nolan@stat.berkeley.edu
   Paul Darius (Belgium) paul.darius@biw.kuleuven.be
10. An International Perspective on Statistics Education.
    Delia North (South Africa) northd@ukzn.ac.za
    Enriqueta Reston (Phillipines) edreston@usc.edu.ph

Session themes within each Topic are currently being discussed. The themes and Session organizers with email contact are available on the ICOTS-8 web site http://icots8.org/, under "Scientific Programme." Those interested in submitting an invited paper should contact the appropriate Session Organiser before December 1, 2008.

CONTRIBUTED PAPERS

Contributed paper sessions will be arranged in a variety of areas. Those interested in submitting a contributed paper should contact either Gilberte Schuyten (Gilberte.Schuyten@UGent.be), John McKenzie (mckenzie@babson.edu), or Flavia Jolliffe (F.Jolliffe@kent.ac.uk) before September 1, 2009.

POSTERS

Those interested in submitting a poster should contact Mojca Bavdaz (mojca.bavdaz@ef.uni-lj.si) or Alesa Lotric Dolinar (alesa.lotric.dolinar@ef.uni-lj.si) before January 15, 2010.

GENERAL ISSUES

More information is available from the ICOTS-8 web site at http://icots8.org/ which will continue to be updated over the next three years, or from the ICOTS IPC Chair John Harraway, (jharraway@maths.otago.ac.nz), the Programme Chair, Roxy Peck (rpeck@calpoly.edu), and the Scientific Secretary, Helen MacGillivray (h.macgillivray@qut.edu.au).

OTHER FORTHCOMING CONFERENCES

USCOTS 2009
UNITED STATES CONFERENCE ON TEACHING STATISTICS
"LETTING GO TO GROW"
Columbus, OH, USA, June 25 - 27, 2009

The third biennial United States Conference on Teaching Statistics (USCOTS 09) will be held on June 25-27, 2009 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 09 is Letting Go to Grow. "Letting Go" has many interpretations, such as letting go of some classic course content in order to better align with course goals, letting go of some old ideas about pedagogy in order to use more effective methods, or letting go of old notions about the students we teach in order to better facilitate their learning. USCOTS is a "working conference" with many opportunities for hands-on activities, demonstrations, networking, sharing ideas, and receiving the latest information on research and best practices in teaching statistics. Leaders in statistics education and assessment will give plenary talks, including Dani Ben-Zvi (Haifa, Israel), George Cobb (USA), Peter Ewell (USA), Ronald Wasserstein (USA), and Chris Wild (Auckland, New Zealand).

Details are available at USCOTS web page: http://www.causeweb.org/uscots

INNOVATIVE APPROACHES TO TURN STATISTICS INTO KNOWLEDGE
Washington, D.C., USA, July 15 – 16, 2009

This two-day seminar and a plenum should contribute to the development of tools to help people transform statistics into knowledge and decisions. A first condition for statistics to be used this way is that relevant statistics become known, available, and understood by wider audiences. The seminar is held in the context of the OECD Global Project on "Measuring the Progress of Societies." It should contribute to one of the goals quoted in the Istanbul Declaration: "Produce a broader, shared, public understanding of changing conditions, while highlighting areas of significant change or inadequate knowledge."

The seminar can be seen as a continuation of the previous seminars organized in Rome and Stockholm and of the first International Exhibition on "Innovative Tools to Transform Information into Knowledge," organised during the second OECD World Forum on "Statistics, Knowledge and Policy" (Istanbul, 27-30 June 2007).

We want to look at tools and applications for making statistics more popular, while avoiding the pitfalls of populism, over-simplification, or propaganda. We want to base all these initiatives on scientific standards, observing the basic principles of objectivity and good communication. We would therefore welcome experts in statistical methodology, cognitive science, and communication as active participants in the workshop.

Details are available at seminar web page:
http://www.oecd.org/progress/ict/statknowledge

10TH INTERNATIONAL CONFERENCE OF THE MATHEMATICS EDUCATION INTO THE 21ST CENTURY PROJECT MODELS IN DEVELOPING MATHEMATICS EDUCATION
Dresden, Saxony, Germany, September 11 – 17, 2009

The Mathematics Education into the 21st Century Project was founded in 1986 and is dedicated to the planning, writing and disseminating of innovative ideas and materials in Mathematics and Statistics Education. You are invited to attend our 10th anniversary project conference to be held in the historic city of Dresden, Germany. The conference is organized in full cooperation with the Saxony Ministry of Education. All our conferences have a string Statistics Education component. You are warmly invited to attend our conference in the heart of the historic city of Dresden.

INTERNATIONAL ORGANISERS
Dr. Alan Rogerson, Coordinator of the Mathematics in Society Project (Poland), Prof. Fayez Mina, Faculty of Education, Ain Shams University (Egypt)

CHAIR OF THE LOCAL ORGANISING COMMITTEE
Prof. Dr. Ludwig Paditz, Dresden University of Applied Sciences.

Further information: Alan Rogerson, arogerson@inetia.pl
Website: http://math.unipa.it/~grim/21project.htm

2009 JOINT STATISTICAL MEETINGS
Washington, D.C., USA, August 1-6, 2009

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 of the meeting include oral presentations, panel sessions, poster presentations, continuing education courses, 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.
More information: jsm@amstat.org
Website: http://www.amstat.org/meetings/jsm/2009/