Statistics Education Research Journal

Volume 2 Number 2 November 2003

International Association for Statistical Education
http://www.cbs.nl/isi/iase.htm

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

Statistics Education Research Journal

Statistics Education Research Journal is published by the International Association for Statistical Education and the International Statistical Institute to:

- encourage research activity in statistics education;
- advance knowledge about students' attitudes, conceptions, and difficulties as regards stochastic knowledge;
- improve the teaching of statistics at all educational levels.

The Journal encourages the submission of quality papers, including research reports, theoretical or methodological analyses, and integrative literature surveys, that can advance scholarly knowledge, research methods, and educational practice in any of the broad areas related to statistical education or learning of statistics and probability at all educational levels and in all educational contexts. Contributions in English are recommended. Contributions in French and Spanish will also be accepted. All papers are blind-refereed by at least two referees.

ISSN: 1570-1824

Submissions

Manuscripts should be sent to co-editor Flavia Jolliffe (F.Jolliffe@kent.ac.uk), by email, as an attached document in RTF format. Manuscripts will be submitted in a form ready for blind review to ensure anonymity in the refereeing process, i.e., all identifying information will be removed by the author from all parts of the document. However, the author's name and email will be left only on the first page of the manuscript for file identification, and will be removed by the editor before sending to external referees. Full contact information for all authors will be included in the email message to the editor. Manuscripts should conform to the style and instructions specified in the Guidelines for Authors on the Journal's Web page: http://fehps.une.edu.au/serj.

TABLE OF CONTENTS

Thanks from Carmen Batanero

Carmen Batanero would like to publicly thank the other members of the SERJ editorial board for supporting her while she has implemented the transformation of the *Statistics Education Research Newsletter* (SERN) into SERJ. She is also most grateful to all the referees who helped her in revising papers and in contributing comments to improve the quality of SERJ. She finally thanks all the authors who trusted the SERJ editorial board and submitted their papers to make the project a reality.

Thanks from the SERJ Editorial Board

In turn the other members of the board would like to thank Carmen Batanero for her enthusiasm for this project and the many hours of work she has done to help ensure its success. We are pleased that she will continue as an associate editor and will look after submissions in Spanish.

Welcome to Iddo Gal

The board members are also pleased to welcome associate editor Iddo Gal as editor in place of Carmen Batanero. He has already been working with Carmen Batanero and Flavia Jolliffe in anticipation of his change of role and has made many useful contributions to the way SERJ operates.

EDITORIAL

Welcome to the fourth edition of SERJ, the last edition with which Carmen Batanero has been involved as editor. We have three papers in this issue. The first of these, the paper by Jane Watson and Rosemary Callingham, is very comprehensive and addresses the issues of measuring statistical literacy. This paper is likely to become a key reference for other work in this field. The second paper is by Flavia Jolliffe and addresses an important problem, organising an easily accessible database of research in statistical education. The survey described in her paper is expected to go live before the end of 2003. We note that when a paper is submitted by a member of the editorial board, as in this case, it has to pass our usual blind refereeing process. The third paper, by Jamie Mills, summarises the research he has done into texts which incorporate SPSS.

The editorial board discussed various matters relating to SERJ during the 54th Session of the International Statistical Institute (ISI54) in Berlin in August and afterwards using email. Much of the discussion was on how to make the operation of the editorial board more efficient, but some of the decisions have affected the content of this issue. As we receive a growing number of papers which are suitable for publication, we feel that we are now in a position to concentrate on making SERJ a top quality refereed journal. Thus we are no longer including sections on summaries of publications and recent dissertations which we have published during the transition period while the newsletter *Statistics Education Research Newsletter* (SERN) has been changing into SERJ. We are now reporting and announcing in full only conferences which have been organised or supported by IASE, and for other forthcoming conferences of interest to SERJ are giving the main information only.

The summaries of dissertations with a statistical education content were thought to be useful and we are pleased to announce that Joan Garfield will be organising the collection of such summaries and hosting a web page. Further details will be announced in SERJ 3(1).

A special issue of SERJ is planned for the second issue in 2004. This will be on statistical reasoning about variability and will be based on papers presented at the third international research forum on statistical reasoning, thinking and literacy held in 2003. The authors of the papers presented in this forum have been invited to submit their papers to SERJ for refereeing. Joan Garfield and Dani Ben-Zvi will be guest editors for this issue. The editorial board of SERJ is pleased to be co-operating with the SRTL-3 organisers in this way.

Chris Wild took over the presidency of IASE from Carmen Batanero at ISI54. We are pleased that we shall continue to have the active involvement of an IASE president in SERJ and Chris Wild has already made some changes to the IASE web page making the links to SERJ more obvious.

Finally, Chris Reading now has a new title of Assistant Editor to acknowledge all the input she makes to SERJ, both in looking after the web page and in overseeing the final stages of production of each issue.

CARMEN BATANERO AND FLAVIA JOLLIFFE

STATISTICAL LITERACY: A COMPLEX HIERARCHICAL CONSTRUCT

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

ROSEMARY CALLINGHAM
University of New England
rcalling@pobox.une.edu.au

SUMMARY

The aim of this study was, first, to provide evidence to support the notion of statistical literacy as a hierarchical construct and, second, to identify levels of this hierarchy across the construct. The study used archived data collected from two large-scale research projects that studied aspects of statistical understanding of over 3000 school students in grades 3 to 9, based on 80 questionnaire items. Rasch analysis was used to explore an hypothesised underlying construct associated with statistical literacy. The analysis supported the hypothesis of a unidimensional construct and suggested six levels of understanding: Idiosyncratic, Informal, Inconsistent, Consistent non-critical, Critical, and Critical mathematical. These levels could be used by teachers and curriculum developers to incorporate appropriate aspects of statistical literacy into the existing curriculum.

Keywords: Statistical literacy; school students; Rasch analysis; conceptual hierarchy

1. INTRODUCTION

Historically there are two antecedents to this study. One is the growing interest, from the middle of the last century, in numeracy as a co-agent with literacy becoming an essential education foundation for all school students. The other is the specific inclusion of data handling and chance in the mathematics curricula of many countries toward the end of the century.

1.1. EMERGENCE OF NUMERACY

Although the term literacy has been accepted for a long time as describing the ability to read and write, the acceptance of a similar term has been more elusive in describing number or mathematical skills. The term numeracy was introduced in the Crowther Report in the United Kingdom in 1959:

> *A word to represent the mirror image of literacy… On the one hand an understanding of the scientific approach to the study of phenomena — observation, hypothesis, experiment, verification. On the other hand… the need in the modern world to think quantitatively, to realize how far our problems are problems of degree even when they appear to be problems of kind. Statistical ignorance and statistical fallacies are quite as widespread and quite as dangerous as the logical fallacies that come under the heading of illiteracy* (quoted in Cockcroft, 1982, para. 36).

Statistics Education Research Journal 2(2), 3-46, http://fehps.une.edu.au/serj

For those interested in statistical literacy it may be regretted that this definition was not immediately and universally accepted. The dilution of the term numeracy to include only basic skills with numbers, has led some people into the multiple literacies milieu, using the phrase "quantitative literacy" to describe the broad range of understanding required when students leave school. This is particularly true in the United States where Quantitative Literacy is in the title of recent significant reports on the need to improve mathematical and statistical understanding of the general population (Steen, 1997, 2001). In Australia, however, the term numeracy continues to have a broad meaning. For example, the following description of numeracy was endorsed in 1997 at a conference organized by the Australian Association of Mathematics Teachers Inc (AAMT).

To be numerate is to use mathematics effectively to meet the general demands of life at home, in paid work, and for participation in community and civic life. In school education, numeracy is a fundamental component of learning, performance, discourse and critique across all areas of the curriculum. It involves the disposition to use, in context, a combination of:

- *underpinning mathematical concepts and skills from across the discipline (numerical, spatial, graphical, statistical and algebraic);*
- *mathematical thinking and strategies;*
- *general thinking skills; and*
- *grounded appreciation of context* (AAMT, 1997, p. 15).

1.2. STATISTICS IN THE SCHOOL CURRICULUM

The second antecedent to interest in statistical literacy was the introduction of revised mathematics curricula in the early 1990s around the world (e.g., National Council of Teachers of Mathematics [NCTM], 1989; Ministry of Education, 1992). In Australian states these curricula were based to a greater or less extent on *A National Statement on Mathematics for Australian Schools* (Australian Education Council [AEC], 1991), which included chance and data as one of five content areas of the mathematics curriculum. Written with the advice of statisticians and educators, the chance and data curriculum could not benefit from previous research into school students' understanding of probability and statistics concepts as, with the exception of some work in the area of probability (e.g., Fischbein, 1975; Green, 1982, 1983b), there had been virtually none.

Most national curriculum documents (e.g., AEC, 1991; Ministry of Education, 1992; NCTM, 1989) reflected the five components that comprise a statistical investigation based on a question of interest, as suggested by Holmes (1980): data collection, data tabulation and representation, data reduction, probability, and interpretation and inference. In Australia Holmes' categories were aggregated into three major subheadings of the mathematics curriculum: chance, data handling, and statistical inference. In the latest standards of the NCTM (2000) in the United States these ideas are included in the Data Analysis and Probability Strand:

Instructional programs from pre-kindergarten through grade 12 should enable all students to:

1. *Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them;*
2. *Select and use appropriate statistical methods to analyze data;*
3. *Develop and evaluate inferences and predictions that are based on data;*
4. *Understand and apply basic concepts of probability* (NCTM, 2000, p. 48).

1.3. STUDENT UNDERSTANDING

The implications of the new curricula for the professional development of teachers and for the development of materials and activities for different grade levels were many. What did students know about the new topics? How did understanding develop over time? What alternative conceptions existed to complicate learning? These and similar questions provided the motivation for research on students' initial and developing understanding of the topics in the curriculum over the next decade.

This research has added much to our understanding of students' developing ideas on particular topics, such as average (Cai, 1995, 1998; Mokros & Russell, 1995), sampling (Watson & Moritz, 2000a), chance measurement (Metz, 1998), inference (Gal & Wagner, 1992), association (Batanero, Estepa, Godino, & Green, 1996; Moritz, 2000) and graphing (Friel, Curcio, & Bright, 2001; Mevarech & Kramarsky, 1997), as well as for the areas of probability and data handling more generally (Jones, Langrall, Thornton, & Mogill, 1997; Jones, Thornton, Langrall, Mooney, Perry, & Putt, 2000). These studies were based on interviews, moderate-sized surveys, or reviews of the literature. As well, large-scale surveys have added to the store of information on understanding of these topics (Watson & Moritz, 1998, 1999a, 2000b, 2002; Zawojewski & Shaughnessy, 2000).

In addition to research specifically related to topics listed in the curriculum, there were calls from Green (1993) and Shaughnessy (1997) for research into students' understanding of variation as the underlying factor that creates the need for the chance and data curriculum in the first place (Moore, 1990; Wild & Pfannkuch, 1999). Work in this area (e.g., Shaughnessy, Watson, Moritz, & Reading, 1999; Reading and Shaughnessy, 2000, in press; Watson & Kelly, 2002) has brought about the development of further instruments specifically to measure how understanding of variation is displayed in association with understanding of the other topics in the chance and data curriculum (e.g., Watson, Kelly, Callingham, & Shaughnessy, 2003).

1.4. FURTHER CURRICULUM CHANGE

Within a decade of the initial changes to the mathematics curriculum in Australia (AEC, 1991, 1994), new, more general moves were taking place in the school curriculum. Under descriptors such as *New Basics* (Education Queensland, 2000) and *Essential Learnings* (Department of Education Tasmania, 2002) a more integrated approach to all areas of the curriculum has focussed attention on critical skills linking literacy, numeracy, and information technology through a curriculum addressing cultural, aesthetic, scientific, and social issues in a holistic fashion. These moves force a serious consideration of the links between the two antecedents described in the opening paragraphs. The place of chance and data in the mathematics curriculum suggests a consequent need for a wider appreciation of numeracy and of the part to be played by statistical skills in relation to social and scientific thinking based on both literacy and numeracy. As statistical understanding is the foundation for many of the decisions made in society today (Wallman, 1993), aspects of statistical literacy – the application of statistical understanding in context – will be placed at the intersection of literacy and numeracy and will be essential to meeting the goals of the new curricula, particularly those associated with citizenship (Steen, 2001).

2. THE CONSTRUCT OF STATISTICAL LITERACY

The emergence of a description of a construct of statistical literacy has taken place within the contexts described in the previous section. It has links to the parallel development of descriptions of numeracy, quantitative literacy, critical literacy, and adult literacy, as well as the chance and data elements of the school curriculum. Although the building blocks for statistical literacy are found within the mathematics, chance, data, and literacy components of the school curriculum, for students to become statistically literate they also need to interact with a variety of contexts as they take their place as consumers of information in the adult world. Hence contexts also must be included in a final theoretical description of the construct. This provides a link from statistical literacy to the other notions of quantitative literacy, adult literacy, and critical literacy. As well, a final description is likely to acknowledge that growth is hierarchical in nature.

2.1. EMERGENCE OF THE STATISTICAL LITERACY GOAL

A somewhat apocryphal prophecy from H. G. Wells at the beginning of the twentieth century has been a starting point for many who argue for a high status for statistical literacy: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write" (quoted in Castles, 1992, p. v). This claim is highly significant in terms of the move to an information society in the last third of the twentieth century as recognized in traditionally agriculture-dependent Australia by Jones (1982, p. 173): "Australia is an information society in which more people are employed in collecting, storing, retrieving, amending, and disseminating data than producing food, fibres and minerals and manufacturing products." Steen (1997, p. xv) reflected a similar view for a North American audience: "As information becomes ever more quantitative and as society relies increasingly on computers and the data they produce, an innumerate citizen today is as vulnerable as the illiterate peasant of Gutenberg's time."

These concerns were echoed in the desires of national governments, for example in Canada, the United States, and Australia, to survey the adult literacy skills of their citizens (Statistics Canada and the OECD, 1996; Dossey, 1997; McLennan, 1997). Although using the phrase "adult literacy" these surveys had three components: prose literacy, document literacy, and quantitative literacy. The terminology implies that numeracy skills would only be required for the last literacy, however, as is seen in the items used (e.g., McLennan, 1997), numerical information is embedded in many prose tasks and in most graphical and table-based documents to be interpreted. Further, basic statistical interpretation skills are required for many of the document and quantitative literacy tasks, for example summarizing information in pie charts and bar graphs, and working with percents and averages (e.g., Dossey, 1997; McLennan, 1997). Interest in adult literacy, at least implicitly, acknowledges a debt to statistical thinking/literacy. The stage is hence set for a consideration of the nature of the contribution of statistical literacy to the educational milieu.

For her Presidential Address to the American Statistical Association a decade ago, Katherine Wallman chose the topic of statistical literacy. In her brief definitional summary she focused on the application of understanding of the type developed during the school years for people as users rather than creators of statistics.

> *'Statistical Literacy' is the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions*. (1993, p.1)

The two dimensions of statistical literacy – public and private – introduced by Wallman are significant when thinking of motivating learning. Statistical literacy is not only important to our society as a whole; it is also relevant to the individual members of society as they make decisions in their personal lives based on information and risk analysis provided by others in the community. Decisions related to where to live, what type of employment to seek, whether to gamble, or what car to buy may be influenced by data provided from outside of one's individual experience.

More recently in summarizing the current state of understanding concerning adult statistical literacy, Gal (2002) suggested that the requirements are contained in the following two components:

> (a) *people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant,*
>
> *(b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions*. (pp. 2-3)

These components do not rely on sophisticated topics from the senior secondary curriculum such as standard deviation or hypothesis testing. They are built up as part of the chance and data curriculum, leading toward an expectation of critical thinking in many contexts across the school curriculum. This is envisaged in Australia's *National Statement*, which claims that "students should learn to question the assumptions underlying data collection, analysis and interpretation and the reasonableness of inferences and conclusions" (AEC, 1991, p. 164). It is also found in three extracts from the first NCTM statement of standards (1989):

> *In particular, citizens must be able to read and interpret complex, and sometimes conflicting information* (p. 5).

An understanding of probability and the related area of statistics is essential to being an informed citizen (p. 109).

A knowledge of statistics is necessary if students are to become intelligent consumers who can make critical and informed decisions (p. 105).

The transition from the components of a statistical investigation as set out, for example in the NCTM (2000) standards, to what is needed for survival in the world outside the classroom is illustrated in the New Zealand mathematics curriculum (Ministry of Education, 1992). "Interpreting statistical results" is one of three themes for statistics education and among the learning experiences is the following: "Investigating ways in which statistical information is presented in the media and other sources, and recognizing and identifying sources of deception in misleading graphs and their accompanying statements" (p. 189).

The use of the phrase "critically evaluate" by both Wallman (1993) and Gal (2002) in their descriptions of statistical literacy suggests a link to the area of critical literacy more generally. Adapting to statistical literacy the discussion of Luke and Freebody (1997), which related to the social practice of reading, four roles of a statistical text user are evident: the role as a code breaker, e.g., understanding the basic terminology of statistics; the role as a text participant, e.g., using knowledge to make sense of data, graphs, and chance claims embedded in text; the role as a text user, e.g., using data, graphs and chance concepts in particular social contexts; and the role as a text analyst, e.g., critically reading and seeing text as ideologically framed and constructing a position in relation to data driven claims. These categories of involvement mirror the views of Frankenstein (2001) in discussing the goals of a "critical mathematical literacy" curriculum. They are also useful in considering progression through a hierarchy of understanding.

2.2. THE HIERARCHICAL NATURE OF STATISTICAL LITERACY

Taking into account the rich and complex description of statistical literacy in the preceding section, the aim of this study is to explore and document the hierarchical stages associated with the goal as described for example by Gal (2002). Based on responses to tasks associated with statistical literacy, there are two aspects of understanding that can be described: increasing structural complexity and increasing statistical appropriateness. Two models were employed in this regard for the current study. The first was based on a cognitive framework (Biggs & Collis, 1982, 1991), which provides a structural hierarchy for responses: (i) *Prestructural* responses do not address elements of the task; (ii) *Unistructural* responses employ single elements of the task and do not recognize conflict should it occur; (iii) *Multistructural* responses employ elements in a sequential fashion and recognize conflict if it occurs but are unable to resolve it; (iv) *Relational* responses create connections among elements to form an integrated whole and resolve conflict should it occur.

The second framework was based on the work of Watson (1997) and Gal (2000) in relation to the expectations for statistical literacy of students when they leave school to participate in society. Watson suggested a three-tiered framework for statistical literacy including (i) the understanding of basic statistical terminology, (ii) the understanding of terminology when it appears in social contexts, and (iii) the ability to question claims that are made in context without proper statistical justification. This framework, in conjunction with the cognitive hierarchy, was employed for example to describe student understanding of sampling (Watson & Moritz, 2000b) and, within each tier, understanding was displayed in terms of (i) more sophisticated definitions of sample, (ii) greater engagement with samples in social contexts, and (iii) the emerging ability to question inappropriate claims made based on samples in newspaper articles. Gal provided a similar framework that included the element of motivation and the ability to communicate reactions. In the current study motivation was not addressed because the format of the items did not lend itself to any substantive measure of motivation. Overall the tasks used in this study, however, directly addressed terminology, or basic statistical skills required to address issues, associated with the first tier of Watson's framework; or were grounded in school-based curriculum or social contexts, providing the opportunity for students to demonstrate their understanding, consistent with the second tier of Watson's framework; or were based on articles from the media, which gave opportunity for critical questioning of claims, the goal of the third tier of the framework.

In this study, it was felt that the ability to consider variation as well as concepts associated with the components of the chance and data curriculum (Holmes, 1980), and ability to interact with the contexts presented would be significant aspects of the statistical literacy construct, reflecting Gal's (2002) description of statistical literacy. This includes the requisite mathematical terminology and statistical skills appropriate to the task. Previous studies (e.g., Watson et al., 2003; Watson & Moritz, 1998, 1999a, 2000b) suggested that the expectation of a hierarchical ordering of observed responses was reasonable. Gal's (2000) concern about ability to communicate reactions was felt to be handled in a hierarchical fashion by the structural model (Biggs & Collis, 1982) that was used.

3. MEASUREMENT ASPECTS OF THE RESEARCH

Assessment instruments, such as the questionnaires used in this study, are designed to measure a particular attribute or ability. All of the items used in this study, for example, address students' understanding of aspects of chance and data, albeit in different situations. The aim of this current study is to explore the mapping of all of the items onto a hypothesized single underlying variable of statistical literacy. One approach to this problem is to use Rasch modelling techniques (Rasch, 1980).

Rasch models are a set of measurement models coming under the general heading of Item Response Theory (Stocking, 1999) that have been widely used in surveys such as the Third International Mathematics and Science Survey (TIMSS) (Lokan, Ford, & Greenwood, 1997). They use the interaction between persons and items to estimate the probabilities of response of each person to each item. This process produces a set of scores that defines the position of each item and each person against the underlying variable or construct. The unit of measurement is the logit, the logarithm of the odds of success. The specific model used in this study is the Partial Credit Model (Masters, 1982), which allows for items that have a number of hierarchical scoring categories. This model has been shown to be appropriate for use with items that have been coded using hierarchical cognitive taxonomies, such as that of Biggs and Collis (1982, 1991) (e.g., Wilson, 1990, 1992).

Although Rasch methods have been widely used in the area of school mathematics generally (e.g., Lokan, Ford, & Greenwood, 1997; Wilson, 1990, 1992) and in the area of adult literacy including quantitative aspects (Kirsch, 1997), very little research has used Rasch methods specifically in relation to statistical concepts. Izard (1992) considered some of Green's (1982) data on students' responses to probability items and confirmed Green's hypothesised hierarchical structure for an English data set, although with the suggestion of a possible additional level based on data gathered in Quebec, Brazil, and Hungary. Gagatsis, Kyriakides, and Panaoura (2001) used Rasch techniques to suggest levels of understanding in relation to probabilistic concepts found in the Cypriot school curriculum. Description of items suggested a theoretical basis for their construction. Reading (2002) also used Rasch methods to confirm a profile of statistical understanding that had been developed using the cognitive development model of Biggs and Collis (1982, 1991). Her profile included aspects related to data collection, data tabulation and representation, data reduction, and interpretation and inference.

Earlier research involving some of the items to be used in the current study employed the Partial Credit Model (Masters, 1982) on 44 items measuring understanding of chance and data with a particular emphasis on variation (Watson et al., 2003). Based on responses from 746 students in grades 3, 5, 7, and 9, the fit of the items to a unidimensional model of variation was acceptable. Analysis of the variable map, a diagram of the item difficulty and student ability distributions produced by the Quest software (Adams & Khoo, 1996), suggested four levels of understanding associated with appreciation of variation in chance, variation in data, and variation in sampling. All of these studies confirmed specific hierarchies identified through qualitative analysis. The current study differs, however, in that it hypothesises a variable, statistical literacy, which comprises a wide range of statistical understanding and skills. In this sense it is an exploratory rather than a confirmatory study, which aims to postulate the existence of an hypothesized variable, rather than confirm the presence of a construct previously identified by other means.

3.1. TEST EQUATING

Rasch modelling techniques provide a way of linking, or equating, tests through the use of common items or common persons (Bond & Fox, 2001; Kolen, 1999). Generally this is done to ensure that two tests that purport to measure the same construct can be used interchangeably to measure student performance. In this study, however, the techniques were used to bring together questionnaires that were designed to measure both common and varying aspects of a theorised construct of statistical literacy. If it can be shown that when different tests are combined, the items in the various tests work together in a consistent and predictable fashion, that is fit the model, this provides evidence of a single underlying dominant variable and it can be argued that they are likely to be measuring the same construct (Bond & Fox, 2001). Placing all items together on the same scale then provides an opportunity to examine the nature and validity of the underlying theorised construct.

Where different tests are used to measure the same underlying construct, equating is used to align these different measures so that all the items, and the students who attempt them, can be consistently described with the same measures. In addition, in relation to students, their performance on different tests can be mapped onto a single scale. In this study, the different questionnaires are hypothesised to address component parts of a variable termed “statistical literacy”. To examine the hypothesis that this is indeed a single, unitary variable, the responses to different items across the questionnaires needed to be linked, or equated, in order to provide a basis for the assertion that these are part of the same construct, and for the interpretation of that construct. The process is described in some detail because the later variable interpretation is dependent on this.

Tests can be equated through Rasch techniques if an anchor or link-set of items can be found or if the same people undertake the questions (Kolen, 1999). The link set of items needs to be common across two or more questionnaires but it does not need to be common to every questionnaire. Linacre (1997) refers to this as connectedness: Providing that all items, regardless of which questionnaire they appear in, are directly or indirectly connected to every other test item in the pool, then all items can be mapped onto a single underlying scale. Links between tests that have some common items enable a “common item equating design”.

In this study, 24 common items were used in the 1993, 1995 and 1997 questionnaires. Of these, eight were used across all grades, and all others were used across at least two grades. These items provided a means of linking all the different questionnaire forms from these years across the range of school grades. Within the 2000 questionnaire, 20 items were used across all grades, and, apart from one, all others were presented across at least two grades. Hence the different questionnaire forms within this year were also linked across all grades. In addition, there were four items that were common across all questionnaires, 1993, 1995, 1997, and 2000. These four common items linked the 2000 questionnaire to the earlier questionnaires and provided a means of meeting Linacre’s (1997) connectedness criterion. There is no recommended minimum number of items for linking using the Rasch partial credit model, although it is usually suggested that the link items have a range of difficulty (Kolen, 1999). In this instance the range of task-step difficulty was from -1.72 logits to 3.02 logits, providing a wide difficulty span along the variable. Although more common items would provide a stronger link, the link established here appears satisfactory for this initial exploration. Every item from every questionnaire form administered across the years was hence connected directly or indirectly to every other item. The process of test equating for this study is shown diagrammatically in Figure 1.

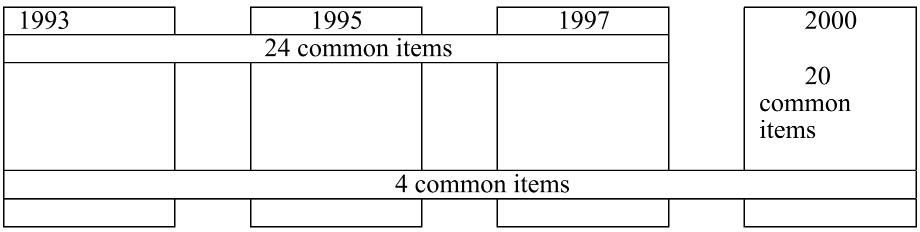


Figure 1. Linking process for equating questionnaires

3.2. VALIDITY OF THE UNDERLYING CONSTRUCT

Validity is the extent to which the inferences drawn from scores on a test, or any other form of assessment, are supported empirically and theoretically. Traditionally, evidence of three kinds of validity was gathered (Messick, 1989). Criterion validity compared the scores on a test with a different measure of the same variable within a short time frame (concurrent validity) or over time (predictive validity) (Anastasi, 1988). Content validity was considered in relation to the possible universe of content that could be included in the test, with the aim of avoiding construct under-representation (Messick, 1989). Construct validity was concerned with the underlying cognitive processes, and Messick (1989) argued, should be considered to be of overarching importance.

Many different sources of evidence may contribute to the establishment of construct validity, but the validity is strongest when the fit of the information gathered from the assessment is explicitly related to an underlying theoretical perspective (Messick, 1989). To establish construct validity two aspects must be considered. First, there should be some underlying substantive theory on which the construct is based. Second, the measuring instrument should address this explicitly – that is, it should be designed to measure a theoretical construct. In this situation, it has been claimed that the Rasch model (Rasch, 1980) is a useful approach to establishing construct validity (Fisher, 1994).

Rasch models make three assumptions. The first of these is that the underpinning construct is unidimensional. Although different domains or categories may be part of the assessment, these all form a single dimension. Second, the variable is hierarchical, or has direction. The construct is measurable with an additive unit of measure that is repeated along the variable; that is, the scoring or coding assigned to each item describes an increasing "quantity" of the construct. The third assumption is that each item is independent of all other items. This means that responses to later items do not depend on a correct response to an earlier item. The extent to which these three assumptions are met is a measure of the validity of the underlying construct (Wright & Masters, 1982).

The purpose of this study is to establish support for the existence of a unidimensional scale that provides interpretable information about hierarchical levels of a hypothesized construct of statistical literacy. In order to do this, the three assumptions must be tested. There are a number of measures provided by Rasch measurement techniques that may be used to ascertain the measurement characteristics of the variable, and test the assumptions outlined above. Two that are widely used are Infit Mean Square (IMSQ) and Item Separation Reliability (R_I). If the measurement characteristics suggest that the assumptions are met, then a substantive interpretation of the construct provides further information about its nature. These processes together provide evidence of construct validity (Wright & Masters, 1982).

The Infit Mean Square is a measure of fit to the model and has an expected value of 1. In practice, values within the range 0.77 to 1.3 provide acceptable fit to the model (Adams & Khoo, 1996; Keeves & Alagumalai, 1999). There are two ways in which items can misfit. The first is "overfit" – the items discriminate too sharply. A group of items behaving in this manner is indicative of a different construct being measured; that is, the scale is not measuring a single construct, and it thus violates the first assumption. A second reason for overfit is that there is dependence among the items, violating the third assumption. This was of particular interest in this study as many of the items shared a stimulus. These were, however, constructed as "superitems" (Collis, Romberg, & Jurdak, 1986; Cureton, 1965) in which each question asked, although referring to the same stem, did not depend on a correct response to a previous question. It was necessary, however, to check this aspect. Finally, items may "underfit" – behave in a random fashion. This occurs if responses are random, or raters apply the scoring scale inconsistently. In the situation in this study, the items had been scored by a small number of expert raters so that inconsistency would be expected to be minimal. Consistent misfit could, however, also indicate that the second assumption of an additive unit of measure is not met.

The Item Separation Reliability measures the extent to which the items are separated along the scale. Items that all cluster closely together do not provide sufficient information about the direction and meaning of the variable, the second assumption. This measure indicates how well the items are separated in difficulty and can thus provide information about a range of development along the variable (Wright & Masters, 1982).

Once a scale has been produced, it is necessary to interpret the underlying construct. This is a process of criterion referencing (Glaser, 1963, 1981). A criterion is the point at which a student moves from one level of competence to another and is sometimes called the step or threshold. By considering the common demands made on knowledge and understanding by items that appear near each other on the scale produced by Rasch analysis, required levels of performance at particular points along the variable can be identified. The thresholds or steps at which pertinent aspects of competence evolve from one level to the next can then be defined. Since the variable describes a continuum of competence, determining the thresholds becomes a process of judgement that must be justified and understood in terms of the relevance of the criteria to the underlying variable (Eisner, 1993; Wright & Masters, 1982). This provides a conceptual interpretation of the variable that can be compared with the theorised construct. This process of evaluating the fit of the information obtained from a test, or, in this instance questionnaires, to the theoretical rationale for the interpretation of the test outcomes, or scores, establishes strong construct validity (Messick, 1989).

4. METHODOLOGY

4.1. SAMPLE

The data used were collected from questionnaires completed by 3852 students in Tasmania, conducted in 1993, 1995, 1997, or 2000. Although it could be argued that the teaching of chance and data within the state could have changed over this period, there were indications (e.g., Watson & Moritz, 1998) that there was no change at least for some items between 1993 and 1997.

As this was a preliminary study to explore the nature of the underlying variable, it was the individual understanding shown on items that was of interest, not a comparison among students. As such the research design was cross-sectional rather than longitudinal, and this initial study did not attempt to map the rate of students' progress along the variable from year to year. Rather it was intended to establish the nature and validity of the underlying construct of statistical literacy. Although some students were involved in longitudinal data collection, only data from initial questionnaires were used in this study. The sample distribution across grades is given in Table 1.

Table 1. Sample Size for Each Grade

Grade	Frequency	Percent
3	1039	27.0
5	421	10.9
6	881	22.9
7	239	6.2
8	207	5.3
9	1065	27.6

The uneven distribution reflects the target grades of the previous questionnaires. Overall, however, the sample covers a wide range of ages and school curriculum experience. The number of students in each grade answering each item is given in Appendix B.

4.2. ITEMS/TASKS

The items used in this study were devised to measure various components of the chance and data curriculum as it was introduced to Australian schools in the 1990s. They reflect specific straightforward aspects of content, such as sampling, average, chance, and graphs. They also reflect the contexts within which the subject matter is applied, for example, school-curriculum based social contexts, such as conducting a survey in school, and less familiar media-based contexts specifically related to statistical literacy. Issues of acknowledging variation, drawing inferences, and questioning claims were interwoven with the content and context for some of the items. The items hence covered the range of potential contributing elements to the construct of statistical literacy as outlined earlier.

Altogether 80 items were used, 48 of which were used in the previously noted study of Watson et al. (2003). There were 44 items reported in Watson et al. (2003), four of which appeared in two forms with different numbers of imagined trials used for different grades (see e.g., SP2A, SP2B in Appendices B and C). Thirty-six of the items were used in data collections that took place in 1993, 1995, and 1997. The items are detailed in the various studies that used them to explore particular topics in the chance and data curriculum. They included curriculum content-based items, such as items addressing basic probabilities, basic table reading, variation in a spinner scenario, understanding of stacked dot plots, and interpretation of a pictograph. Other items were closely related to the school curriculum but placed in some type of social setting, for example, items on conditional and conjunction probabilities, on conducting a survey of children in a school, on risk from taking a medicine, on average number of children in a family, on the median (or average) of a set of science measurements, on selecting a new car, on actors' performances, and on the story conveyed by a stacked dot plot.

These items were adapted from items used in previous research (Fischbein & Gazit, 1984; Garfield, 2003; Green 1982, 1983a; Jacobs, 1999; Konold & Higgins, 2002; Konold & Garfield, 1992; Nisbett, Krantz, Jepson, & Kunda, 1983; Pollatsek, Well, Konold, Hardiman, & Cobb, 1987; Tversky & Kahneman, 1983; Watson, Collis, & Moritz, 1994, 1997), or developed as part of previous local research (Torok, 2000; Watson, 1998a; Watson et al., 2003). Other items were based on contexts taken directly from newspaper extracts, such as the "average" house price, biased sampling of populations, odds and independent binomial events in sporting contexts, interpretation of a misleading picto-bar graph, interpretation of an incorrect pie graph, ranking chance language used in headlines, creating a graph to represent a cause-effect claim and questioning the claim, and interpreting conditional language about the effects of smoking (Moritz & Watson, 1997, 2000; Moritz, Watson, & Collis, 1996; Watson, 1998b, 2000; Watson & Moritz, 2003). Finally the scale included questions on the definitions of the terms "average", "sample", "random", and "variation" (Mortiz, Watson, & Pereira-Mendoza, 1996; Watson et al., 2003). All items, codes, and example responses are given in Appendix C. The codes reflect the theoretical frameworks introduced earlier (Biggs & Collis, 1982, 1991; Watson, 1997).

For clarity, in this study the term "task" is used to refer to the total item as it was presented to the students, and "task-step", to refer to the level of response denoted by the coding. Thus for an item that asked about guns in United States high schools, M7CH is the task, and the levels of response, coded as 0–4, are the task-steps, shown as M7CH.1 to M7CH.4, that appear on the variable map in Appendix A.

4.3. ANALYSIS

The process for calibrating and equating the questionnaires, undertaken with Quest software (Adams & Khoo, 1996) is summarized as follows:

1. The difficulty levels of four items that had the highest number of respondents, SMP4, DIE7, HAT8, and BOX9, were calibrated using a data set that included no missing data. The purpose of this step was to provide a stable set of difficulties as an "anchor set" against which all other difficulties could be estimated.
2. Using the difficulty levels from these items as an anchor, and the full data set of all 80 items and 3852 students, the different questionnaire forms were calibrated and equated in one operation (Adams & Khoo, 1996). The four items common across all questionnaires (DIE7, BOX9, RAN3, and M4DR) linked the earlier questionnaires to the 2000 questionnaire.
3. Difficulty levels and fit statistics of all steps on all items were obtained and written into a file that was used for subsequent analyses.

This procedure produced a variable map, showing the task-steps and persons distributed on a single scale, a fit map of the items to the model, and a classical item analysis. Common characteristics of task content and skill were identified by undertaking an audit of the demands of the task-steps that were close together on the map. It was then possible to describe trends in the development of concepts and cluster these together with descriptions across the topics covered in the questionnaires. This process is one of professional judgment and discussion similar to procedures suggested by Miles and Huberman (1994). Six hierarchical levels were identified as a convenient way of distinguishing overall steps in the progress along the variable for the underlying construct of statistical literacy. These levels are identified on the map in Appendix A.

5. RESULTS

5.1. MEASUREMENT CHARACTERISTICS OF THE VARIABLE

Table 2 provides estimates of Item and Case Separation Reliabilities and overall fit measures for the 80-item instrument. These separation reliabilities describe how adequately the tasks describe the underlying variable, and the extent to which the student cohort is spread out across the continuum (Wright & Masters, 1982). The Item Separation Reliability is high, suggesting that the tasks do indeed describe a spread of difficulty along the variable that will allow the underlying construct to be described. The overall fit to the model is also high, at the expected value of 1.00, suggesting that the tasks form a hierarchical unidimensional scale. Similar comments apply to the Case Separation Reliability and fit statistics, suggesting that the tasks used are appropriate for this cohort of students. The Cronbach Alpha is a measure of internal consistency taken from classical psychometric theory, and is also obtained from Quest analyses. This value is high at 0.85, showing that the scale meets not only Rasch measurement standards but also a classical standard of reliability.

Table 2. Reliability and Fit Indices for 80 Items

Item Separation Reliability	0.99
Item Infit Mean Square	1.00 (S.D. 0.13)
Case Separation Reliability	0.86
Case Infit Mean Square	1.00 (S.D. 0.42)
Cronbach Alpha	0.85

Appendix A contains the variable map of student ability and item difficulty. The measurement scale, from –4.0 to +4.0, on the left hand side is in logits, the logarithm of the odds of success. The crosses on the left hand side each represent 12 students and the tasks are presented on the right hand side. This map shows the relationship between the tasks and the students who undertook them. The representation is limited by space and printing requirements, and thus students are grouped together, in this case in groups of 12, according to their estimates of ability. Where there were fewer than 12 students at any particular ability estimate, no cross is shown. The map does not suggest that where there are no crosses there were no students – only that there were fewer than 12 students at any particular level of ability (in logits) to be grouped together and thus they do not appear in this representation. Each task having more than one coding level is described in terms of m.n, where m is a 4-letter-digit task identifier and n represents the coding level or task-step. Thus M2PI.2 is media task 2, Pie chart, coding level 2. Task identifiers, along with a statement of the question, information on codes, and typical responses, are found in Appendix C. Although the variable map in Appendix A presents both item and student data, the analysis presented here focuses on the right hand side of the map, which contains the item distribution along the variable. Only two tasks fell outside the acceptable limits of fit, both showing underfit. The first of these, AV12, required students to select from a list of alternatives to explain the meaning of “2.2 children per family”. Randomness in this situation is likely to be because students were guessing. The second task showing misfit, DIE2, asked students to predict the outcomes of 60 tosses of a single die and explain their answers. The most likely reason for underfit here is inconsistent coding patterns. Detailed examination of individual item statistics confirmed the high level of item fit for all tasks except the two problematic ones, and measurement errors were very small for each task at each step of difficulty (which corresponded to a change in coding). In general, tasks fitted the model well, providing a basis for interpretation of the scale. In particular, there was no evidence of overfit, as would be expected for a multi-dimensional construct.

5.2. QUALITATIVE INTERPRETATION OF THE UNDERLYING VARIABLE

Having provided evidence suggesting that the variable was unidimensional, the next step was to interpret it in relation to the tasks that were used to measure it. Six levels were distinguished along the continuum, shown by horizontal lines in Appendix A, indicating clusters of task-steps that when analysed by content (e.g., mathematical skills, statistical concepts, context) suggested common characteristics in a hierarchical sequence of statistical literacy. In some instances, two task-steps appeared within the same level for the same task (e.g., TBL5.2 and TBL5.3 both appear in Level 3 of the continuum). Usual measurement practice would be to recode these and collapse the two categories into one. Here, however, we have chosen to maintain the original coding because the qualitative interpretation suggested that students were responding in different ways. The implications of this are addressed further in the Discussion section. In Appendix C brief descriptors of the code levels are shown and these are amplified in the following sub-sections with examples of content and context. Summary descriptions of the characteristics of statistical literacy displayed in the tasks at the six levels are presented in Table 3.

Table 3. Statistical Literacy Construct

Level	Brief characterization of step levels of tasks
6. Critical Mathematical	Task-steps at this level demand critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.
5. Critical	Task-steps require critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance, and appreciation of variation.
4. Consistent Non-critical	Task-steps require appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics.
3. Inconsistent	Task-steps at this level, often in supportive formats, expect selective engagement with context, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas.
2. Informal	Task-steps require only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, graph, and chance calculations.
1. Idiosyncratic	Task-steps at this level suggest idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

5.3. LEVEL 1 – IDIOSYNCRATIC

Task-steps associated with concepts such as average or with the definition of terminology do not appear at Level 1, indicating lack of engagement with their associated ideas and contexts. At this level personal beliefs and experience dominate, for example, with the task-step to identify unusual features of an article about guns in high schools in the United States satisfied by “people should not have guns” (M7CH.1) and a task-step on new car selection satisfied by the alternative “rely on friends” (SM19.1). In terms of data representation in straightforward contexts, task-steps demand only reading specific values from a simple two-way un-nested table (Q10A.2, TBL1), choosing the highest value from a row or column in a table (Q10C.2, TBL2), and determining a frequency and a difference from a pictograph (TRV1, TRV2).

Chance task-steps at this level suggest idiosyncratic beliefs, for example in drawing names from a hat of boys and girls (HAT8.1), "a girl because the teacher is a girl"; in describing odds from a newspaper article on a sporting event (M3OD.1), "it is the current score in the game"; in describing the chance of a 50-50 spinner landing on a certain half (SP1.1), "bad chance" or "1 in 10". For a task to interpret "a 15% chance of getting a rash" (CH11.1), colloquial alternatives "good chance" or "hardly any chance" appear at this level.

The only task-steps associated with inference appearing at Level 1 suggest concern for certainty rather than uncertainty (TRV5.1), refusal to predict or belief in "no change" (TRV6.1). Similarly for tasks related to variation, only those task-steps which require basic acknowledgement of change, as in a travel-to-school graph "won't look the same every day" (TRV3), or idiosyncratic predictions of chance outcomes without justification (SP2A.1, SP2B.1, DIE2.1), appear at this level.

5.4. LEVEL 2 – INFORMAL

Although the task-steps appearing at Level 2 demand engagement with more contexts, the engagement is still intuitive, non-statistical, or reflective of irrelevant aspects of the task context. Some task-steps require single ideas, for example in terminology associated with sampling (SMP3.1, SMP4.1) and average (AVG2.2, ME13.1). Other single aspects of sample are encompassed in a survey planning task in a school-based context, where the task-step requires features such as "ask 400", "ask everyone", or "ask the people I meet", without considering the need to represent the population (MVE1.1). For a task on new car selection (SM19.2), the alternative "it doesn't matter whether a person uses friends' advice or data from a consumers' report" appears at this level.

At Level 2, table reading task-steps demand comparing cells to determine the highest or most even counts (Q10D.2, Q10E.2) and finding a total greater than 100 (TBL4). Graphing task-steps require identifying the smallest data value in a stacked dot plot (SP6) but only idiosyncratic arithmetic strategies for working out prices from a picto-bar graph from the media (M9C.1, M9D.1). In relation to chance, a simple 50-50 spinner task (SP1.2) demands a correct response equivalent to a half, but no recognition of variation. "Anything can happen" is a justification accepted for task-steps associated with picking a boy's or girl's name from a hat (HAT8.2) or with comparing two boxes of marbles for the chance of choosing a single marble of a particular color (BOX9.2). In a media task of ordering chance headlines by likelihood, the step at this level only requires placement of phrases in the appropriate half of the 0-1 number line (M1CH.1).

In terms of inference at this level, task-steps are satisfied by story-telling (TRV6.2) or pattern recognition (TRV4.1) in predicting from a pictograph. A task to judge the better of two stacked dot plots for telling a story of how long families in a class have lived in a town (TWN3.1) accepts the inappropriate choice with reasoning such as "it is well set out". Task-steps associated with variation require only appropriate "surprising" results for repeated spins of a 50-50 spinner (SP4B); too much, too little, or lop-sided predictions for repeated spinner trials (SP5A.1, SP5B.1); and patterns or strict chance in predicting 60 die outcomes (DIE2.2).

5.5. LEVEL 3 – INCONSISTENT

At this level task-steps require more engagement with context than at the previous two levels but this is dependent to some extent on the format of items, which may provide added support. Although more features are demanded, the statistical ideas required are qualitative rather than quantitative and appropriate conclusions may not be accompanied by suitable justifications.

In relation to sampling, the task-steps for tasks associated with judging plans for a school survey require suggestions in context but focusing on peripheral rather than salient features, for example a method is good because it is "easy" (MVE5.1, MVE6.1), "not too many" (MVE3.1), or "large"

(MVE4.1); or a method is bad because "more people are needed" (MVE2.1), "the wrong people might get picked" (MVE2.1), or "they're a bit young" (MVE4.2). For a task asking for methods of selecting four students to lead a parade, a representation say of boys and girls (TBL5.2) or a random method (TBL5.3) is acceptable at this level. A task-step for commenting on a voluntary poll about legalizing marijuana only requires recognition that people could be lying or the sample size is too large (M4DR.1).

For data representation the task-steps at this level demand at least one summary statement when interpreting stacked dot plots (TWN1.2, TWN2.2), a basic unlabelled graph or a labeled graph with no association when association is intended (M8GR.1), or recognition of non-salient unusual features of a media bar graph (BT1A.1). Average task-steps require colloquial expressions in an open-ended format (AVG1.1, M5AV.1).

Chance task-steps at this level generally demand qualitative rather than quantitative reasoning. Although a simple 50-50 spinner task with repeated spins requires an answer equivalent to "half" of the spins (SP2A.2), a task about drawing names from a hat requires recognition of "more girls' names in the hat" (HAT8.3), and a task about equality of dice outcomes only requires a justification of "anything can happen" (DIE7.2). A task-step supported by a selection of alternatives to interpret a "15% chance of getting a rash" demands an exact numerical interpretation (CH11.2). The language associated with ordering conjunction events appropriately is needed for two tasks at this level (CF15 and CP18), whereas demands for quantifying outcomes from four coin tosses at a sporting match are less stringent, values greater than a half being acceptable (M10A.1). A task-step for defining "random" requires single or multiple elements (RAN3.1, RAN3.2).

Few task-steps associated with inference or variation appear at Level 3. Although the appropriate choice is required in determining which of two stacked dot plots is better at telling a data-based story, justification for the choice is not needed (TWN3.2). Only recognition of chance, not variation, is required in predicting repeated outcomes with a 50-50 spinner (SP2A.2) and only a single aspect is demanded in defining the term variation (VAR.1). Improvement required in quantitative skills at this level is associated with task-steps requiring recognition rather than creation of appropriate responses (e.g., AV12.2, CH11.2).

5.6. LEVEL 4 – CONSISTENT NON-CRITICAL

The task-steps appearing at Level 4 demand a consolidation of appropriate contextual but non-critical engagement by students in various contexts. In terms of the Statistical Literacy Hierarchy discussed in Section 2.2, the task-steps require an understanding of social contexts that are not associated with critical questioning or partial context-only reasoning where critical thinking is the ultimate aim.

For the definition of sample, two aspects are required, such as "you have a small piece of something" (SMP3.2). The task-step associated with suggestions for surveying a school demands representative but not random methods (MVE1.2). Task-steps for evaluating other survey methods require peripheral or partial recognition of salient features associated with appropriate "good" or "bad" judgements (MVE2.2, MVE7.2, MVE4.3). The media task based on the less familiar context of a non-representative sample of United States high schools demands only contextual recognition, such as that people could be lying or the whole United States would be the same (M7CH.2).

Graph recognition task-steps demand the highest data value (SP7), the range of the data (SP8), a qualitative description of the shape (SP10), and appropriate reasoning for selection of the scale as the better of two stacked dot plots (TWN3.3). For media-based graph tasks, however, only partial recognition or representation is required, for example in criticizing a pie chart summing to 128.8% (M2PI.1) or graphing an association of heart deaths and car usage (M8GR.2). Average task-steps at this level require describing the mean or middle appropriately (AVG2.3, M5AV.2), and finding the mean of a small data set (AVG1.2), without recognition of the effect of an outlier.

Chance task-steps present a variety of contexts and demands at Level 4. A task to select which of two boxes with the same ratio but different numbers of marbles is more likely to produce a certain outcome demands appropriate proportional reasoning (BOX9.3), whereas one to justify belief about die outcomes requires only "same" chance reasoning (DIE7.3). Probability tasks set in a media context require a correct response for a single coin toss (M10B.2) but the same answer for four tosses (M10A.2), indicating lack of knowledge of compound events. An odds task-step accepts predicted scores (M3OD.2). Where language rather than numerical calculation is involved, task-steps demand appropriate ordering of chance newspaper headlines on a number line (M1CH.2) and appropriate interpretation of straightforward conditional statements (M6AB.2, M6D.1).

Except for the task of distinguishing between the appropriateness of two stacked dot plots (TWN3.3), the task-steps for inference at this level require limited recognition of the implications of representations, for example balancing information presented in terms of boys and girls in a pictograph (TRV4.2, TRV6.3) or reflecting a majority (TRV6.4). A media task-step for a suspicious cause-effect relationship demands only engagement with the context and questioning of data collection rather than questioning of the association (M8QU.1).

Task-steps dealing with variation in chance settings that appear at Level 4 demand a reason associated with variance in explaining differences in repeated sets of trials with a 50-50 spinner (SP3A.3, SP3B.3), and realistic variation in numerical predictions of outcomes for six sets of repeated trials (SP5B.2) and of outcomes for 60 tosses of a die (DIE2.3). The task-step for deciding the authenticity of sets of spinner trials requires both appropriate choices and reasoning (SP11.2). For media tasks, however, task-steps demand less sophisticated reasoning, with a media bar graph interpretation task requiring focus on single columns rather than comparisons across columns (BT1B.1) and the definition task requiring multiple relevant features, such as "variation means to change something" and "the weather is going to vary over the next few days" (VAR.2).

The task-steps at this level that demand consolidation of the mathematical and statistical skills include those associated with the mean, simple probabilities, and graph characteristics, all in straightforward settings. Task-steps require appreciation of setting but rarely critical questioning.

5.7. LEVEL 5 – CRITICAL

Task-steps at the top two levels of the statistical literacy construct demand similar critical thinking skills associated with the third tier goal of the Statistical Literacy Hierarchy. What distinguishes them is the level of mathematical skill required to engage in critical questioning. At Level 5 sophisticated use of proportional reasoning is not required, but in contexts, particularly familiar ones, critical thinking is otherwise expected, as in appropriate use of terminology, appreciation of variation, and qualitative interpretation of chance.

The task-steps related to defining a sample require the relating of several elements in describing a sample and its purpose (SMP3.3, SMP4.3). The task on surveying a school demands random methods or random methods combined with representation, such as "10 from each grade, 5 boys and 5 girls picked at random" (MVE1.3). For task-steps to evaluate three other suggested surveying methods – a random method, a choice of friends, and a booth for volunteers – appropriate decisions and statistical justifications are required (MVE2.3, MVE5.3, MVE6.3). For the task of selecting a car, the appropriate task-step of using the report on 800 cases is needed (SM19.3). For the task of assessing a voluntary poll on legalizing marijuana, focusing on the central issues, for example, the type of listeners to the radio or that only motivated people telephone the station, is required (M4DR.2). The task-step for assessing an article about access to guns by school students in the United States, however, only requires recognition of the non-representative nature of the sample with the support of an additional question about other regions of the United States (M7CH.3).

In terms of graphing at this level, task-steps require appropriate representations for a claim about the association of heart deaths and car usage (M8GR.3), representing the ability to handle two variables at the same time and show corresponding increases, or recognition of the error in a pie graph that sums to 128.8%, focusing on the total percent or the shapes of the segments of the graph in comparison to the percents they represent (M2PI.2). At this level, for the idea of average, there is the demand to find the median or mean of a small data set (ME13.3).

Chance task-steps at this level demand a consolidation of ordered estimates of conditional statements (CP16) and of giving appropriate "if ... then ..." statements for an embedded conditional statement in a newspaper article on smoking and wrinkles (M6D.2). In media contexts with mathematical skills required, however, task-steps require qualitative rather than quantitative recognition (M10A.3) or use of ratio without appropriate interpretation (M3OD.3). Few tasks on inference appear at Level 5, with a task-step for selecting actors by audition who later perform less well than expected demanding the choice of an alternative reflecting regression to the mean (Q20).

Two task-steps require appreciation of variation at Level 5. For a task predicting the outcomes of spinning a 50-50 spinner repeatedly, responses must spontaneously use words like "about" or "probably" in suggesting numbers of successes or phrases like "it will be close to half" (SP2A.3, SP2B.3). For a task to describe unusual features of bar graphs in a report on boating deaths, an increase or change in the data over time or acknowledgment of variation explicitly in the visual appearance of the graphs is required (BT1B.3).

5.8. LEVEL 6 – CRITICAL MATHEMATICAL

As noted previously proportional reasoning skills are demanded by many of the task-steps that appear at Level 6, particularly in chance or media contexts. As well task-steps require sensitivity to the need for uncertainty in making predictions and appreciation of subtle aspects of the language for some tasks.

In relation to sampling, detection of the two flaws in a survey method suggesting 10 students from a computer club, for example, "there are not enough people and they are selectively picked," is required (MVE3.3). The task-step concerning a sample from Chicago in relation to the United States (M7CH.4) requires the recognition of the non-representative nature of the sample, without any support. A task to suggest two methods to select children to lead a parade demands either two different random methods or a combination of random and representative methods (TBL5.4).

In terms of graphing, two summary statements involving the context, rather than just data reading, are required to describe stacked dot plots about how long families have lived in a town (TWN1.3, TWN2.3). The mode must be recognized in relation to a stacked dot plot (SP9). Finding errors in bar graphs about boating deaths is required (BT1A.2), as are appropriate rate calculations associated with a complex picto-bar graph (M9C.2, M9D.2). Recognition of outliers is demanded when calculating a mean (AVG1.3) and suggesting the median as the appropriate measure of middle in relation to house prices in the context of a newspaper article (M5AV.3).

At the highest level of statistical literacy task-steps require quantitative reasoning for chance tasks. For straightforward task-steps such as those involving outcomes for a single die and drawing names from a hat, numerical (e.g., fractions) rather than qualitative descriptions are demanded (DIE7.4, HAT8.4). For a classic fish-tagging task, proportional reasoning to obtain the solution of "2000" is required (Q17). For a task from the media on explaining odds, proportional reasoning and the correct direction for interpreting the result appear in the response (M3OD.4). For a task based on an article on tossing coins at the start of a cricket match, independence and correct calculations are required (M10A.4). Integrated descriptions for the term random are also demanded (RAN3.3).

Task-steps related to inference at this level reveal subtleties in thinking. Task-steps requiring predictions for a pictograph on how children travel to school, for example, demand inclusion of expressions of uncertainty, such as "probably a [new] girl comes by car – more girls get a car" (TRV4.3). A task concerning a newspaper article about heart deaths and car usage requires responses that ask the salient question about a cause-effect relationship (M8QU.2).

The mathematical/statistical skills demanded by task-steps at the highest level include proportional reasoning associated with ratio and appropriate part-whole interpretations, the ability to use rates in calculating costs, understanding of independence and its implications for calculating probabilities, an overall quantitative view of chance as probability, and a memory for terms such as "mode". Further some task-steps require an ability to account for subtleties in language and context.

These extended summaries of the levels of statistical literacy based on the tasks employed are intended to portray the detail and richness of the information obtained from the questionnaires. As every task-step code is described with examples in Appendix C, it is further possible to link task demands for every task-step displayed in Appendix A. At the other extreme, the summaries in Table 3 and the level labels in Appendix A are intended to provide brief indications of the differences among levels.

6. DISCUSSION

Following comments on the limitations of the study, the discussion will focus on five aspects of the outcomes of this research: the identification and exploration of a hierarchical construct of statistical literacy; the relationship of this construct to previous research; the complex nature of the statistical literacy construct, particularly in relation to context; implications for future research; and implications for classroom planning.

6.1. LIMITATIONS OF THE STUDY

The data used in this study were collected as parts of other research studies into student understanding of the chance and data curriculum over an eight-year period (Watson, 1994; Watson et al., 2003). Although initial indications over the first four years were that curriculum implementation produced no improvement on average performance on many items (e.g., Watson & Moritz, 1998, 1999a, 2000b, 2003), there is no corresponding analysis for items used again in the final year covered by this study. The purpose of this study, however, was to document the hierarchical nature of the statistical literacy construct using all available data from the studies, not to consider changes across cohorts, years, or individuals. Longitudinal data, for example collected from students in 1995 and 1997, were not included in this study.

Although the data used for this study reflected a wide range of age, ability, and socio-economic status, they do only represent the Australian state of Tasmania. Other cultural settings may result in students responding differently, particularly to context-based items. It is the belief of the authors, that the school students used in this study are likely to have experiences similar to other Australian students and to students in most western countries.

The relatively small number of linking items used in the Rasch analysis means that some caution needs to be exercised in interpreting the results of the analysis. Two other factors help to mitigate this concern. First, there is a large number of responses for these linking items across all years. Second, the structure of the resulting variable map in Appendix A is very similar to the structure of the corresponding variable map found by Watson et al. (2003) for the subset of items used in 2000. The relative placement of common items engenders confidence that the Rasch analysis produced as part of the current study is a reasonable suggestion of the hierarchical nature of the construct.

6.2. IDENTIFICATION AND VALIDATION OF A HIERARCHICAL CONSTRUCT OF STATISTICAL LITERACY

The findings from the application of the Rasch model suggest a unidimensional character of the variable. Fit to the model was excellent overall, and individual items also showed no overfit, which might have been expected if a multi-dimensional construct was being addressed.

The scale established from the 80 items had a high Item Separation Reliability (R_I = 0.99) and provided sufficient information to give a criterion-based hierarchical profile of the underlying construct, hypothesised as statistical literacy. The large item pool provided considerable detail about the variable without over-sampling particular concepts, and confirmed that mathematical skills and understanding of contexts, as well as content from the school curriculum, were all aspects of the same construct. These aspects are summarized in Table 3 and further discussed below.

This good model fit, together with the coherence of the interpretation of the underlying variable with the hypothesised construct of statistical literacy discussed earlier in this paper, suggests strong construct validity (Messick, 1989). It seems that the questionnaires that targeted varying aspects of the chance and data curriculum can, when combined, provide useful and interpretable information about hierarchical levels of statistical literacy.

6.3. RELATIONSHIP TO PREVIOUS RESEARCH

In comparing the results of the current study with the previous analysis of Watson et al. (2003) that focused on the construct of variation, several points can be made. In the earlier analysis a subset of the current 80 items was used and the particular focus on variation led to the identification of four levels of the underlying variation variable rather than six as in the current study. The spread of items along the variable was similar, however, with the top 11 items in the variation map appearing at Level 6 in this study, and the bottom 6 items appearing at Level 1. This, together with the good fit to the model, indicated that variation is a sub-domain of statistical literacy appearing across difficulty levels. The greater number of items used in the current study, particularly reflecting more curriculum-based chance tasks and more media-based social contexts, gave greater opportunity to distinguish characteristics of increasingly sophisticated performance. This allowed for the more detailed and complex description than earlier.

There is a close relationship of the characteristics associated with levels of statistical literacy and Watson's (1997) three-tiered framework of statistical literacy. The mathematical and statistical skills noted at the different levels reflect the terminology of statistical ideas and its usage, which are suggested as essential in Tier 1 of the framework. The engagement with the context of statistical inquiry reflects Tiers 2 and 3 of the statistical hierarchy. Applying terminology in interpreting a context, which is the goal of Tier 2, appears from Level 3 in this profile, and thinking critically to question inappropriate claims and methods, the goal of Tier 3, appears from Level 5 onwards. The use of open-ended tasks that allowed for the identification of bias or errors in subtle settings gave students the opportunity to display these understandings at increasingly higher levels of the construct. In particular these tasks reflect the transition to the needs of adults in society as users of statistical information that were recognized by Wallman (1993) and Gal (2002). The written nature of the questionnaire further satisfies at least one dimension of Gal's requirement to communicate reactions to statistical information.

6.4. THE COMPLEX NATURE OF STATISTICAL LITERACY

The title of this paper reflects our view that statistical literacy is indeed a complex construct. Interpretation of the variable suggests that it encompasses all individual components of the chance and data curriculum (AEC, 1991, 1994), as well as the foundational aspect of variation (Wild & Pfannkuch, 1999). Beyond these characteristics is the realisation of the importance of engagement with context in defining the underlying construct for statistical literacy. The emergence of context was a distinguishing feature in the higher levels of the construct. The interaction of mathematical skills from the curriculum with the increasingly subtle contexts involving statistical bias or misinformation, creates situations that only students at the highest ability level can interpret successfully. In saying this we realise it is important to recognise that this is as much related to the opportunity to learn as it is to innate ability. The reasoning associated with the application of high-level mathematical skills in a subtle social context is unlikely to emerge through happenstance.

By including many tasks embedded in social settings that require interpretation, this study has identified an important factor leading to high achievement in the realms of statistical literacy. Statistical literacy is not just knowing curriculum-based formulas and definitions but integrating these with an understanding of the increasingly sophisticated and often subtle settings within which statistical questions arise. Using a metaphor suggested by Tognolini (1996) it appears that statistical literacy is a complex construct that may be thought of as a thick thread or rope comprising two interwoven and essential strands: mathematical/statistical understanding of the content and engagement with context in exploiting this understanding. In the past, assessment has focussed almost exclusively on curriculum-based mathematical skills. This study suggests that measurement of statistical literacy is incomplete without the opportunity to engage with genuine social contexts, particularly such as those found in the media items.

6.5. IMPLICATIONS FOR FUTURE RESEARCH

Although the unistructural, multistructural, and relational aspects of responses to individual items (Biggs & Collis, 1982) could be identified in many cases to aid in developing hierarchical codings, the combination of mathematical skills and engagement with context provided the opportunity to describe six rather than three levels of the overall statistical literacy construct. In terms of the earlier work of Campbell, Watson, and Collis (1992) on students' understanding of volume measurement, the categorisation of levels depends to some extent on the strength of the microscope used to view the phenomenon. Viewing the statistical literacy construct from "afar", it is possible to speculate on the existence of two unistructural-multistructural-relational cycles, similar to those identified by Campbell et al. and for beginning inference by Watson and Moritz (1999b). At Levels 1, 2, and 3, success on items reflects the increasingly structured use of data and information in a highly organised task environment. At Levels 4, 5, and 6, open-ended tasks and less familiar settings provide contexts where success is associated with using more complex mathematical skills and engagement within increasingly complex settings. These levels progressively appear to reflect simple single classroom settings (like using dice), multiple aspects of settings (such as surveys within the school environment), and complex relational settings (such as finding bias in unfamiliar social settings presented in the media). More research, however, is required to provide convincing evidence of this hypothesis.

The appearance of different task-steps for the same task at the same level of difficulty in some instances also provides some insight into students' achievement. It suggests that higher levels of sophistication in thinking are not always related to higher ability. Rather, students appear to be drawing on different ways of conceptualizing the question, and thus, in some situations, students at the same ability level have two ways of responding to a particular question. Improved identification of different conceptual frameworks could provide useful information to teachers about appropriate interventions for students at the same level of understanding.

Another step in future research is to analyse longitudinal data on individuals to explore the hypothesis that indeed the hierarchical structure observed in this study represents a developmental sequence that could be expected to be observed over the years of schooling. Several studies based on subsets of items included in this study suggest that such a hypothesis is reasonable (Watson & Moritz, 1998, 1999a, 2000b).

Although the characterization of the underlying statistical literacy construct appears sound based on the items used and the data collected in this study, using all of the 80 items would be impractical for an instrument to establish statistical literacy standards or benchmarks in the classroom. Items providing redundant data could be eliminated from any new instrument designed to assess statistical literacy. The choice of which items to leave out is dependent on the test writers but any new questionnaire or test purporting to measure statistical literacy should have a test specification that includes items that address both mathematical skills and contextually based application of these. This is consistent both with the conceptualisation of statistical literacy and the findings of this study.

Some gaps in the content covered with respect to topics in the curriculum also emerged when the overall scale was considered. For example, there were few difficult items relating to tables or to more complex graph types, such as those with non-linear association. New items will be needed to attend to this. The next stage of the research will address the preparation of an improved test for the construction of such a scale and its trialling in schools. Associated with this will be the identification of expectations for particular grade levels within the overall hierarchy of the statistical literacy construct. The current research has provided a foundation for future work, and confidence that statistical literacy is a single hierarchical construct that can be measured as students progress through school.

6.6. IMPLICATIONS FOR CLASSROOM PLANNING

Even before a shorter instrument is developed, recommendations for curriculum planners and teachers can be made based on the observed statistical literacy construct. We feel there needs to be more use of context, particularly socially-based media examples, in teaching statistics, both within mathematics and in other curriculum areas. We would support a concerted effort to devise activities specifically to assist students to move from non-context based application of statistical skills, such as "add them up and divide" interpretations of average, to an appreciation of context, and then to an awareness of its importance in decision making, including developing the skills to identify bias and misrepresentation. Some of these activities could be based on media items and interview protocols used in research (e.g., Watson & Moritz, 1999b, 2000a). Explicit discussion of the interwoven nature of the two strands of statistical literacy may help students appreciate its importance. We feel curriculum planners need to develop materials that enhance mathematical and statistical skills at the same time as the qualitative understanding of statistical reasoning.

It should be noted that we are not suggesting that teachers should neglect developing the separate underlying concepts, such as average, chance, variation, or sampling. Nor are we suggesting that improving a student's understanding of one of these concepts will also improve understanding of another, different idea. Rather, this research indicates not only that the underlying ideas are important, but also that students need to have opportunities to address these ideas in a range of contexts, including non-school-based ones. This requires a balance of concept and skill development and application of the ideas in authentic situations, and makes increased demands on teachers and curriculum planners.

In the light of changed curriculum expectations (e.g., Education Queensland, 2000) and extended social expectations for quantitative literacy generally (e.g., Steen, 2001), we believe that teachers across the curriculum will also have increased expectations placed on them in terms of appreciating statistical literacy and how to develop it. It is likely that professional development for teachers will be needed if they are to assist their students to achieve the highest levels of statistical literacy observed here before they leave formal schooling.

ACKNOWLEDGMENTS

This research was funded by the Institutional Research Grant Scheme at the University of Tasmania in 2001. At the time of the initial research Rosemary Callingham was employed by the University of Tasmania. The authors thank Dr. John Izard for helpful comments during the initial preparation of this paper and the anonymous referees for suggestions for necessary revisions.

REFERENCES

Adams, R. J., & Khoo, S. T. (1996). *Quest: Interactive item analysis system. Version 2.1* [Computer software]. Melbourne: Australian Council for Educational Research.

Anastasi, A. (1988). *Psychological testing*. Macmillan: New York.

Australian Association of Mathematics Teachers (AAMT) (1997). *Numeracy = everyone's business. Report of the Numeracy Education Strategy Development Conference. May 1997*. Adelaide: Author.

Australian Education Council (1991). *A national statement on mathematics for Australian schools.* Carlton, Vic.: Author.

Australian Education Council (1994). *Mathematics - A curriculum profile for Australian schools.* Carlton, Vic.: Curriculum Corporation.

Batanero, 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*, 151-169.

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

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, NJ: Lawrence Erlbaum.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Cai, J. (1995). Beyond the computational algorithm: Students' understanding of the arithmetic average concept. In L. Meira & D. Carraher (Eds.), *Proceedings of the 19th Psychology of Mathematics Education Conference* (Vol. 3*,* pp. 144-151). São Paulo, Brazil: PME Program Committee.

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics, 98*, 93-98.

Campbell, K. J., Watson, J. M., & Collis, K. F. (1992). Volume measurement and intellectual development. *Journal of Structural Learning*, *11*, 279-298.

Castles, I. (1992). *Surviving statistics: A user's guide to the basics.* Canberra: Australian Bureau of Statistics.

Cockcroft, W. H. (1982). *Mathematics counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools*. London: HMSO.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem solving ability. *Journal for Research in Mathematics Education, 17*, 206-221.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Cureton, E. E. (1965). Reliability and validity: Basic assumptions and experimental designs. *Educational and Psychological Measurement, 25*, 326-346.

Department of Education Tasmania. (2002). *Essential learnings framework 1*. Hobart: Author.

Dossey, J. A. (1997). National indicators of quantitative literacy. In L. A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 45-59). New York : The College Board.

Education Queensland (2000). *New Basics Project technical paper*. Retrieved January 11, 2002, from http://education.qld.gov.au/corporate/newbasics/html/library.html

Eisner, E. W. (1993). Reshaping assessment in education: Some criteria in search of practice. *Journal of Curriculum Studies, 25*(3), 219-233.

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

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? An exploratory research study. *Educational Studies in Mathematics, 15*, 1-24.

Fisher, W. P. (1994). The Rasch debate: Validity and revolution in educational measurement. In M. Wilson (Ed.), *Objective measurement: Vol. 2* (pp. 36-72). Norwood, NJ: Ablex.

Frankenstein, M. (2001). Reading the world with math: Goals for a critical mathematical literacy curriculum. In *Mathematics shaping Australia* (Proceedings of the 18th Biennial Conference of the Australian Association of Mathematics Teachers, Inc.). [CDROM] Canberra: AAMT.

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

Gagatsis, A., Kyriakides, L., & Panaoura, A. (2001). Construct validity of a developmental assessment on probabilities: A Rasch measurement model analysis. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 449-456). Utrecht, The Netherlands: Freudenthal Institute.

Gal, I. (2000). Statistical literacy: Conceptual and instructional issues. In D. Coben, J. O'Donoghue, & G. E. Fitzsimons (Eds.), *Perspectives on adults learning mathematics: Research and practice* (pp. 135-150). Dordrecht: Kluwer.

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

Gal, I., & Wagner, D. A. (1992). *Project STARC: Statistical reasoning in the classroom*. (Annual Report: Year 2, NSF Grant No. MDR90-50006). Philadelphia, PA: Literacy Research Center, University of Pennsylvania.

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 23-38.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*, 923-936.

Green, D. R. (1982). *Probability concepts in 11-16 year old pupils.* Loughborough, UK: Center for Advancement of Mathematical Education in Technology, University of Technology.

Green, D. (1983a). Shaking a six. *Mathematics in Schools, 12*(5), 29-32.

Green, D. R. (1983b). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the 1st International Conference on Teaching Statistics* (Vol. 2, pp. 766-783). Sheffield, England: Teaching Statistics Trust.

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 and how?* (pp. 219-239). Voorburg, The Netherlands: International Statistical Institute.

Holmes, P. (1980). *Teaching statistics 11-16.* Slough, UK: Schools Council and Foulsham Educational.

Izard, J. (1992). Patterns of development with probability concepts: Assessment for informative purposes. In M. Stephens & J. Izard (Eds.), *Reshaping assessment practices: Assessment in the mathematical sciences under challenge. Proceedings from the First National Conference on Assessment in the Mathematical Sciences* (pp. 355-367). Melbourne, VIC: Australian Council for Educational Research.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School, 5*(4), 240-263.

Jones, B. (1982). *Sleepers wake: Technology and the future of work.* Melbourne: Oxford University Press.

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

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

Keeves, J. P., & Alagumalai, S. (1999). New approaches to measurement. In G. N. Masters and J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 23-42). Oxford: Pergamon.

Kirsch, I. W. (1997). Literacy performance on three scales: Definitions and results. In W. McLennan, *Aspects of literacy: Assessed skill levels Australia 1996* (pp. 98-124). Canberra: Australian Bureau of Statistics.

Kolen, M. J. (1999). Equating of tests. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 164-175). New York: Pergamon.

Konold, C., & Garfield, J. (1992). *Statistical reasoning assessment: Part 1. Intuitive Thinking.* Scientific Reasoning Research Institute, University of Massachusetts, Amherst, MA.

Konold, C., & Higgins, T. L. (2002). Working with data: Highlights related to research. In S. J. Russell, D. Schifter, & V. Bastable (Eds.), *Developing mathematical ideas: Collecting, representing, and analyzing data* (pp. 165-201). Parsippany, NJ: Dale Seymour Publications.

Linacre, M. (1997, August). *Judging plans and facets*. MESA Research Note 3. Retrieved January 8, 2003, from http://www.rasch.org/m3.htm

Lokan, J. Ford, P., & Greenwood, L. (1997). *Maths and science on the line: Australian middle primary students' performance in the Third International Mathematics and Science Survey (TIMSS)*. Melbourne: Australian Council for Educational Research.

Luke, A., & Freebody, P. (1997). Shaping the social practices of reading. In S. Musprati, A. Luke, & P. Freebody (Eds.), *Constructing critical literacies: Teaching and learning textual practice* (pp. 185-225). St. Leonards, NSW: Allen & Unwin.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

McLennan, W. (1997). *Aspects of literacy: Assessed Skill Levels Australia 1996*. Canberra: Commonwealth of Australia.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: American Council on Education and Macmillan Publishing Company.

Metz, K. E. (1998). Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction, 16*, 285-365.

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 Publications.

Ministry of Education (1992). *Mathematics in the New Zealand curriculum*. Wellington: Author.

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

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

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 23rd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 440-447). Perth, WA: MERGA.

Moritz, J. B., & Watson, J. M. (1997). Graphs: Communication lines to students? In F. Biddulph & K. Carr (Eds.), *People in mathematics education: Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 344-351). Rotorua, NZ: MERGA.

Moritz, J. B., & Watson, J. M. (2000). Reasoning and expressing probability in students' judgements of coin tossing. In J. Bana & A. Chapman (Eds.), *Mathematics education beyond 2000: Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 448-455). Perth, WA: MERGA.

Moritz, J. B., Watson, J. M., & Collis, K. F. (1996). Odds: Chance measurement in three contexts. In P. C. Clarkson (Ed.), *Technology in mathematics education: Proceedings of the 19th annual conference of the Mathematics Education Research Group of Australasia* (pp. 390-397). Melbourne: MERGA.

Moritz, J. B., Watson, J.M., & Pereira-Mendoza, L. (1996, November). *The language of statistical understanding: An investigation in two countries*. Paper presented at the Joint ERA/AARE Conference, Singapore. [Online: http://www.swin.edu.au/aare/96pap/morij96.280]

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.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Pollatsek, A., Well, A. D., Konold, C., Hardiman, P., & Cobb, G. (1987). Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes, 40*, 255-269.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published 1960).

Reading, C. (2002). Profile for statistical understanding. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute.

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

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

Romberg, T. A., Jurdak, M. E., Collis, K. F., & Buchanan, A. E. (1982). *Construct validity of a set of mathematical superitems*. Madison, Wisconsin: Wisconsin Center for Education Research.

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 20th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6-22). Rotorua, NZ: MERGA.

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

Statistics Canada and Organisation for Economic Cooperation and Development (OECD) (1996). *Literacy, economy, and society: First results from the International Adult Literacy Survey*. Ottawa: Author.

Steen, L. A. (Ed.) (1997). *Why numbers count: Quantitative literacy for tomorrow's America*. New York: College Entrance Examination Board.

Steen, L. A. (Ed.) (2001). *Mathematics and democracy: The case for quantitative literacy.* Washington, DC: National Council on Education and the Disciplines.

Stocking, M. L. (1999). Item response theory. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 55-63). New York: Pergamon.

Tognolini, J. (1996). Rasch modelling: Advantages and limitations. In *Session notes National Meeting on Assessment and Reporting 25-26 November 1996*. Manly: NSW Department of School Education.

Torok, R. (2000). Putting the variation into chance and data. *Australian Mathematics Teacher*, *56(2),* 25-31.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90,* 293-315.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*, No. 421, 1-8.

Watson, J. M. (1994). Instruments to assess statistical concepts in the school curriculum. In National Organizing Committee (Ed.), *Proceedings of the Fourth International Conference on Teaching Statistics. Volume 1* (Vol. 1, pp. 73-80). Rabat, Morocco: National Institute of Statistics and Applied Economics

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

Watson, J. M. (1998a). Numeracy benchmarks for years 3 and 5: What about chance and data? In C. Kanes, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times.* (Proceedings of the 21st annual conference of the Mathematics Education Research Group of Australasia, Vol. 2, pp. 669-676). Brisbane: MERGA.

Watson, J. M. (1998b). The role of statistical literacy in decisions about risk: Where to start. *For the Learning of Mathematics*, *18*(3), 25-27.

Watson, J. M. (2000). Statistics in context. *Mathematics Teacher*, *93*, 54-58.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1994). Assessing statistical understanding in Grades 3, 6 and 9 using a short answer questionnaire. In G. Bell, B. Wright, N. Leeson, & G. Geake (Eds.), *Challenges in Mathematics Education: Constraints on Construction Proceedings of the 17th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 675-682). Lismore, NSW: MERGA.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, *9*, 60-82.

Watson, J. M., & Kelly, B. A. (2002). 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*. Voorburg, The Netherlands: International Statistical Institute.

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

Watson, J. M., & Moritz, J. B. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal*, *10*(2), 103-127.

Watson, J. M., & Moritz, J. B. (1999a). The development of concepts of average. *Focus on Learning Problems in Mathematics*, *21*(4), 15-39.

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

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

Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, *19*, 109-136.

Watson, J. M., & Moritz, J. B. (2002). School students' reasoning about conjunction and conditional events. *International Journal of Mathematical Education in Science and Technology, 33,* 59-84.

Watson, J. M., & Moritz, J. B. (2003). The development of comprehension of chance language: Evaluation and interpretation. *School Science and Mathematics, 103*, 65-80.

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

Wilson, M. (1990). Investigation of structured problem-solving items. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 187-203). Washington, DC: American Association for the Advancement of Science.

Wilson, M. (1992). Measuring levels of mathematical understanding. In T. A. Romberg (Ed.), *Mathematics assessment and evaluation: Imperatives for mathematics educators* (pp. 213-241). Albany, NY: State University of New York Press.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress* (235-268). Reston, VA: NCTM.

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

APPENDIX A: VARIABLE MAP FOR STATISTICAL LITERACY CONSTRUCT

```
 4.0                               |
                                   |    HAT8.4
                                   |
                                   |    MVE3.3
                                   |
                                   |    M5AV.3
                                   |    Q17    M10A.4                              Level 6  Critical mathematical
                                   |    M3OD.4
                                   |    BT1A.2
 3.0                               |    RAN3.3 TRV5.2 TRV6.5
                                   |
                                   |    M7CH.4
                                   |
                                   |    TBL5.4
                                   |    M9D .2 TRV4.3
                                   |    M9C .2 TWN2.3
                                   |    M8QU.2 SP9
                                   |    DIE7.4 TWN1.3 AVG1.3
                                 X |
-----------------------------------------------------------------------------------------------------------
 2.0                               |
                                 X |    SP2A.3 MVE5.3 BT1B.3
                                 X |    CP16   SM19.3 M2PI.2 M10A.3 MVE2.3 MVE6.3 VAR .3      Level 5  Critical
                                XX |    SMP4.3 ME13.3 Q20    MVE1.3 SMP3.3 BT1B.2 DRG1
                                XX |    M7CH.3 Q10E.4 M6D .2
                               XXX |    M3OD.3 SP2B.3
                              XXXX |    M4DR.2 MVE8.3
                             XXXXX |    M8GR.3 MVE6.2
-----------------------------------------------------------------------------------------------------------
                              XXXX |    BOX9.3 M1CH.2 M7CH.2 SP5B.2 TWN3.3
 1.0                        XXXXXX |    M2PI.1 M6C    DIE2.4 TRV6.4 MVE4.3 SP10     Level 4  Consistent non-critical
                            XXXXXX |    M5AV.2 M6AB.2 M6D .1 SP3A.3
                        XXXXXXXXXX |    SP3B.3 MVE1.2 SP8
                           XXXXXXX |    M8GR.2 M8QU.1 AVG1.2 MVE8.2
                      XXXXXXXXXXXX |    AVG2.3 DIE2.3 TRV4.2 TRV6.3 MVE2.2 MVE7.2 SMP3.2 SP7
                    XXXXXXXXXXXXXX |    DIE7.3 M10A.2 M10B.2 SP11.2 VAR .2
                   XXXXXXXXXXXXXXX |    ME13.2 M3OD.2 BT1B.1
-----------------------------------------------------------------------------------------------------------
                   XXXXXXXXXXXXXXX |    RAN3.2 SP5A.2 TBL5.3 MVE5.2 BT1A.1 AVG1.1
                  XXXXXXXXXXXXXXXX |    CF15   SP3A.2 MVE3.2 MVE4.2 TBL5.2 TWN2.2 TWN3.2
  .0              XXXXXXXXXXXXXXXX |    M4DR.1 M10A.1 M6AB.1 SP2A.2 MVE2.1 TWN1.2 VAR .1
                      XXXXXXXXXXXX |    SMP4.2 CH11.2 AV12.2 CF18   RAN3.1
             XXXXXXXXXXXXXXXXXXXXX |    M5AV.1 MVE3.1 MVE6.1                          Level 3  Inconsistent
                    XXXXXXXXXXXXXX |    DIE7.2 SP4A   MVE5.1 TWN2.1
                     XXXXXXXXXXXXX |    HAT8.3 AV12.1 SP3B.2 MVE4.1
                  XXXXXXXXXXXXXXXX |    M8GR.1
-----------------------------------------------------------------------------------------------------------
                      XXXXXXXXXXXX |    CP14   M9D .1 TRV4.1 SMP3.1 TWN1.1 TWN3.1
                     XXXXXXXXXXXXX |    M1CH.1 M9C .1 M10B.1 SP1 .2 SP3A.1 SP5A.1 MVE7.1
                      XXXXXXXXXXXX |    AVG2.2 ME13.1 DIE2.2 TBL5.1
-1.0                      XXXXXXXX |    BOX9.2 TRV6.2 MVE1.1 TBL3.2
                        XXXXXXXXXX |    SP5B.1                                         Level 2  Informal
                        XXXXXXXXXX |    SMP4.1 SM19.2
                               XXX |    AVG2.1 HAT8.2 SP3B.1 TBL4   MVE8.1
                          XXXXXXXX |    BOX9.1 Q10E.3 SP2B.2 SP4B   TBL3.1 SP6
                             XXXXX |    Q10D.2
                                XX |
-----------------------------------------------------------------------------------------------------------
                              XXXX |    DIE7.1 CH11.1 TRV2   SP11.1
                                XX |    HAT8.1 TRV6.1 TBL1
                               XXX |    TBL2                                           Level 1  Idiosyncratic
-2.0                             X |    M3OD.1
                               XXX |    M7CH.1 Q10E.2 TRV3
                                 X |    SM19.1
                                XX |    SP1 .1 SP2A.1 TRV5.1
                                 X |
                                   |
                                 X |    DIE2.1
                                   |
                                   |
-3.0                             X |
                                   |    Q10C.2 Q10D.1 SP2B.1
                                 X |    Q10E.1
                                   |
                                   |    TRV1
                                   |    Q10A.2
                                   |
                                   |    Q10B.2
                                   |
-4.0                               |
                                   |
                                   |    Q10B.1
                                   |    Q10C.1
                                   |
                                   |
                                   |    Q10A.1
                                   |
                                   |
                                   |
-5.0                               |
Each X represents   12 students
```

APPENDIX B: NUMBER OF STUDENTS IN EACH GRADE ANSWERING EACH ITEM

	GRADE					
ITEM	**3**	**5**	**6**	**7**	**8**	**9**
AV12		218	518		167	641
AVG1				189		197
AVG2	626	218	518		167	641
BOX9	1039	421	875	239	196	1034
BT1A, BT1B				189		197
CF15, CF18, CP14, CP16		238	861	46	196	837
CH11		238	875	50	196	837
DIE2	176	183		189		197
DIE7	1039	421	875	239	196	1034
DRG1				189		197
HAT8	863	238	875	50	196	837
M10A, M10B			395		185	403
M1CH, M2PI, M3OD			695		185	746
M4DR				189	185	943
M5AV			517		165	618
M6AB, M6C, M6D					184	746
M7CH			695		185	746
M8GR, M8QU			396		185	746
M9C, M9D			521		176	647
ME13		218	518		167	641
MVE1 to MVE4, MVE7	176	183		189		197
MVE5		183		189		197
MVE6				189		197
MVE8						197
Q10A to Q10E	854	238	875	50	196	837
Q17, Q20		238	861	46	196	837
RAN3	863	238	875	239	196	1034
SM19		238	861	46	196	837
SMP3	176	183		189		197
SMP4	863	238	875	50	196	837
SP1, SP2A	176	183		189		197
SP2B				189		197
SP3A	176	183				
SP3B				189		197
SP4A	176	183				
SP4B				189		197
SP5A	176	183				
SP5B				189		197
SP6 to SP11				189		197
TBL1 to TBL5	176	183		189		197
TRV1 to TRV6	176	183		189		197
TWN1 to TWN3		183		189		197
VAR				189		197

APPENDIX C: ITEM STATEMENT AND RESPONSE CODE EXAMPLES

AV12. To get the average number of children per family in a town, a teacher counted the total number of children in the town. She then divided by 50, the total number of families. The average number of children per family was 2.2.

Tick which of these is certain to be true.

(a) Half of the families in the town have more than 2 children.
(b) More families in the town have 3 children than have 2 children.
(c) There are a total of 110 children in the town.
(d) There are 2.2 children in the town for every adult.
(e) The most common number of children in a family is 2.
(f) None of the above.

Code 3	c
Code 2	d, e, f, multiple
Code 1	a, b
Code 0	NR

AVG1. A small object was weighed on the same scales separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.3 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.3

The "average" value could be calculated in several ways.

1. How would you find the average? ______________________
2. The average weight is _______ grams. [Show your working in the box below.]

Code 3	Mode explained; Median explained and correct; Mean discarding outlier
Code 2	7.18, mean
Code 1	Any of the three measures mentioned but answer incorrect
Code 0	Incorrect or idiosyncratic method with or without unreasonable answer; NR

AVG2. If someone said you were "average", what would it mean?

Code 3	Add and divide, same as most, in the middle between good and bad
Code 2	Add, same as others, okay, normal
Code 1	Example
Code 0	Don't know, etc; No response (NR)

BOX9. Box A and Box B are filled with red and blue marbles as follows. Each box is shaken. You want to get a blue marble, but you are only allowed to pick out one marble without looking. Which box should you choose?

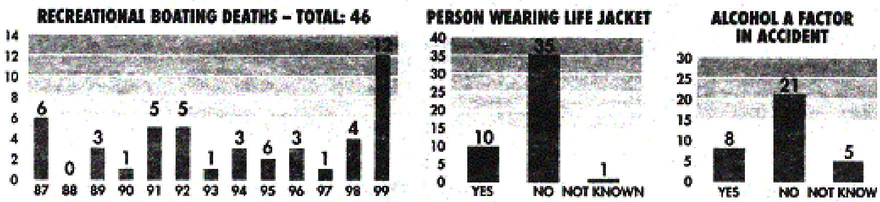
(A) Box A (with 6 red and 4 blue).
(B) Box B (with 60 red and 40 blue).
(=) It doesn't matter.

Please explain your answer.

Code 3	=, 40% chance each, B is 10 times bigger than A, same chance; 40 versus 4, 6 versus 6, similar
Code 2	=, both have more red; A, only 2 more reds, B 20 more reds; B, more blues, more marbles; A less reds, less marbles
Code 1	=, could be anything; A, B, = idiosyncratic reason
Code 0	NR, no reason

BT1. These graphs were part of a newspaper story reporting on boating deaths in Tasmania. Comment on any unusual features of the graphs. (2 spaces provided)

BOATIE'S SAFETY FAILURE



BT1A.

Code 2	Recognises mistakes: "The axes aren't named" & "The total boating deaths of 46 is not the same shown in the graph"
Code 1	Statistically specific comments about graphing elements; perhaps includes some incorrect comments
Code 0	Incorrect graph interpretation of unusual data; Inferring from graph; Advice; idiosyncratic

BT1B.

Code 3	Acknowledges variation: "In some years there were heaps and they dropped to none"
Code 2	Focuses on increase: "As the years progress the amount of years grew"
Code 1	Focuses on the highest column, a single column or only 2 columns
Code 0	No focus: "Numbers are above the shaded area"

CF15. Please estimate:
(a) The probability that you will miss a whole week of school next year.
(b) The probability that you will get a cold next year.
(c) The probability that you will get a cold causing you to miss a whole week of school next year.

Code 1	c<min(a,b); c=min(a,b)
Code 0	min<c≤max, c>max(a,b), undefined

CF18. A health survey was conducted in a sample of 100 men in Australia of all ages and occupations. Please estimate:
(a) How many of the 100 men have had one or more heart attacks.
(b) How many of the 100 men are over 55 years old.
(c) How many of the 100 men both are over 55 years old and have had one or more heart attacks.

Code 1	c<min(a,b); c=min(a,b)
Code 0	min<c≤max, c>max(a,b), undefined

CH11. A bottle of medicine has printed on it*: WARNING: For applications to skin areas there is a 15% chance of getting a rash. If you get a rash, consult your doctor.* What does this mean?
(a) Don't use the medicine on your skin - there's a good chance of getting a rash.
(b) For application to the skin, apply only 15% of the recommended dose.
(c) If you get a rash, it will probably involve only 15% of the skin.
(d) About 15 out of every 100 people who use this medicine get a rash.
(e) There is hardly any chance of getting a rash using this medicine.

Code 2	d, a and d, d and e
Code 1	e, a
Code 0	b, c, NR, multiple selections

CP14. Please estimate:
(a) Out of 100 men, how many are left-handed.
(b) Out of 100 left-handed adults, how many are men.

Code 1	b>a
Code 0	b=a/2, b=a, b<a, undefined

CP16. Please estimate:
(a) The probability that a woman is a school teacher.
(b) The probability that a school teacher is a woman.

Code 1	b>a
Code 0	b=a/2, b=a, b<a, undefined

DIE2. Imagine you threw the dice 60 times. In the table below, fill in how many times you think each number might come up. Why do you think these numbers are reasonable? [Appeared immediately after DIE7]

Number on Dice	How many times it might come up
1	
2	
3	
4	
5	
6	
TOTAL	60

Code 4	Appropriate variability in prediction and reason
Code 3	Strict probability with reason reflecting variation; Too much or too little variation with reason reflecting chance
Code 2	Strict probability with reasons reflecting classical chance or aspects of geometry; multiples of 5 with reasons reflecting chance; Did not add to 60 with reasons reflecting equality or chance
Code 1	That's what I think.
Code 0	Sums ≠ 60 and odd distributions with no reasoning

DIE7. Consider rolling one six-sided die. Is it easier to throw
(1) a one or
(6) a six or
(=) are both a one and a six equally easy to throw?
Please explain your answer.

Code 4	=, 1/6 chance every number
Code 3	=, only one of each number, same chance, cube
Code 2	=, could be anything, never know outcome
Code 1	1, 6, = idiosyncratic reason
Code 0	NR

DRG1. What was the size of the sample in this article?

Decriminalise drug use: poll
SOME 96 percent of callers to youth radio station Triple J have said marijuana use should be decriminalised in Australia. The phone-in listener poll, which closed yesterday, showed 9924 - out of the 10,000-plus callers - favoured decriminalisation, the station said. Only 389 believed possession of the drug should remain a criminal offence. Many callers stressed they did not smoke marijuana but still believed in decriminalising its use, a Triple J statement said.

Code 1	10,313; 10,000+; 10,000
Code 0	9924 out of 10 000+ callers; 96%; very small

HAT8. A mathematics class has 13 boys and 16 girls in it. Each pupil's name is written on a piece of paper. All the names are put in a hat. The teacher picks out one name without looking. Is it more likely that
(b) the name is a boy or
(g) the name is a girl or
(=) are both a girl and a boy equally likely?
Please explain your answer.

Code 4	g, 16/29 chance
Code 3	g, 13 versus 16, more girls
Code 2	=, depends on mix, same chance, could be anything
Code 1	b, g, = idiosyncratic, such as luck or teacher is a certain sex
Code 0	NR, no reason

M10A. During the recent Australian cricket tour of South Africa, the Hobart Mercury (6/4/1994, p. 52) reported that Allan Border had lost 8 out of 9 tosses in his previous 9 matches as captain. Imagine his situation at this point in time.
Suppose Border decides to choose heads from now on. For the next 4 tosses of the coin, what is the chance of the coin coming up tails (and him losing the tosses) 4 times out of 4?

Code 4	1/16
Code 3	Other number/word
Code 2	50%, 50-50, 2/4 other, word
Code 1	Value > 0.5
Code 0	NR

M10B. Suppose tails came up 4 times out of 4. For the 5th toss, should Border choose
__Heads
__Tails
__Doesn't matter
What is the probability of getting heads on this next toss?
What is the probability of getting tails on this next toss?

Code 2	=, each value 0.5
Code 1	= other values; H/T, each value 0.5
Code 0	H/T other values

M1CH. Here are eight chance words or phrases from headlines.
A. 58 per cent success at SkillShare
B. Impossible
C. It's a sure thing
D. Jack looking good for big one

E. Holden an unlikely American hero
F. No worries
G. Smith in doubt to play
H. There's a 50-50 chance

Please mark on the scale below the likelihood expressed by each of the seven phrases A to G. H is done as an example.

Likelihood

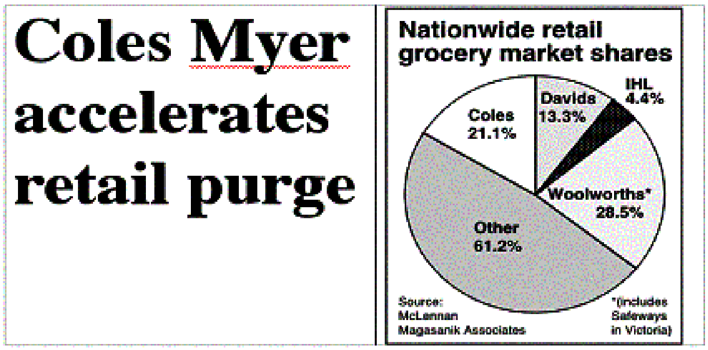
(low) (high)

|________________________|________________________|

0 1

Code 2	Letter in position and correct halves
Code 1	Up to 2 half errors
Code 0	>2 errors on halves

M2PI. Explain the meaning of this pie chart. Is there anything unusual about it?



Code 2	>100%; 61.2% less than half
Code 1	Large other; not explained % does match headline; company names
Code 0	Colours, no/yes, NR

M3OD. What does "7-2" mean in this headline about the North against South football match? Give as much detail as you can. From the numbers, who would be expected to win the game?

North at 7-2 **But we can still win match, says coach**

Code 4	South, 2/9
Code 3	North, 7/9; South, 2/7, if play, win 2/7; Chance 3.5 to 1, won 2/7
Code 2	% chance, 2 pts to 7 pts, odds; Chance, predicted score
Code 1	Current score, betting
Code 0	NR, no reason

M4DR. Is the sample reported here a reliable way of finding out public support for the decriminalisation of marijuana? Why or why not? [see DRG1]

Code 2	Only JJJ listeners, Only motivated ones phone in
Code 1	No, small sample; Yes, large sample; No, not everyone; Yes, everyone listens to radio; Reliability of measurement; No, could be lying
Code 0	Shouldn't have marijuana, NR

M5AV. What does “average” mean in this article? What does “median” mean in this article?
Why would the median have been used?

Hobart defies homes trend

AGAINST a national trend, Hobart’s median house price rose to $88,200 in the March quarter - but, Australia-wide, the average wage-earner finally can afford to buy the average home after almost two years of mortgage pain.

Code 3	Median not influenced by outliers as mean is; Good contrast median and mean
Code 2	Add and divide, middle value
Code 1	Normal
Code 0	Tautology/irrelevant, NR

M6. Each of the <u>four sentences</u> in the following article sets a condition and describes an associated outcome. In each case, state what these are.

Wrinkles ultimate smoking deterrent.

1. A study found that those who smoked a pack of cigarettes a day for less than 49 years doubled the risk of premature wrinkling.
2. For more than 50 years, the risk was 4.7 times greater than those who do not smoke.
3. He said he was not sure if the wrinkling could be reversed if people quit smoking.
4. “ ‘You’re going to be old and ugly before your time if you smoke,’ may be just the message that leads them to throw away their cigarettes for good," he said.

Condition Outcome
Q1. ______________________________ ______________________________
Q2. ______________________________ ______________________________
Q3. ______________________________ ______________________________
Q4. ______________________________ ______________________________

M6AB.

Code 2	Both Q1 and Q2 correct: {Smoke, Cigarette, Pack} →{Wrinkle, Premature}{50} → {Wrinkle, Risk, 4.7}
Code 1	One of Q1 and Q2 correct
Code 0	Incorrect, NR

M6C.

Code 1	Q3 correct: {Smoke, Quit} → {Wrinkle, Reverse}
Code 0	Incorrect, NR

M6D.

Code 2	Q4 correct: {Message, Wrinkle, Old, Ugly} → {Quit, Not Smoke}
Code 1	Q4 correct: {Smoke} → {Wrinkle, Old, Ugly}
Code 0	Incorrect, NR

M7CH. Would you make any criticisms of the claims in this article? If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona? Why or why not?

ABOUT six in 10 United States high school students say they could get a handgun if they wanted one, a third of them within an hour, a survey shows. The poll of 2508 junior and senior high school students in Chicago also found 15 per cent had actually carried a handgun within the past 30 days, with 4 per cent taking one to school.

Code 4	(a) Only Chicago has been asked
Code 3	No, 2508 is small sample; (b) Maybe not in Arizona
Code 2	No, not everyone; Reliability of measurement: No, could be lying; Whole of USA would be the same
Code 1	Shouldn't have guns
Code 0	NR

M8

Family car is killing us, says Tasmanian researcher

Twenty years of research has convinced Mr Robinson that motoring is a health hazard. Mr Robinson has graphs which show quite dramatically an almost perfect relationship between the increase in heart deaths and the increase in use of motor vehicles. Similar relationships are shown to exist between lung cancer, leukaemia, stroke and diabetes.

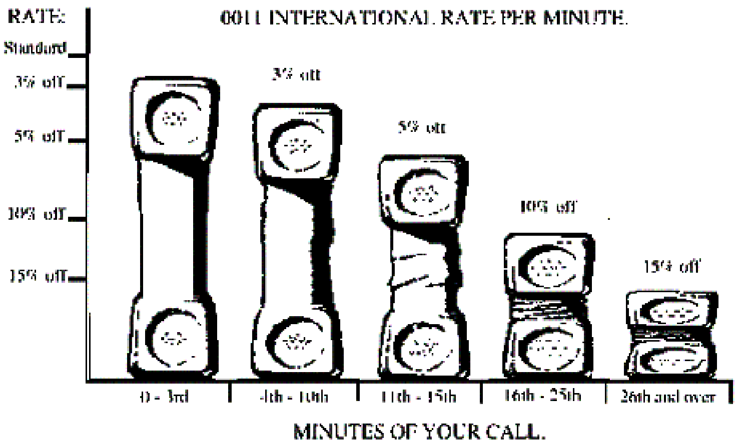
M8GR. Draw and label a sketch of what one of Mr. Robinson's graphs might look like.

Code 3	Bivariate or Series Comparison Graph
Code 2	Trend or Double comparison Graph
Code 1	Labeled or Single Comparison; Basic Graph
Code 0	No graph

M8QU. What questions would you ask about his research?

Code 2	Other causes? How linked?
Code 1	Sample size, & location; Location, size, age groups
Code 0	Can it be prevented?

M9. The longer your overseas call, the cheaper the rate.



M9a Explain the meaning of this graph. [Not coded]

M9b Is there anything unusual about it? [Not coded]

M9C. Suppose the standard rate is $1.00 for 1 minute.You have already talked for 30 minutes How much would the next 10 minutes cost?

Code 2	$8.50
Code 1	15%, $1.50, 85c; 3%, $10, $40, Other$
Code 0	NR

M9D. How much did the first 30 minutes of the phone call cost?

Code 2	\$27.79 or calculation by right method; \$25.50
Code 1	15%, \$4.50; \$30, Other\$, Other %
Code 0	NR

ME13. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.
6.3 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.3
The median of this set of data is
(a) the most common value.
(b) the middle value.
(c) the most accurate value.
(d) the average value.
So, the median value is _________________ grams.

Code 3	b & 6.2; d & 7.18
Code 2	a & 6.3; b & 6.1; d and 6≤#≤8
Code 1	c, other selections not above
Code 0	NR

MVE. MOVIEWORLD
A class wanted to raise money for their school trip to Movieworld on the Gold Coast. They could raise money by selling raffle tickets for a Nintendo Game system. But before they decided to have a raffle they wanted to estimate how many students in their whole school would buy a ticket. So they decided to do a survey to find out first. The school has 600 students in grades 1-6 with 100 students in each grade.

MVE1. How many students would you survey and how would you choose them? Why?

Code 3	Representative & random; Random only
Code 2	Based on one or more factors
Code 1	Just the students I meet; take them all
Code 0	Misinterpretation

MVE2. Three students in the school conducted surveys. Shannon got the names of all 600 children in the school and put them in a hat, and then pulled out 60 of them. What do you think of Shannon's survey?
__Good __ Bad __ Not Sure - Why?

Code 3	Random methods; range
Code 2	Fair chance; sample size; methodology (easy)
Code 1	Method too random, inaccurate; inadequate sample size; unfair; time consumption
Code 0	Misinterpretation; no reason or logic

MVE3. Jake asked 10 children at an after-school meeting of the computer games club. What do you think of Jake's survey?
__Good __ Bad __ Not Sure - Why?

Code 3	Detecting bias & small sample size
Code 2	Bias only, small sample size only; unfair, survey all
Code 1	Creating bias, good sample size; good method
Code 0	Misinterpretation; no reason or logic

MVE4. Adam asked all of the 100 children in Grade 1. What do you think of Adam's survey?
__Good __ Bad __ Not Sure - Why?

Code 3	Detecting bias in groups
Code 2	Sample size too large; unfair; not sure
Code 1	Large sample size good; fair
Code 0	Misinterpretation; no reason or logic

MVE5. Raffi surveyed 60 of his friends. What do you think of Raffi's survey?
__Good __ Bad __ Not Sure - Why?

Code 3	Lack of range &/or variation
Code 2	Unfair; vague friendship factor; uncertainty; adequate sample size
Code 1	Inadequate sample size; 'easy'; good to use friends
Code 0	Misinterpretation; no reason or logic

MVE6. Claire set up a booth outside of the tuck shop. Anyone who wanted to stop and fill out a survey could. She stopped collecting surveys when she got 60 kids to complete them. What do you think of Claire's survey?
__Good __ Bad __ Not Sure - Why?

Code 3	Non-representative
Code 2	Uncertainty; adequate sample size
Code 1	Inadequate sample size; fairness; free choice; assuming range and variation; 'easy'
Code 0	Misinterpretation; no reason or logic

MVE7. Who do you think has the best survey method? - Why?

Code 2	Shannon or Shannon plus another
Code 1	Raffi, Claire, etc., with reason
Code 0	Raffi, Claire, etc., with no reason or logic

MVE8. What percent of students in the whole school will buy a raffle ticket? - (Circle one)

a. 35% (Shannon's result) because ____________________
b. 90% (Jake's result) because ________________________
c. 50% (Adam's result) because _______________________
d. 75% (Raffi's result) because _______________________
e. 95% (Claire's result) because _______________________
f. I think it is best to average the 5 surveys. The average of the kids that said they would buy a raffle ticket is 69%.
g. I don't know because Raffi, Shannon, Claire, Jake and Adam all got different results.
h. I think that __percent of the kids in the whole school are willing to buy a raffle ticket because_______

Code 3	Unpersuaded by new information – Shannon; influenced choice from another earlier
Code 2	Average them with or without doubt
Code 1	Uncertainty with or without doubt; Unpersuaded by new information from inappropriate choice
Code 0	Misinterpretation; idiosyncratic; no reason or logic

Q10. A primary school had a sports day where every child could choose a sport to play. Here is what they chose:

	Netball	Soccer	Tennis	Swimming
Girls	30	5	15	10
Boys	0	20	18	20

Q10A. How many girls chose tennis?

Code 2	15
Code 1	Other
Code 0	NR

Q10B. How many boys chose netball?

Code 2	0
Code 1	Other
Code 0	NR

Q10C. How many children chose swimming?

Code 2	30
Code 1	Other
Code 0	NR

Q10D. In which sport were boys and girls most evenly divided?

Code 2	Tennis
Code 1	Other
Code 0	NR

Q10E. Were there more girls or more boys at the sports day? How do you know?

Code 4	(girls) 60 vs 58 totals correct
Code 3	(girls/boys/other) totals error add/count
Code 2	(g) other reason
Code 1	(boys/other) other reason
Code 0	NR

Q17. A farmer wants to know how many fish there are in his dam. He took out 200 fish and tagged each of them, with a coloured sign. He put the tagged fish back in the dam and let them get mixed with the others. On the second day, he took out 250 fish in a random manner, and found that 25 of them were tagged. Estimate how many fish are in the dam.

Code 1	2000
Code 0	Other response, NR

Q20. Every year, Susan selects about 5 young actors for the drama team who perform brilliantly at audition. Unfortunately, most of these kids turn out to be no better than the rest. Why do you suppose that Susan usually finds that they don't turn out to be as brilliant as she first thought?

(a) In her eagerness to find new talent, Susan may exaggerate the brilliance of the performances she sees at the audition.
(b) The actors probably just made some nice acts at the audition that were much better than usual for them.
(c) The actors probably coast on their talent alone without putting in the effort for a consistently excellent performance.
(d) The actors who did so well at the audition may find that the others are jealous, and so they slack off.
(e) The actors who did so well are likely to be students with other interests, so they don't put all their energies into acting after the audition.

Code 1	b, b + another
Code 0	a, c, d, e, other (multiple letter not including b, or including all)

RAN3. What things happen in a "random" way?

Code 3	Definition + Example; "To pick without any pattern"
Code 2	Definition – No order, choose any, unpredictable; Multiple Examples from different aspects below
Code 1	Example – Natural (Weather), Human design (Breath testing), Game/selection (Tattslotto)
Code 0	Inappropriate (ransom, fighting, everything); Chosen (weak), in order, random numbers/alphabet, NR

SM19. Mrs. Jones wants to buy a new car, either a Honda or a Toyota. She wants whichever car will break down the least. First she read in Consumer Reports that for 400 cars of each type, the Toyota had more break-downs than the Honda. Then she talked to three friends. Two were Toyota owners, who had no major break-downs. The other friend used to own a Honda, but it had lots of break-downs, so he sold it. He said he'd never buy another Honda.

Which car should Mrs. Jones buy?
(T) Mrs. Jones should buy the Toyota, because her friend had so much trouble with his Honda, while her other friends had no trouble with their Toyotas.
(H) She should buy the Honda, because the information about break-downs in Consumer Reports is based on many cases, not just one or two cases.
(=) It doesn't matter which car she buys. Whichever type she gets, she could still be unlucky and get stuck with a particular car that would need a lot of repairs.

Code 3	H
Code 2	=
Code 1	T
Code 0	NR

SMP3. What does "sample" mean? Give an example of a "sample".

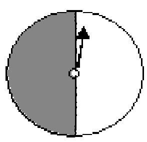
Code 3	Small part of whole to test of taste
Code 2	Small part of whole, part to test
Code 1	Test; try, piece, part
Code 0	Inappropriate; idiosyncratic; no response

SMP4. If you were given a "sample", what would you have?

Code 3	Small part of something to test
Code 2	Part of something, test of something, piece of carpet, taste of cheese
Code 1	Part, piece, test, example carpet
Code 0	NR

SP1. A class used this spinner. If you were to spin it once, what is the chance that it will land on the shaded part?

Code 2	50%, 1/2, 5/10, 1 in 2 chances, 50/50, half, same as white
Code 1	1 in 10, 80%, 20 out of 50, alright, any chance, bad chance
Code 0	I don't know



SP2A & B. Out of 10 (50) spins, how many times do you think the spinner will land on the shaded part? Why do you think this?

Code 3	Variation in one or both of response & answer
Code 2	Strict probability, implicit chance, at least 25; you can't tell (theoretically correct)
Code 1	Illogical or no reason with reasonable number
Code 0	NR

SP3A & B. If you were to spin it 10 (50) times again, would you expect to get the same number out of 10 (50) to land on the shaded part next time? Why do you think this?

Code 3	Sophisticated or simple recognition of variation
Code 2	Anything can happen, strict chance, implicit chance, contradiction
Code 1	Intuitive & primitive theories; Personal ideas & experiences
Code 0	Yes, just guessing

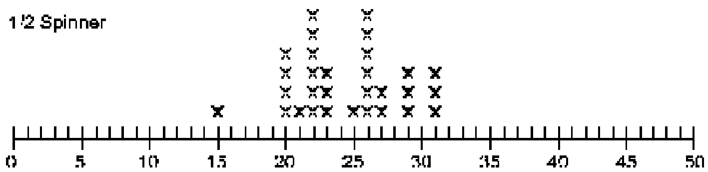
SP4A & B. How many times out of 10 (50) spins, landing on the shaded part, would surprise you?

Code 1	Grade 3 & 5 – 0,1,2,8,9,10; Grade 7 & 9 – <20, >30
Code 0	Grade 3 & 5 – 3,4,5,6,7; Grade 7 & 9 –20 to 30, Ambiguous, misinterpretation

SP5A & B. Suppose that you were to do 6 sets of 10 (50) spins. Write a list that would describe what might happen for the number of times the spinner would land on the shaded part?

Code 2	SD = 0.6-2.3 (10), SD = 1.3-5.0 (50)
Code 1	SD <0.6 >2.3 (10), SD <1.3 >5.0 (50), strict probability
Code 0	Out of range, misinterpretation

SP6. A class did 50 spins of the above spinner many times and the results for the number of times it landed on the shaded part are recorded below.
What is the lowest value?



Code 1	15
Code 0	Values with only one X above them; "1", values that have no X's above them; "0"; NR

SP7. What is the highest value?

Code 1	31
Code 0	Values with 6 X's above them; "6", "50"; no apparent logic

SP8. What is the range?

Code 1	16, 15-31
Code 0	20-27; 31; 50; 3; don't know

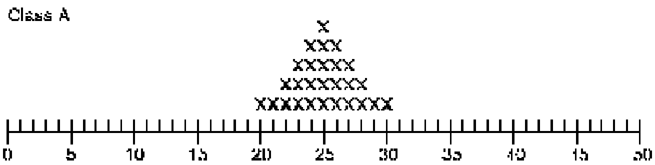
SP9. What is the mode?

Code 1	One or both of 22 and 26
Code 0	No logical reason; don’t know

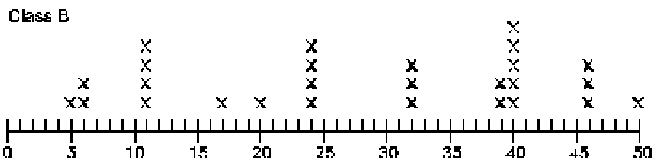
SP10. How would you describe the shape of the graph?

Code 1	Acknowledges variation; Focuses on physical objects, geometric shapes
Code 0	Focus on graph type or axes; illogical; NR

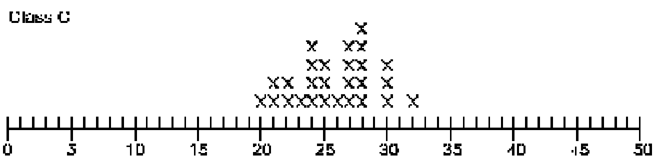
SP11. Imagine that three other classes produced graphs for the spinner. In some cases, the results were just made up without actually doing the experiment.



a) Do you think class A’s results are made up or really from the experiment?
__Made up
__Real from experiment
Explain why you think this.



b) Do you think class B’s results are made up or really from the experiment?
__Made up
__Real from experiment
Explain why you think this.



c) Do you think class C’s results are made up or really from the experiment?
__Made up
__Real from experiment
Explain why you think this.

Code 2	“Made up: It would never be so even. Made up: Too spaced out. Real: Cause it’s right.” (3 correct)
Code 1	1 part incorrect with reasons; 2 parts incorrect with somewhat sensible reasoning
Code 0	Anything can happen; no reasoning for choices; NR

TBL. A primary school had a sports day where every child could choose a sport to play. Here is what they chose.

	Netball	Soccer	Tennis	Swimming	TOTAL
Boys	0	20	20	10	50
Girls	40	10	15	10	75

TBL1. How many girls chose Tennis?

Code 1	15
Code 0	Number other than 15; idiosyncratic

TBL2. What was the most popular sport for girls?

Code 1	Netball
Code 0	Partly correct (2 sports including Netball, 40); idiosyncratic, NR

TBL3. What was the most popular sport for boys?

Code 2	Soccer & Tennis
Code 1	Soccer or Tennis
Code 0	20; Sport other than Soccer or Tennis; or a number other than 20; NR

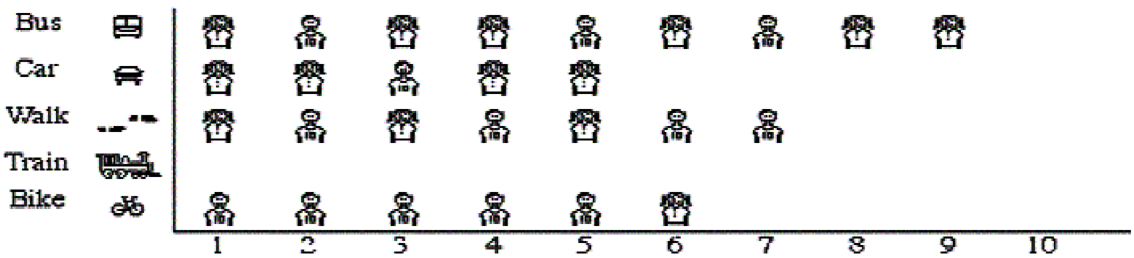
TBL4. How many children were at the sports day?

Code 1	125; Table reading only (50, 75)
Code 0	Computational fault; idiosyncratic comment; NR

TBL5. The teacher wanted to choose four children to lead the closing parade. Suggest two fair ways she could have chosen them.

Code 4	At least one method combining random selection & stratification; At least one combining random selection & stratification
Code 3	One chance method and various other possibilities
Code 2	At least one like "2 out of swimming, girl & boy"
Code 1	E.g., "The winners of each events or captains"
Code 0	Idiosyncratic methods like "They play the girls game first"

TRV. How children get to school one day
Number of students



TRV1. How many children walk to school?

Code 1	7
Code 0	Incorrect within range; odd comments

TRV2. How many more children come by bus than by car?

Code 1	4
Code 0	Bus; 9, 5, 14, a few

TRV3. Would the graph look the same everyday? Why or why not?

Code 1	Realistic or potential recognition of variation
Code 0	No variation or no reasoning

TRV4. A new student came to school by car. Is the new student a boy or a girl? How do you know?

Code 3	Explicit uncertainty - Probably a girl – More girls get a car to school; implicit - Girl – There is more chance of it being a girl
Code 2	Majority (local or global); Balance (local or global)
Code 1	Pattern in graph, could be either
Code 0	Not enough information, misinterpretation

TRV5. What does the row with the Train tell about how the children get to school?

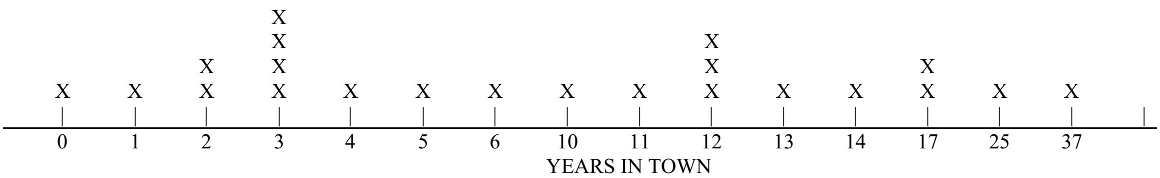
Code 2	"You can get to school by it."
Code 1	Direct interpretations, Geographical / historical assumptions, likes & dislikes
Code 0	Misinterpretations / Idiosyncratic, NR

TRV6. Tom is not at school today. How do you think he will get to school tomorrow? Why?

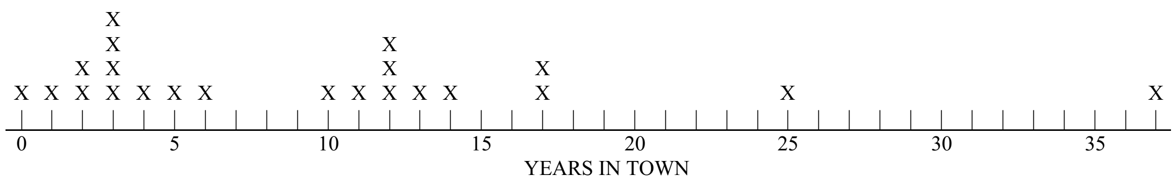
Code 5	"Probably by bus – Because 1/3 of the children caught it today." (uncertainty stated)
Code 4	Gendered or non-gendered majority; Bike, majority of boys; bus or walk are more common.
Code 3	Balancing using train or other transport; anything can happen
Code 2	Placing Tom (patterns) / Finding Tom
Code 1	No variation: Same as yesterday; Not enough information: Can't tell
Code 0	Misinterpretation, NR

TWN. A class of students recorded the number of years their families had lived in their town. Here are two graphs that students drew to tell the story.

Graph 1



Graph 2



TWN1. What can you tell by looking at Graph 1? – (2 spaces provided)

Code 3	2 summary comments; e.g., "Well, some people haven't lived there long" & "Some people have lived up to 17 years"
Code 2	1 summary plus perhaps data reading
Code 1	2 data reading comments (e.g., "There is only one in column 1" & "Column 3 has four crosses") or one appropriate and one inappropriate comment
Code 0	Graphing or idiosyncratic (e.g., "The graph is very spacey")

TWN2. What can you tell by looking at Graph 2? (2 spaces provided)

Code 3	2 summary comments
Code 2	1 summary plus perhaps data reading
Code 1	2 data reading comments or one appropriate and one inappropriate comment
Code 0	Graphing or idiosyncratic

TWN3. Which of these graphs tells the story better? - Why?

Code 3	"Graph 2 – You can see the difference between years more clearly and the graph is more spaced out"
Code 2	Indifference ; Personal preference; lack of logical reasoning
Code 1	Focused on graph spread / lay out; personal preference
Code 0	Statistically inappropriate choice with inappropriate or no reasoning, NR

VAR. What does "variation" mean? Use the word "variation" in a sentence. Give an example of something that "varies".

Code 3	"Varying is when something doesn't stay the same all the time – it varies" "That dress is a variation of the one I bought here last summer" "Clothes vary"
Code 2	More sophisticated definition with inappropriate sentence usage, or Simple but clear understanding reflected in definition
Code 1	Definition attempted, or Example given only with confused definition
Code 0	Idiosyncratic / Tautological, NR

TOWARDS A DATABASE OF RESEARCH IN STATISTICAL EDUCATION

FLAVIA JOLLIFFE
University of Kent
F.Jolliffe@kent.ac.uk

SUMMARY

Definitions of research in statistical education are discussed. A system of keywords for categorising statistical education research is outlined. Proposals for a Web-based survey of statistical education researchers to collect details of their research activities in statistical education, and the design of a database to store these details are described.

***Keywords*: Statistics education research; database; Web survey**

1. INTRODUCTION

The development of a project to collect information about statistical education researchers and their research, and to store it in a database, is described in this paper. This project is the author's response to an invitation to teachers in higher education to get involved in activities of interest to the Learning and Teaching Support Network (LTSN) Centre in Mathematics, Statistics and Operational Research in the United Kingdom (Davies, 2000). One of the suggestions made in the invitation was that an international survey of research into pedagogic issues in statistics and operational research would be useful and productive. The dictionary definition of pedagogy is that it is the science of teaching. Teaching and training go hand in hand with learning, and neither teaching nor learning can be completely separated from assessment. Statistical education researchers are concerned with all of teaching. learning, and assessment, so a survey of statistical education research is in fact a survey of research into pedagogic issues in statistics.

In the literature both of the terms "statistics education research" and "statistical education research" tend to be used with similar meanings. However, statistical education research includes research into such topics as statistical reasoning and thinking and so is wider in scope than statistics education research, which implies research into education in statistics. Statistical education research therefore includes statistics education research. The project described in this paper is concerned with statistical education research in the wider sense, but both terms are used in this paper as appropriate. A distinction is sometimes made between statistics and probability, but for the purposes of this project statistical education is taken to include probability education. Indeed many papers on probability education are presented in the meetings organised by the International Association for Statistical Education (IASE).

There are as yet relatively few outlets for publication and presentation of statistical education research activities and results so that researchers tend to feel isolated, and finding out what others are doing is partly a matter of chance. In consequence important research findings do not always get as widely disseminated as they deserve, and unbeknown to one another, researchers could be doing similar studies (ignorant duplication), but without the advantages offered by the discussion and comparison of methodology and by the comparison and pooling of results (deliberate replication). An easily accessible and widely available electronic database giving details of research and researchers in statistical education would be a valuable resource for both current and future researchers in this field, and for those who teach statistics. An international survey to find out who the statistical education

Statistics Education Research Journal 2(2), 47-58, http://fehps.une.edu.au/serj

researchers are, and details of their past, current and planned statistical education research could provide the initial input to such a database, and has become the main aim of the survey.

This is an ambitious project and presents several challenges. These include defining statistical education research, consideration of the database, and the implementation of the survey, as discussed in this paper in sections 2, 3, and 4 respectively. It might be argued that in itself the project constitutes statistical education research. Certainly some of the methodology is relevant to statistical education researchers and has wider applicability than in the context presented here. The results of the survey will indicate the directions which research is taking, and perhaps where research is needed, that is, will indicate what shape statistical education research might take in the future. Both the survey and the database will help strengthen the statistical education research community, will help make statistical education researchers more visible as a group, and will facilitate communication and collaboration among researchers.

2. STATISTICAL EDUCATION RESEARCH

2.1. CAN STATISTICAL EDUCATION RESEARCH BE DEFINED?

The author posed the question "What is research in statistics education?" and some supplementary questions at a round table discussion meeting held at the fifth International Conference on Teaching Statistics (ICOTS5) (Jolliffe, 1998). Similar questions were also being asked at the 39th session of the Italian Statistical Society by Ottaviani (1998) at about the same time. Although those who are interested in research in statistical education are likely to recognise when a study falls under this heading, the author is unaware of a definitive statement or description of what such research is. Statistical education research might be defined by examples of the research which is done under this heading, and of areas where further research is needed as suggested in Batanero, Garfield, Ottaviani and Truran (2000) and in the reactions to their paper in the January 2001 issue of the S*tatistical Education Research Newsletter* (SERN). These papers also contain other ideas which help with a definition and this section makes reference to several of these. These papers are publicly available on the Web site http://www.ugr.es/~batanero/sergroup.htm. The kinds of papers suggested as suitable for the *Statistics Education Research Journal* (see the SERJ Web page), and the classification theme outlined in section 3.2 of this paper, are also indications as to what constitutes statistical education research.

Defining what is meant by research in statistical education is not easy as can be seen by considering the study, which was set up in 1993 by the International Commission of Mathematical Instruction (ICMI), on the nature of research in the related field of mathematics education. This resulted in a 576 page book (Sierpinska & Kilpatrick, 1997). As those involved in the ICMI study found in the case of mathematics education research, asking what research in statistics education is leads immediately to a number of supplementary questions, and further thought about these leads to yet more questions. As suggested in Jolliffe (1998), some of the general questions are: "Is there a unity to the different activities which take place under the heading of research in statistics education?", "What are the research questions in statistics education?", "Are there research questions which are specific to statistics education?", "What are the results of research in statistics education?", "What criteria should be used to evaluate research in statistics education?". Questions more directly concerned with teachers of statistics are: "How many teachers have access to research results and how many read these?", "Do research results influence training in the teaching of statistics?", "Are statistics teachers competent to do research in statistics education?", and "Does research positively inform practice?".

In Jolliffe (1998) the author stated that some statisticians are reluctant to recognise statistics education as a field of academic research, and she also suggested that it was timely to proclaim the existence of research in statistics education as a research discipline in its own right. Pfannkuch (2001) put forward the view that statistics is a relatively new discipline in academic programmes and some groups are still "grappling" with it. She thought that this could be a barrier to the acceptance of the new discipline of statistics education research. Batanero et al. (2000) feel that some academics believe that education as a discipline has nothing to contribute to knowledge. On the other hand Glencross (2001) feels that statistics education is already recognised internationally in its own right and not just as a subset of either statistics or education.

One of the chief aims of statistical education research is that findings would be used to improve teaching practice and students' understanding of statistics and their performance in it (Jolliffe 1998). For Glencross (2001) research in statistics education is research about the teaching and learning of statistics, and he suggests that one of the features of good research is that its results can be implemented in the teaching-learning arena. Konold (2001) feels that teachers ought to be the primary audience for research and that papers presenting the research should be about the educational implications. Batanero et al. (2000) comment that academics working in education need to understand and be understood by politicians and by leaders in industry and schools. It would be interesting to ask respondents to the survey for their views on what constitutes research in statistical education.

Ottaviani (2000) comments that research into statistical education requires the development and use of appropriate statistical methods. This gives researchers in statistical education the opportunity to advance the discipline of statistics itself, and hence become recognised as educational statisticians. If statisticians are involved in educational research in other disciplines (as well as, or instead of, statistics education) the quality of educational research is likely to be higher and the status of such research would be raised. One area where researchers in statistical education might make a useful contribution is in developing methods for evaluating the effectiveness of a particular teaching approach. Batanero et al. (2000) remark that research methods and philosophies change over time, partly because research questions or available techniques change. They note that there is considerable experience as regards undertaking research in statistical education and so we might now be at a stage where it is possible to develop some general principles, including what background knowledge is needed in order to conduct quality research in statistical education. Bacelar-Nicolau (2001) comments that the question of what is research in statistics education has much in common with the more general question of "What is research in other sciences education?" She suggests that a good research topic would be finding features which are common to research in "other sciences" (and statistics) education, and this might lead us to suitable common methods of research and help us search for new and richer approaches. Statistical education researchers might well find it useful to look at research methods used in other areas of educational research.

Glencross (2001) points out that any research takes place within a research paradigm, and that educational researchers tend to distinguish the scientific, critical-theoretic, and interpretative paradigms among the many ones possible. He says that because of the multi-faceted nature of both statistics education and statistics education research, we must accept and tolerate the diversity of research traditions and methodologies that exist.

In a plenary address at ICOTS6 Watson (2002) gave examples of three types of research in statistics education which she described as theoretical, qualitative, and quantitative. She remarked that research in statistics education is more broadly based than classical statistics applied to science. She also commented that when teachers of statistics turn the focus on themselves as statistics educators they run the risk of forgetting some of the fundamental principles of good research.

It is the author's opinion that research in statistical education must be involved with at least one of learning, teaching, and assessment of statistical methods or statistical thinking. Assessment includes studies probing people's understanding of concepts as well as assessment of those studying statistics. More importantly, there obviously has to be an element of research in statistical education research. This could be, for example, an experiment (for example, Hilton & Christensen, 2002), a survey of people or of publications (for example, a survey of internet sites is described in Gal, 2003), an observational study, development of a model or instrument (Garfield, 2003), or even a well thought-out proposal for a research study. The keywords for methodology/type of research shown in the Appendix provide a more comprehensive list of types of research. A paper suggesting that, for example, a particular method of teaching a statistical topic is a good method, is not in itself research, but a report on the implementation of the method in the classroom and some attempt at evaluating its effectiveness, with perhaps suggestions of a research design to evaluate the method more fully, could well count as statistical education research. It is perhaps worth mentioning that classroom teachers do not always have the opportunity to do randomised experiments and the only evaluation possible might be a comparison with a different method used with different students in a different time period. Such studies can still be of value and are still research. Chance and Garfield (2002) discuss a number of ways of obtaining research data in statistics education research.

2.2. WHO DOES STATISTICAL EDUCATION RESEARCH?

Researchers in statistical education do not yet have a clear identity, although they are starting to make an impact, and with the growth of IASE activities are gaining recognition as a group. Statistics teachers, particularly those at the tertiary level, are probably the most active group of researchers in statistical education, but it is thought that not all of those who teach statistics would describe themselves primarily as statisticians, as, for example, statistics might be taught by a psychologist or a mathematician. It might be said that any teacher of statistics who thinks about how and what they want to teach in statistics courses is a potential researcher in statistical education. The distinction between engaging in the activity of teaching, which often involves developing courses and trying different methods of presentation, and being an active researcher into teaching and learning, is small. Thus if statistics teachers say that they have research interests in statistical education it is not immediately obvious that this is any different from what teachers of statistics might do as part of their normal work, or whether their interest is mainly as consumers, rather than producers, of research.

Statistical education researchers tend to have backgrounds in at least one of mathematics, statistics, psychology, or education. They work in a variety of departments reflecting the multi-disciplinary nature of statistics, and research groups have no natural academic home. A statistical education researcher is often the only one in his or her institution. Researchers also come from many different countries and belong to different cultures. Characteristics of respondents to the survey will be a guide as to what kinds of people do statistical education research, subject to finding potential respondents and to the response rate and resulting biases (see sections 4.1 to 4.3 for further discussion of this).

Some researchers in education, particularly researchers in mathematics education, now participate also in statistics education research, the research tending to concentrate on children and on the learning of concepts. Research done by psychologists on stochastic reasoning is very relevant to statistics education. Some researchers in the area of educational technology, which tends to mean computer assisted learning and, more recently, multi-media learning, are concerned with statistical education. Some developers of software for doing or learning statistics, or for assessing statistical knowledge, do research into the requirements of software and into its effectiveness, and as such are statistical education researchers. Thus just as statistics and education are multi-disciplinary in nature those who do statistical education research are a varied group reflecting this.

To be successful in statistical education research, researchers need to be good at research, but not necessarily good at research in the theory and methods of statistics. They need knowledge of the methodology they are using in their research and of the statistical topics they are researching, and familiarity with the processes of teaching and/or learning, which could be from practical experience or from a theoretical perspective. They also need to work with educators across all disciplines where statistics is used or taught. Cooperation and contact with other statistics education researchers is important, and is likely to improve the quality of the research and make it easier to get funding.

3. THE NATURE OF THE DATABASE

3.1. GENERAL

As the main aim of the proposed survey is to provide initial input to an electronic database, the potential contents of the database and the anticipated requirements of users help determine many other aspects of the study. Society as a whole is fast becoming internet dependent, and although access to the internet is difficult or impossible in some parts of the world, it is thought that the majority of those interested in statistical education research have the facility to obtain information from the World Wide Web. This is certainly the case for those based in the UK, for whom the LTSN has been set up. This points to placing a database of statistical education research and researchers on the Web, with all the advantages that offers over a printed version, such as ease of searching for items, and flexibility in making changes. This will be hosted on the LTSN Centre in Mathematics, Statistics and Operational Web site (http://ltsn.mathstore.gla.ac.uk).

The author has looked at two databases in fields related to statistical education from a user's perspective. One is the MATHDI (Mathematical didactics) database (http://www.enis.de/MATH/DI/en/quick.html). This is a large database, but a subscription has to be paid in order to have unlimited use of it and it is necessary to enter search-words (König, 2003). This means that users have to have a good idea what they are looking for. The other database is of reviewed educational research in mathematics, statistics, and operational research (http://ltsn/mathstore.gla.ac.uk/resource_collection). It is hosted on the same LTSN Web site as the database proposed in this paper will be. It does not yet have many entries, but it is free and can be searched by title, and by keywords, and contains a list of authors. Texts can also be searched. This is the form envisaged for the statistical education database.

It is thought likely that mostly users would want to search the statistical education database by at least one of researcher, topic of the research, and the methodology used in the research. Consideration of these requirements suggests that the information collected in the survey would best be stored in a relational database having two main tables, one for researchers and one for publications and products. These would be linked through a table containing only pairs of unique researcher and unique publication identification codes (see Figure 1). Publication has to be interpreted broadly to cover both hard copy and electronic versions of research activities. To avoid repetition of data such as an address of an institution common to more than one researcher, or details of conference proceedings containing several papers, the data structure would be simplified by storing data relating to items of this nature in separate files. The design would need to allow for the addition of details of new researchers and new research, and of details relating to past research as they come to light, and to take account of updating, for example of changes of address, and of the need to retain historical data. It would be useful at some stage to extend the database to include research by those who were not active at the time of the survey. Certainly there is no intention to delete details relating to researchers who become inactive.

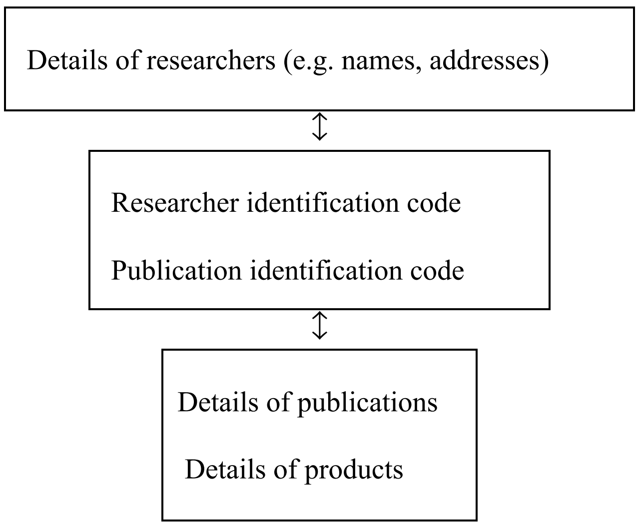


Figure 1: The form of the relational database

3.2 KEYWORDS

It is envisaged that users would search the database via a user interface on a Web page, selecting from given keywords. Information would then be extracted from tables in the database. Respondents to the survey would be asked to use the same set of keywords to classify their research into categories and to suggest other keywords if those provided were inadequate. In the current working model the main headings are: *Teaching level/stage, Type of student, Syllabus/curriculum, Focus of research, Technology,* and *Research details*. These go from classification of the type of person receiving statistical education (includes statistical literacy), through the content of the education and matters related to delivery, to matters more directly related to the research. Each of these areas is broken down into categories and some categories are further broken down. For example, *Focus of research* has as categories: *Method of delivery/learning, Teaching approach, Assessment*, and *Pedagogic issues*. The Appendix shows the proposed breakdown for *Research details* at the time of writing. This is not thought to be in the final form yet.

The order in which key-words appear in the proposed scheme is fairly arbitrary and the numbering is for administrative convenience only. This classification scheme is experimental to some extent. It is expected that it will evolve over time in the light of experience in using it and in response to users' comments. Some categories might be added. Some fine details have yet to be filled in. For example under *Teaching level/stage* specific stages, ages or age-groups will be determined by the research studies, in continuing professional development for professions other than statistics specific areas of work will be added as required. Finding appropriate keywords has not always been easy and some might be changed. Taking account of the terms used in different cultures and in different countries is particularly difficult as researchers with different backgrounds do not necessarily understand the same word in the same way, and some terms are unique to specific groups. Alternative terms will be given as keywords where these are known. Short explanations of terms will also be given in the hope that this will reduce misclassifications of research and misunderstandings.

To show how the scheme might work two papers given at the 53rd session of the ISI (Blumberg, 2001; Ben-Zvi & Arcavi, 2001) have been classified by the author and depend on her interpretation of their research in consultation with the authors of these papers. Suggested key-words are in italics. The paper by Blumberg (2001) is relatively easy to classify, but showed how it is important to take account of differences in terminology used in different countries. Blumberg did a *survey of post-secondary teachers* and the focus of her research was the *syllabus* followed by *undergraduates* (teaching level/stage) who were *not statistics or mathematics specialists* (type of student). The paper by Ben-Zvi and Arcavi (2001) is less easy to classify and is an example of the use of newer research methods. It involves both a *case study* and *testing* of *seventh grade* students (age 13), is looking at the *method of learning* by the students and also a *data based teaching approach*. From the report in the paper it appears that the students were participants in an *experiment without controls*. The paper also touches on the *curriculum*, and mentions that the research had relied on *cognitive* and *socio-cultural* perspectives. Researchers would be free to choose as many or as few keywords to describe their research as they wished, but would be encouraged to ask if they were unsure about any matters connected with this.

It should be emphasised that no evaluations of the quality of the research will be given on the database, nor, initially, will there be descriptions of the research apart from key words as outlined above. In this sense the database would be a kind of directory. It would complement bibliographies on related topics such as Sahai, Khurshid, and Misra (1996) and Holmes (2002), and would serve a different purpose from the summaries of publications which were published in the *Statistical Education Research Newsletters* and which have been published in this journal. However, the Web offers the opportunity of linking to other sites, so that an obvious development would be to have links to other Web pages, perhaps to those of the researchers, where abstracts or full versions of publications might be available.

4. THE SURVEY

Surveys are usually done for a purpose so it is advisable to think about the final product and dissemination of results at planning stage. The different stages involved are inter-related, and consideration has to be given also to the target population, to the sample design and sampling frame, and to the design and method of administration of the questionnaire. As already stated, the aims of the survey under discussion are to find out who the statistical education researchers are and details of their research, and the results would be disseminated via an electronic database as discussed in section 3. The implementation of the survey is discussed in this section. As the survey is primarily a fact-finding exercise, questions of inference are of secondary interest.

Ideally anyone who is currently undertaking or planning statistical education research or who has done such research in the past is a potential contributor to the database, which implies a census rather than a survey. In practice a census is impossible as there are no ready-made lists of statistical education researchers. As participation will be voluntary those who are included on the database will be a self selected sample, using their own definitions of statistical education research. However, a broad definition of research would be accepted (see section 2.1 for some discussion of this issue).

4.1. FINDING THE MOST ACTIVE STATISTICAL EDUCATION RESEARCHERS

The first stage in obtaining contact details of statistical education researchers in this project is making announcements on email lists such as Stat-Ed and the Teaching Statistics discussion list (hosted by JISC). Those interested in contributing details of their research will be asked to get in touch. Other announcements might be made in publications such as *Teaching Statistics* and the on-line *Journal of Statistical Education.* Those responding to announcements will be directed to a questionnaire asking for basic details (see 4.3).

As it is possible that not all those who have an interest in statistical education will see the announcements, other lists will be used at a later stage to supplement the list of statistical education researchers obtained through announcements. These include lists of members of the International Association for Statistical Education (IASE) and participants in the International Conferences on Teaching Statistics (ICOTS). Published research is yet another way of finding out who does statistical education research (see also 4.2). Initial contact with potential respondents provides a good opportunity to tell them about the survey and to get their agreement to participate.

Most of the sources of finding statistical education researchers mentioned are biased towards the English speaking world, and this could mean that some important research is excluded. As the project is sponsored by a UK based organisation this is not inappropriate, even if a little insular. It should be emphasised that there is no intention to exclude researchers who write in a language other than English, but it might be difficult to make contact with them in order to ask them to contribute to the database.

Initial contact with potential respondents is planned to be by email. However, email addresses are often out of date and sometimes even a slight difference in an email address can result in non-contact. Another frustration is that servers do not always communicate with one another and can be unreliable, so those who cannot be contacted by email, for whatever reason, would be contacted by post or telephone. Some non-response might occur at this stage, for example because the person is too busy to respond, or is not interested in participating. Further attempts would be made to reach these people.

4.2. FINDING THE ELUSIVE STATISTICAL EDUCATION RESEARCHERS

The more obvious ways of making contact with statistical education researchers mentioned in 4.1 are likely to reach a large proportion of the most active researchers, but there are still likely to be some who are missed, so at a later stage an attempt to reach others will be made. The majority of those who are visible in doing research in statistical education are university teachers or researchers, and it is relatively easy to access lists of these, although such lists are sometimes out of date and are unlikely to indicate those who are statistical education researchers. Many school teachers probably engage in research-like activities, although they do not necessarily realise this or bring their work to the attention of the larger community of statistical education researchers. Finding members of this elusive but important minority group is not easy due to the difficulty of obtaining lists, and the small number in a large number of school teachers likely to be statistical education researchers. It would be impractical on a world-wide basis.

A first step in this later stage could be to amalgamate all lists thought to contain a high proportion of researchers in statistical education, including those used in the initial trawl. See section 2.2 for a discussion of who might be in this group. Some people will be on more than one list, but sorting names on the combined list into alphabetical order, relatively easy to do with lists in electronic form, will make it easier to spot and eliminate duplicates. Misspelt names and matches of name where there are different addresses are likely causes of failure to match. Human beings are fairly good at spotting matches between records where there are slight differences in the information, possibly using external information to help in deciding whether there is a match. Computerised record matching systems have been less successful, but need to take account of the statistical characteristics of the errors which occur in computer records (Copas & Hilton, 1990). The matching problem is easy to state but hard to solve. It would be necessary to contact those on the list so produced, who had not already responded, to identify the statistical education researchers.

Taking into account the lists that are available and the fact that unknown researchers might be considered to be a rare population, a combination of snowball sampling and of screening lists likely to contain names of statistical education researchers would be an appropriate method of building up a list of members of the target population (Kalton & Anderson, 1986). To use snowball sampling in this survey, known researchers in statistical education would be asked for names and contact details of other statistical education researchers. Further researchers discovered in this way would then be asked for names and contact details of other researchers. The process is repeated until the number of additional researchers discovered is negligible (see Figure 2). It could well be a slow process and not yield many more names, but it is thought that it would be worthwhile to make an attempt. In order to make the implementation of the procedure as smooth as possible only those on email will be asked for names of other researchers.

1. Contact known statistical education researchers for names and contact details of other statistical education researchers.

2. Ask the additional researchers found for names and contact details of other statistical education researchers.

3. Repeat the process until the number of “new” researchers found is negligible.

Figure 2: Using snowball sampling to find researchers

Two obvious problems with snowball sampling are that those researchers who are well known will be on the initial list, and will be named by many other researchers, so that this procedure will be rather wasteful, and those researchers who are isolated and known to only a few are fairly unlikely to be discovered by the procedure. One way of avoiding the naming of those who are already known to be researchers in statistical education is to circulate a list of these when asking for names of other researchers. Checking and dropping duplicates would be a similar process to that done when drawing up the initial list.

It is hoped that occasional announcements publicising the database and a permanent invitation on the Web page for those not already on the database to get in touch would identify additional researchers. It is recognised that there is a risk of bias in this method of finding researchers and non-inclusion of some groups, particularly those without easy access to the internet.

4.3. CAPTURING THE DATA

As pointed out in section 3.1, it is thought that most of the people interested in statistical education research, of whom statistical education researchers are a subset, have access to the internet which suggests that the internet is an appropriate medium for collection of information. It tends to be both quicker and cheaper than the more traditional methods of collecting information by interviewers or by post, and has the advantage of easy conversion to data files. To some extent use of the internet for conducting surveys is still in its infancy, and little has yet been published in the statistical literature regarding internet-based survey methodology (Ohsumi & Yoshimura, 1999; Witmer, Colman, & Katzman, 1999; Mitofsky, 2001). Papers tend to be given at conferences or published in journals for those in the market research industry or computing (see for example websm.org on Web survey methodology). A paper by MacElroy (1999) gives a useful comparison of the various methods of online surveying.

The initial decision is between email interviewing and Web-based interviewing. In email interviewing there is a choice between including the questionnaire as a plain text message or as an attachment. The former tend to look dull, whereas although attachments look more professional they do not always travel well. The advantage of a Web based survey is that data can be read directly into a database when they are received (Payne & Crawford, 2003), and clicking on a Web address in an email to access a questionnaire is no more effort than opening an attachment, but the questionnaire might be affected by both the software and the hardware used to view it.

It is thought that statistical education researchers would want to be included on the database and to provide updates without undue prompting as it would be in their interests to publicise their work, which suggests that response rates would be relatively high. However, there is a tendency these days to be overwhelmed by the volume of incoming emails, and busy people are just as likely to postpone responding to questionnaires sent via email or to ignore them as they are in the case of paper ones. Response rates in Web surveys tend to be low (Vehovar, Manfreda, & Batagelj, 2001). The response rate in a survey of members of the Teaching-statistics email list, which includes statistical education researchers, was about 27% (Green & Fuller, 1999).

The initial questionnaire to which those expressing interest in taking part in the survey will be directed is hosted on the LTSN Web site. The information required from respondents is mainly factual – contact details, and information about the main areas with which their statistical education research is concerned, using a greatly simplified selection of keywords. There are also questions asking if they are willing to provide further details, and for comments and suggestions. Once researchers have been identified a procedure for contacting them at later dates would be developed in order to update their entries on the database. At a later stage those who indicate their willingness to participate further will be asked for details of their statistical education research publications and invited to classify these according to the full predefined list of keywords (see section 3.2). They will also be asked about their future plans, and their agreement to providing updates would be sought.

As the survey is under the auspices of the LTSN the questionnaires are in English although this could mean the exclusion of some groups (see also 4.1). There is some evidence to suggest that the design of the form has some effect on response rates and the quality of response (Vehovar et al., 2001; Dillman, 2003), and this might need to be investigated. In theory it should be relatively easy for researchers to provide details of their research as many are likely to have lists of their publications in electronic form. However, different people will list their publications in different ways. Researchers might be unwilling to standardise their lists and classify their work into categories unless there is some incentive to do so. Would inclusion on the database as a service for the communities of researchers and teachers be sufficient reward?

At each stage work might be needed to put the responses into the form required and even classifying publications at the last stage. Respondents would be given feedback on how the information they had provided had been changed and in what form it was to be made available to others. Once the database was in existence they would be encouraged to look at it. Hopefully this would encourage them to update and provide more details concerning their own entries.

5. CONCLUSIONS

The Internet has made the world a smaller place, and has made it more feasible for researchers in statistical education to be an interactive community. It is in fact essential that researchers take full advantage of the newest technological developments. Statistical education in itself is undergoing substantial and rapid changes and it is important that research is done into the effects of these changes and that those concerned with providing education and assessing the outcomes are familiar with the research. References to the research need to be readily accessible. The proposed survey and the resulting database will therefore have an important contribution to make as regards the future of statistical education and of statistical education research.

ACKNOWLEDGEMENTS

This is an expanded and updated version of a paper given at the 53rd session of the International Statistical Institute in Seoul, Republic of Korea (Jolliffe, 2001). The author was given funding to attend this session when she was a lecturer at the University of Greenwich.

The Learning and Teaching Support Network (LTSN) Centre in Mathematics, Statistics and Operational Research in the UK is funding the project discussed in this paper, and the database will be hosted on the LTSN Web site. Ewan Crawford designed the proforma used for the Web version of the questionnaire (Payne & Crawford, 2003).

REFERENCES

Bacelar-Nicolau, H. (2001). On the paper "Research in statistical education: some priority questions". *Statistical Education Research Newsletter, 2*(1), 3-4. [Online:
http://www.ugr.es/~batanero/sergroup.htm.]

Batanero, C., Garfield, J. B., Ottaviani, M. G., & Truran, J. (2000). Research in statistical education: some priority questions. *Statistical Education Research Newsletter, 1*(2), 2-6. [Online: http://www.ugr.es/~batanero/sergroup.htm]

Ben-Zvi, D., & Arcavi, A. (2001). Developing experts' points of view on local-global approaches to data and data representations. In *Bulletin of the International Statistical Institute, 53rd Session Proceedings, Tome LIX, Book 2* (pp. 215-217). Seoul: International Statistical Institute.

Blumberg, C. J. (2001). Is there life after introductory statistics? In *Bulletin of the International Statistical Institute, 53rd Session Proceedings, Tome LIX, Book 2* (pp. 233-236). Seoul: International Statistical Institute.

Chance, B., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal, 1*(2), 38-41. [Online: http://fehps.une.edu.au/serj]

Copas, J. B., & Hilton, F. J. (1990). Record linkage: statistical models for matching computer records. *Journal of the Royal Statistical Society Series A, 153*, 287-320.

Davies, N. (2000). Promoting effective methods of learning and teaching statistics and operational research. *MathStats&OR,* 2, 2-3.

Dillman, D. (2003). How visual design influences responses to internet and other self-administered questionnaires. In *Bulletin of the International Statistical Institute 54th Session Proceedings* [CD-ROM]. Berlin*:* International Statistical Institute.

Gal, I. (2003). Expanding concepts of statistical literacy: an analysis of products from statistics agencies. *Statistics Education Research Journal, 2*(1), 3-21. [Online: http://fehps.une.edu.au/serj]

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 22-38. [Online: http://fehps.une.edu.au/serj]

Glencross, M. (2001). A reaction to "Research in statistical education: some priority questions". *Statistical Education Research Newsletter, 2*(1), 9-10. [Online:
http://www.ugr.es/~batanero/sergroup.htm].

Green, D., & Fuller, M. (1999). Analysis of the Teaching-statistics mailbase list. *Teaching Statistics*, *21*(2) 56-59.

Hilton, S. C., & Christensen, H. B. (2002). Evaluating the impact of multimedia lectures on student learning and attitudes. In B. Phillips (Ed.), *Developing a Statistically Literate Society.*

Proceedings of the Sixth International Conference on Teaching of Statistics [CD-ROM]. Cape Town: International Statistical Institute.

Holmes, P. (2002) Some basic references for the teaching of undergraduate statistics. *Statistics Education Research Journal*, *1*(2), 49-53. [Online: http://fehps.une.edu.au/serj]

Jolliffe, F. R. (1998). What is research in statistics education? In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. F. Wong (Eds), *Statistical Education – Expanding the Network. Proceedings of the Fifth International Conference on Teaching of Statistics.* (pp. 801-806). Singapore: International Statistical Institute.

Jolliffe, F. (2001). An international survey of research in statistical education. In *Bulletin of the International Statistical Institute, 53rd Session Proceedings, Tome LIX, Book 2* (pp. 201- 204). Seoul: International Statistical Institute.

Kalton, G., & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society Series A, 149*, 65-82.

König, G. (2003). Electronic access to literature in theory and practice of statistics education. In *Bulletin International Statistical Institute, 54th Session Proceedings* [CD-ROM]. Berlin: International Statistical Institute.

Konold, C. (2001). Reaction to "Research in statistical education: some priority questions". *Statistical Education Research Newsletter,* 2(1), 14. [Online: http://www.ugr.es/~batanero/sergroup.htm]

MacElroy, W. (1999). Comparing seven forms of online surveying. *Quirk's Marketing Research Review, article 0510.* [On line: http://www.quirks.com/index.asp]

Mitofsky, W. J. (2001). IPM17: Internet and innovative data collection. In *Bulletin of the International Statistical Institute, 53rd Session Proceedings, Tome LIX, Book 3* (pp. 79- 80). Seoul: International Statistical Institute.

Ohsumi, N., & Yoshimura, N. (1999). The online survey in Japan: an evaluation of emerging technologies. In *International Statistical InsIitute 52nd session, invited papers on IASS topics* (pp. 117-133). Helsinki: International Statistical Institute.

Ottaviani, M. G. (1998). Didattica della statistica. Quale futuro in Italia? (Statistics education: Is there a future in Italy?) In *Società Italiana di Statisitca, Atti della XXXIX Riunione scientifica, Comunicazioni Spontanee, Vol. II.* [CD-ROM]. Sorrento, Italy: Instituto di Statistica e Matematica.

Ottaviani, M. G. (2000). Research into statistical education. *Statistical Education Research Newsletter, 1*(1) 2-4. [On line: http://www.ugr.es/~batanero/sergroup.htm].

Payne, B., & Crawford, E. (2003). A Web questionnaire creator tool. *MSOR Connections*, *3*(3), 7-10.

Pfannkuch, M. (2001). Statistical thinking and research. *Statistical Education Research Newsletter, 2*(1), 18-20. [Online: http://www.ugr.es/~batanero/sergroup.htm]

Sahai, H., Khurshid, A., & Misra, S. C. (1996). A second bibliography on the teaching of probability and statistics. *Journal of Statistical Education*, *4*(3). [Online: http://www.amstat.org/publications/jse/v4n3/sahai.html]

Sierpinska, A., & Kilpatrick, J. (Eds.) (1997). *Mathematics education as a research domain: a search for identity. An ICMI study*. Dordrecht: Kluwer Academic Publishers.

Vehovar, V., Manfreda, K. L., & Batagelj, Z. (2001). Errors in Web surveys. In *Bulletin of the International Statistical Institute, 53rd Session Proceedings, Tome LIX, Book 1* (pp. 243- 246). Seoul: International Statistical Institute.

Watson, J. (2002). Doing research in statistics education: more than just data. In B. Phillips (Ed.), *Developing a Statistically Literate Society. Proceedings of the Sixth International Conference on Teaching of Statistics* [CD-ROM]. Cape Town: International Statistical Institute.

Witmer, D. F., Colman, R. W., & Katzman, S. L. (1999). From paper-and-pencil to screen-and-keyboard. Toward a methodology for survey research on the Internet. In S. Jones (Ed.), *Doing Internet research: critical issues and methods for examining the Net*, Thousand Oaks, London, New Delhi: Sage publications.

FLAVIA JOLLIFFE
Institute of Mathematics and Statistics
University of Kent
Canterbury
Kent, CT2 7NF
United Kingdom

APPENDIX: KEYWORDS FOR THE CLASSIFICATION OF "RESEARCH DETAILS"

1. Researcher and Contact Details

2. Methodology/Type of Research
    - 2.1. Experimental without controls
    - 2.2. Comparative study
        - 2.2.1. Concurrent (parallel groups)
        - 2.2.2. Before/after
    - 2.3. Observational
    - 2.4. Case study
    - 2.5. Exploratory e.g. of understanding
    - 2.6. Report and evaluation e.g. of an activity
    - 2.7. Survey
        - 2.7.1. of Students
        - 2.7.2. of Pre-secondary teachers
        - 2.7.3. of Post-secondary teachers
        - 2.7.4. Other
    - 2.8. Literature review
    - 2.9. Expository
    - 2.10. Critique
    - 2.11. Secondary analysis
    - 2.12. Reflective study
    - 2.13. Evidence based
    - 2.14. Systematic review
    - 2.15. Ethnographic

3. Statistical Methods Used in the Research
    - 3.1. Descriptive
    - 3.2. Inferential
    - 3.3. Modelling
    - 3.4. Multivariate
    - 3.5. Qualitative

4. Products and Details
    - 4.1. Written report
        - 4.1.1. Dissertation
        - 4.1.2. Paper
        - 4.1.3. Book/monograph
    - 4.2. Assessment instrument
    - 4.3. Software
    - 4.4. Equipment
    - 4.5. Teaching resource
        - 4.5.1. Text
        - 4.5.2. Student notes
        - 4.5.3. Data
        - 4.5.4. Case study
    - 4.6. Theoretical knowledge
        - 4.6.1. Model of cognitive development, learning or reasoning
        - 4.6.2. Categories of students' errors, conceptions etc.

5. Status of Research and Details
    - 5.1. Published
    - 5.2. Presented orally
    - 5.3. Ongoing
    - 5.4. Planned

SPSS TEXTBOOKS: A REVIEW FOR TEACHERS

JAMIE D. MILLS
University of Alabama
jmills@bamaed.ua.edu

SUMMARY

Many teachers and researchers use the Statistical Product and Service Solutions (SPSS) software for instructional and/ or research purposes. Because of the comprehensive nature and features of this program, there are various textbooks available that may offer teachers and practitioners a more concise way to analyze and discuss many of the topics that are typically taught in statistics courses. These textbooks differ on many different features, such as level of the audience, complexity of statistical procedures discussed, degree of interpretation of statistics/output, amount of detail discussed on the basic mechanics, accessibility of data files, and student exercises. This paper is written to offer teachers and researchers a review of some of the most popular SPSS textbooks that are available today by utilizing evaluation criteria previously discussed in the literature. This review can provide a starting point for teachers to explore features of the various SPSS textbooks as well as to consider what book is most appropriate based on their own teaching style. Comments from teachers who use the software, limitations of the review, and a table of other ancillary textbook data conclude the paper.

***Keywords*:** SPSS textbooks; evaluation criteria; teaching; statistics

1. INTRODUCTION

Over the past few years, there appears to have been a shift on how to teach statistics in a variety of different fields. In 1992, the American Statistical Association (ASA) and the Mathematical Association of America (MAA) formed a joint committee to study the teaching of introductory statistics. The main recommendations were to emphasize statistical thinking, incorporate data and emphasize concepts using less theory and fewer 'recipes', and to foster active learning (Cobb, 1992). According to Moore (1997), the most effective learning takes place when content, pedagogy, and technology reinforce each other. In particular, students should be active participants assigned with structured activities that focus on statistical concepts and ideas that are nonmathematical in nature.

In response to these and the many other recommendations regarding the teaching of statistics, a plethora of statistics textbooks designed to communicate and foster this style of teaching is very visible and widely available to teachers of statistics (Kendrick, 2000; Lomax, 2001; Moore, 2000; Weinberg & Abramowitz, 2002). Instead of focusing heavily on formulas and mathematical and statistical theory, these textbooks report to promote conceptual learning and understanding of statistics concepts and thus, a fundamental understanding of basic mathematics for the introductory student is adequate.

Another possible way to achieve the goal of conceptual learning for students is to perhaps use some type of software program to perform the calculations. Using the computer in this way may not only allow students to focus more on concepts by freeing them of computational tasks, but it may also help students to become active in their own learning, another important recommendation. Moore (1997) suggests that the use of technology helps to automate many routine operations which in turn facilitate the learning process. Many researchers and teachers agree (Garfield, 1997; Hoerl, Hahn, & Doganaksoy, 1997; Scheaffer, 1997) and others have reported that their students benefit academically

Statistics Education Research Journal 2(2), 59-70, http://fehps.une.edu.au/serj

when they have included assignments, in conjunction with instruction, that have involved utilizing data analysis procedures (Giesbrecht, 1996; Goodman, 1986; Gratz, Volpe, & Kind, 1993; Velleman & Moore, 1996; Weinberg & Abramowitz, 2002).

Because of the integration of the computer into statistics instruction, there have been many software programs to consider (SPSS, SAS, S-Plus, Minitab, Excel, Systat). In education and the behavioral and social sciences, the Statistical Product and Service Solutions (SPSS, formerly known as the Statistical Package for the Social Sciences) is a popular choice. SPSS is a fairly user-friendly statistics software program that is windows-driven, and offers users a point-and-click way to generate output. The program offers a base feature that provides the most basic statistical procedures and it also offers some advanced features, which allow the user to tackle more sophisticated statistical techniques. Users may also choose to utilize the syntax editor to write 'code' to target specific analyses as opposed to the point-and-click method of generating the output.

Almost as important as selecting an appropriate primary textbook for a statistics course is the now challenging task of selecting a suitable software textbook. A software 'textbook', in this case, refers to any book or manual that is designed to teach any aspects of the software program as well as any book that also emphasizes other objectives, such as teaching concepts. Considering only the books designed to help students learn and apply SPSS, there may be numerous. Many offer students a basic understanding of the mechanics of SPSS, emphasizing data entry and generation of the output for basic descriptive and inferential statistics. Others may offer some theory behind the methods discussed and the authors may recommend their book as an alternative to the traditional introductory statistics textbook. Still others may focus less on mechanics and theory, but more on utilizing 'real' data sets from their own substantive area while at the same time emphasizing interpretation, writing, and reporting results.

Because the textbooks vary on many different factors, it may be useful to review and summarize some of the most popular textbooks. Factors such as level of the audience, complexity of statistical procedures discussed, degree of interpretation of statistics/output, amount of detail discussed on the basic mechanics, accessibility of data files, and student exercises are important considerations. The purpose of this paper is to offer teachers and researchers a brief review of some of the most popular SPSS textbooks that are available today by utilizing evaluation criteria previously discussed in the literature. This review can provide a starting point to explore features of the various SPSS textbooks as well as to consider what book is most appropriate based on their own teaching style. At the same time, this paper continues the tradition in *SERJ* and *SERN* by also providing selected bibliographies for specific topics available to statistics educators (see Holmes, 2002).

2. METHOD

2.1. CRITERIA FOR EVALUATION

The criteria selected for evaluation was developed based on criteria previously selected in the statistics education literature. Huberty and Barton (1990) used coverage, how-to-do, readability, and exercises to evaluate the quality of multivariate statistics text while Cobb (1987) included technical level and exposition, topics covered, and quality of exercises to evaluate 16 introductory statistics textbooks. Harwell, Herrick, Curtis, Mundfrom, and Gold (1996) used variables such as text math and reading level, reading ease, emphasis on theory and computations, writing style, text resource versus learning resource, breath, depth, and overall description of statistical inference, and problem solving. These criteria were utilized to construct and pilot instruments that would provide objective and fair evaluations among students, instructors, and expert evaluators.

Considering the previous research, the criteria used for evaluation of the SPSS textbooks were defined in three major categories: mechanics, content, and classroom/student activities. The mechanics category describes how the books address the 'how to' in terms of SPSS processes (input and output) by providing a summary about how concepts, procedures, and input and output are discussed (i.e., basic details about how to generate output, detailed explanation of interpretation of output). Content includes more discussion about the concepts and procedures, coverage, pedagogy if appropriate, information helpful to determine the appropriate level of audience, and any other useful information. Classroom/student activities include information about the exercises, examples, appendices, and data files that may accompany the text. Other notable information about the textbooks is included in a table in Appendix 1.

2.2. SURVEY TO TEACHERS

Thirty-five teachers and researchers who were members of either the Measurement and Research Methodology Division or the Educational Statistician Special Interest Group of the American Educational Research Association responded to a brief survey about their teaching practices involving the software and software textbook used in their quantitative courses (see Appendix 2 for the questionnaire used in the survey). Two college teachers who did not belong to either group also responded, for a total of n=37 respondents. Approximately 81% reported using SPSS, with others reporting using SAS, Systat, Excel, Statview, LISREL, EQS, AMOS, or no software. Because this paper's primary focus is on the use of SPSS, results reported by teachers who only use SPSS in their teaching will be presented and discussed later.

2.3. SPSS BOOKS REVIEWED

An attempt was made to review the most popular SPSS software textbooks. All but one of the books used by the teachers in this study was reviewed. In addition, other books were selected based on contact made with major publishers for their most popular-selling SPSS books. A total of 17 books were considered.

Table 1. Sample Textbooks and Abbreviations

Textbook	Abbreviation
Carver, R. H., & Nash, J. G. (2000). *Doing data analysis with SPSS 10.0. Pacific Grove.* CA: Duxbury/Thomson Learning.	C&N(2000)
Cronk, B. C. (2002). *How to use SPSS: A step-by-step guide to analysis and interpretation.* (2nd ed.). Los Angeles, CA: Pyrczak.	C(2002)
George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference 11.0 update*. (4th ed.). Boston, MA: Allyn and Bacon.	G&M(2003)
Green, S. B., & Salkind, N. J. (2003). *Using SPSS for windows and macintosh: Analyzing and understanding data.* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.	G&S(2003)
Kendrick, J. R. (2000). *Social statistics: An introduction using SPSS for windows.* Mountain View, CA: Mayfield.	K(2000)
Kirkpatrick, L. A., & Feeney, B. C. (2003). *A simple guide to SPSS for windows for versions 8.0, 9.0, 10.0, & 11.0* (Rev. ed.). Belmont, CA: Wadsworth/Thomson Learning.	K&F(2003)
Norusis, M. J. (2002*). SPSS 11.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.	N(2002)
Pavkov, T. W., & Pierce, K. A. (2003). *Ready, set, go! A student guide to SPSS 11.0 for windows*. New York, NY: McGraw-Hill.	P&P(2003)
Shannon, D. M., & Davenport, M. A. (2001). *Using SPSS to solve statistical problems: A self-instruction guide.* Upper Saddle River, NJ: Prentice-Hall.	S&D(2001)
Sweet, S. A., & Grace-Martin, K. (2003). *Data analysis with SPSS : A first course in applied statistics* (2nd ed.). Boston, MA: Allyn and Bacon.	S&G-M(2003)
Weinberg, S. L., & Abramowitz, S. K. (2002). *Data analysis for the behavioral sciences using SPSS.* New York, NY: Cambridge.	W&A(2002)

As versions 8.0-11.5 are functionally identical, older versions of SPSS (prior to 8.0) were not reviewed. In addition, only the latest edition of a book for the same author or authors was reviewed. Six of the 17 books were excluded due to the previous two reasons; therefore, 11 books were included in the final review. The books and their corresponding abbreviations are presented in Table 1.

All of the books reviewed were alike in many ways. The authors appeared to support and follow the recommendations of the ASA/MAA (Cobb, 1992) by fostering a framework of learning that emphasized conceptual understanding of statistics using SPSS. The majority of the books provided step-by-step illustrations with data editor, screen, and output screen excerpts. The topics typically taught in introductory statistics courses were discussed and contextual examples were followed through from inputting data and defining variable names to generating the output and discussing and interpreting the statistics. In most cases, there were few to very few formulas presented and student exercises and practice data files accompanied the chapters.

A review is presented below by considering the evaluation criteria described earlier. In order to facilitate the reader, the books have been 'clustered' based on their similarities, and distinctions are made within each cluster across each category. For example, K&F(2003), P&P(2003), and C(2002) all appear to offer the very basics in terms of how-to and interpretation therefore, these books were clustered together. Therefore, the following books were clustered: Cluster 1 is K&F(2003), P&P(2003), and C(2002); Cluster 2 is S&G-M(2003) and W&A(2002), Cluster 3 is C&N(2000), K(2000), N(2002), and S&D(2001); and Cluster 4 is G&M(2003) and G&S(2003).

3. RESULTS AND DISCUSSION

3.1. MECHANICS

Cluster 1

For an SPSS manual that teaches the very basics in terms of SPSS mechanics and explanations of output and statistics concepts, K&F(2003) or P&P(2003) could be considered. Both books provide just enough information in order to explain the procedures. K&F(2003) also discuss syntax; however these sections can easily be omitted. The book by C(2002) also provides very basic details of SPSS mechanics but the discussion and explanation of the statistical procedures and concepts are slightly more detailed (i.e., definitions of concepts, assumptions are also discussed).

Cluster 2

The S&G-M(2003) and W&A(2002) books were alike in that they both offered the basics in terms of mechanics or 'how-to' but both authors provide more detailed explanation of the procedures, output, and concepts. Neither book focused on the 'how-to' but more on teaching concepts and interpreting output. According to W&A(2002), their book could perform as a primary statistics textbook. Their book is comprehensive in terms of teaching statistics concepts whereas the S&G-M(2003) book provides more of a balance of teaching and data analysis by briefly discussing definitions of concepts and targeting specific analyses used by social scientists.

Cluster 3

The C&N(2000) book provides straightforward and somewhat detailed discussions of statistical procedures, explanations and discussion of concepts, mechanics, and interpretation of output. Also comprehensive in SPSS mechanics are K(2000), N(2002), and S&D(2001). K(2000) is also comprehensive in the teaching of statistics concepts and procedures as it is recommended to be used as a primary statistics textbook. S&D(2001) offer very detailed explanations of concepts and interpretation of output and the authors contend their book should be used as a supplement. The N(2002) book focuses on SPSS and data analysis, as reported by the author and could be used as a supplement or primary text for an introductory course in data analysis. An introduction to statistics concepts and topics are succinctly discussed with heavy emphasis on generating, discussing and interpreting output.

Cluster 4

Finally, G&M(2003) and G&S(2003) offer the most advanced and comprehensive books in terms of both mechanics and statistical procedures discussed. G &M(2003) provides brief to moderate details of definitions and explanations of statistics concepts and the output is briefly discussed. The same is true of G&S(2003) but more emphasis is given to the interpretation of the output and writing results in APA format.

3.2. CONTENT

Cluster 1

All of the authors reported that their books were appropriate for students studying at the introductory level, or higher. None of the books in cluster 1 would be adequate as a stand-alone statistics textbook; students using these books would need to have some understanding of statistics. For each chapter in K&F(2003), one example taken from a psychology perspective begins the chapter, a discussion on how to generate the output using both the point-and-click and syntax methods are described, and a discussion of the output and a very brief interpretation is provided. The authors do not adequately address important assumptions (i.e., homogeneity of variance for independent samples t-test), indicating that certain tests are "probably not of interest to most readers" (2003, p. 34).

C(2002) chapters begin with a brief description of the topic and assumptions, followed by a discussion of variables, an illustration of how to generate the output, and a section on how to read and interpret the output. Although C(2002) includes a separate section on assumptions, they are not exhaustive. For example, the assumption of equal variances in the population is not mentioned for the one-way ANOVA design (2002, p. 63) and is also not addressed in the output. Nonparametric methods, reliability analyses, and MANOVA are additional topics that are discussed in C(2002).

Finally, the authors P&P(2003) meet their objectives by publishing a supplemental book that includes brief explanation of statistics concepts yet also accurately addresses important assumptions using SPSS. For each chapter in P&P(2003), the research questions and the design are first discussed, the statistical procedure is chosen, and the results are briefly interpreted and summarized. As far as more advanced topics, the repeated measures ANOVA design is also discussed.

Cluster 2

For cluster 2, the W&A(2002) book is adequate as a stand-alone statistics textbook. In addition to providing detailed discussions for concepts and procedures, the book is written to facilitate student pedagogical objectives by reintroducing concepts for reinforcement. For example, although the authors introduce and discuss descriptive statistics early on, they continue to incorporate many of these concepts and the use and misuse of graphical displays in other topics throughout their textbook. In addition to the topics normally covered at the beginning level, data transformations, nonparametric methods, and effect size indicators are also discussed. This book contains more formulas and uses short examples within the chapters to illustrate the formulas immediately followed by SPPS commands and outputs to confirm the hand calculations.

S&G-M(2003) provides accurate but brief discussions of concepts and procedures. Although details of some assumptions are discussed, they are also not exhaustive and not addressed with the SPSS output (see pp. 119-123). Chapters discussing how to write a research report and potential research projects that may be considered using the data sets on disk are included. Logistic regression and multivariate logistic regression are two of the more advanced topics discussed.

Cluster 3

C&N(2000) presents some detailed discussions of concepts and procedures including discussions of assumptions. Although there are presentations devoted to topics such as nonlinear models, forecasting techniques, nonparametric tests, and quality control procedures, some explanations of other important concepts, such as investigating interaction effects were limited (i.e., using descriptive charts see pp. 154-159). Exercises for simulation activities are accessible for probability, normal distribution, sampling distributions, confidence intervals, and the one-sample t-test.

S&D(2001) provide a more detailed explanation of concepts, procedures and assumptions, however the authors provide a similar discussion of interaction effects as C&N(2000) (see pp. 246-247). Reliability analyses and regression with categorical predictors are additional topics that are discussed.

K(2000) presents accurate discussions of concepts and statistical procedures in the manner in which a primary statistics textbook would. In terms of pedagogy, there are skill practices within the chapters where students can evaluate their understanding in a formative manner. The book is very comprehensive in both statistics and providing SPSS information, which may explain why only introductory material (up through one-way ANOVA) for a one-semester course is presented.

Finally, N(2002) begins each chapter with a list of questions, provides a concise but somewhat detailed introduction to statistics concepts, procedures, and assumptions (through multiple regression), and concludes with detailed steps on how to generate and interpret the output. A limited discussion of interaction effects is also provided using a plot of the means (see pp. 330-334). There are chapters devoted to plotting data, nonparametric tests, analyzing residuals, and multiple regression diagnostics.

Cluster 4

G&M(2003) authors report that 95% of the analyses that are conducted in the sciences or business can be accomplished using their book. The first 16 chapters cover introductory statistics topics while the last twelve chapters utilize the most advanced modules of SPSS. Some of these topics include 3-way ANOVA, reliability analyses, multidimensional scaling, factor analysis, cluster analysis, discriminant analysis, repeated measures MANOVA, logistic regression, hierarchical loglinear models, and general loglinear models. Because many topics are discussed, there is brief to moderately detailed explanations of concepts and procedures. For example, specific reports about assumptions are not discussed in the introduction of a chapter but important assumptions tested by SPSS are addressed when referring to the output (see p.140).

G&S(2003) is similar to G&M(2003) in their explanation of concepts and coverage of advanced topics such as factor analysis, discriminant analysis, and reliability analyses. However a more well-rounded presentation of the topics is provided, including discussions of assumptions, definitions of research questions, a description of the research design, measures of effect sizes, discussion of results in APA format as well as alternative ways to analyze the data. Macintosh users for version 10.0 can also utilize the book, for the most part because any features that are not available on Macintosh are pointed out by the authors.

3.3. CLASSROOM/STUDENT ACTIVITIES

Cluster 1

Both C(2002) and P&P(2003) have practice or student exercises but neither provides a list of references. The exercises for C(2002) allow students to manually enter their own data using short data sets provided in the chapters or they can refer to the appendix for longer data sets, which also have to be manually entered. In one case, students were referred to a data set in the SPSS directory program files (CARS.SAV) which was not found in SPSS version 11.5 for Windows (see p. 18). There is an instructor's key for all exercises and an appendix which includes a glossary of terms as well as a decision tree to help students learn how to select the appropriate inferential statistical test.

P&P(2003) use the GSS93 data file (General Social Survey) provided in the SPSS program directory to illustrate their statistical procedures. They also utilize 'general' exercises, (i.e., State your research question based on two variables you or your instructor chooses) that should encourage students to undertake independent computer assignments, according to the authors. The appendix includes how to enter data using files and programs other than SPSS (i.e., text editor, spreadsheet) and four additional data sets are available for students to use for practice and must be entered manually.

Finally, K&F(2003) do not have exercises or references; however the appendices describe other features of the program not discussed in detail in the book such as saving and retrieving data, output, and syntax files, and data transformations (creating and computing new variables, recoding existing variables, selecting cases).

Cluster 2

S&G-M(2003) have no student exercises but provide a data disk that includes 2 data sets, the General Social Survey (1998) as well as a data set that contains variables about the social behavior in the United States. The appendix includes a description of the data sets at the end of each chapter, a summary, key terms, and a list of references.

W&A(2002) includes ample student exercises and a disk accompanies the book that consists of several real data sets. There is one student exercise that uses an SPSS macro syntax file to generate a sampling distribution of means that was not included on the disk (SAMPDIS.sps) provided for this reviewer (All of the other SPSS data files were available). The appendix provides a description of the data sets, statistical tables (i.e., z, t), references, and solutions to all exercises.

Cluster 3

C&N(2000) report that they use mostly real data sets for illustration of procedures and exercises. The files are expected to be in the SPSS program file directory or need to be downloaded from the World Wide Web (WWW). In some cases, either the data was not in the directory or the data was not available on the Web (see p. 274 for AIRLINE.SAV – 'page not found'). Also, in one case, the data to be downloaded is in 'html' format (LONDON1.SAV), a file type that was not addressed in the appendix for use in SPSS in their book. There could be differences in versions in terms of SPSS program files; C&N(2000) considers version 10.0 (this reviewer is using version 11.5). The appendix includes a detailed description of SPSS files, a discussion of how SPSS handles and supports other data files, and an introduction to users working with SPSS 9.0.

K(2000) also reports using version 9.0 and the text is accompanied by an instructor's manual and a data disk (These were not received with this text for the reviewer). The author reports that the disk contains the General Social Survey (1996) data and an instructor's manual that contains solutions to the even-numbered problems. The appendix lists tables, advanced features of SPSS, answers to odd-numbered problems, references, and a glossary.

S&D(2001) uses real data sets provided on a disk collected from students enrolled in an introductory statistics course related to student admission and computer and statistics attitudes. The student exercises utilize these data and all solutions are appended. The appendix also includes a description of the data sets and information about how to use syntax commands.

Finally, N(2002) includes a data disk of real data from the General Social Survey, Impact of the Internet on Library Use study, Stanford Institute for the Quantitative Study of Society, Chicago marathon, opinions of the criminal justice system, and an ABC survey of manners. All of the files are used in the chapters for exercises and extra data analysis problems at the end of each chapter. The appendix includes information on how to obtain high resolution charts, data transformations and case selections, statistical tables, description of the data files, answers to selected exercises, and a list of references. In addition, an instructor's guide accompanies the book which includes teaching hints for each chapter as well as answers to all of the student exercises.

Cluster 4

Data files used to explain procedures in the chapters and student exercises have to be downloaded for G&M(2003). At this site, the files can be downloaded one at a time or in zip format. Also, solutions to selected exercises using the adobe acrobat reader can also be obtained. Information about the instructor's manual is also available and is password protected. Information on the data files, a glossary, and a listing of references is also included with the book.

G&S(2003) provides a disk for illustrating procedures and student exercises. The appendix includes descriptions of data files, solutions to selected exercises, references, and a discussion of the different methods used to control Type I errors for multiple testing.

3.4. COLLECTIVE VIEWPOINT

A survey was used to gather data about the extent to which teachers use SPSS and SPSS software books in their teaching. These teachers reported using SPSS from their introductory courses in research methods, statistics, sampling and survey design to advanced quantitative courses such as multivariate statistics. Also, teachers reported using between 1-50% of their classroom time (some courses used labs devoted exclusively to the teaching of SPSS) related to SPSS.

Of the 81% who reported using SPSS, about half (40%) used software books and all but one reported using these books as a supplementary resource (this one response was missing). The other half reported using their own SPSS handouts. The most popular book was Green & Salkind (2003) with 40% reporting using this book while 13% each indicated using Kirkpatrick & Feeney (2003) and Norusis (2002). One teacher each (6%) reported using George & Mallery (2003) and Pavkov & Pierce (2003) while 13% of the teachers reported using some other book or various books from semester to semester.

There were many reasons why teachers reported selecting their particular book. Cost, readability, detailed description and explanation of analyses and concepts, emphasis on calculation, use, and interpretation of effect sizes, good use of visuals, user-friendly, availability and correctness of concepts, a focus on primary statistical procedures, statistical theory is presented, use of the book as a good reference, easy to understand, clarity, annotated output with sample reports, facilitates conceptual understanding, examples are taken from the behavioral sciences, real data, many examples, ease of use by both teacher and student, minimal formulas, APA format of results, and not too theoretical were the answers provided. In fact, 40% of the teachers indicated that they use their book to help reinforce statistics concepts, 13% reported using their manual to teach concepts, and 40% indicated that they do not rely on their book to teach or reinforce. Also, the majority of the teachers reported using their book to teach mechanics (53%), 33% said they did not use their book to teach mechanics, and two teachers (13%) reported using their books as references only. Finally, the teachers reported using the data files, exercises at the end of the chapters, syntax, and APA formatting either in class or that these features are used by their students.

3.5. COMPLEMENTARY INFORMATION

The table in Appendix 1 provides some summary data as well as other information teachers might find useful. The intended audience level (introductory, intermediate, and advanced), syntax (whether syntax language is discussed), whether a disk is included with the book, information about exercises and solutions, number of pages, and the versions of SPSS needed for the books are reported.

4. FINAL THOUGHTS

Interestingly enough, half of the teachers who reported using SPSS did not require or recommend a software book. There was no question to address why these teachers preferred not to use a book however some researchers provided reasons. One researcher stated that there are often more misconceptions and errors in software books than a primary textbook. Two other researchers commented that their primary textbooks provided enough SPSS information for their purposes. Most other researchers stated that they used their own handouts; one researcher reported that all SPSS (and SAS) documentations were online at his/her university. There could be other reasons why half of the teachers use their own handouts, two of which might be that teachers may not value or have the time to fully utilize SPSS and quick handouts will suffice, or there could be other courses that teach data analysis procedures (i.e., Computer Use in Educational Research) available to their students.

5. LIMITATIONS

The following limitations should be mentioned for this research study and review. First, neither the teachers nor the books were randomly sampled. Even though the books were not randomly selected, I do believe they are in fact representative of books used in many methods courses. Secondly, the sample was not only convenient but also small. Consequently, generalizations to all teachers who use SPSS are quite limited. Most importantly, the books were evaluated by one researcher who is a teacher and a researcher in quantitative methods. Because of the subjective nature of the review, attempts were made to lower bias by using evaluation criteria selected from the statistical education literature, including ideas from other teachers and researchers, and reporting page numbers in cases where more subjective comments were included. However, the review does consist of comments from a single reviewer; therefore, readers should not over-generalize or draw very simple conclusions. Instead, this information is presented to make readers aware of the SPSS software books available today; my hope is that teachers will choose several books that seem appealing and ultimately make their own assessments. Any oversight related to any of the features or criteria of the software books was unintentional.

Finally, the evaluation of statistics textbooks has important consequences. An evaluation of a textbook should provide objective information to readers and may often help one decide whether it is ‘suitable’ for their purposes. Although this review focused on evaluation criteria selected from the statistics education literature, there was little guidance as to how objective evaluations should be conducted. Subjective evaluations undoubtedly have merit however there appears to be a lack of research in the literature that suggests what criteria to consider and how to use it. Harwell et. al (1996) stated “This lack of guidance may be at least partly responsible for the fact that published evaluations of statistics texts almost invariably employ evaluation criteria that lack any theory-based rationale” (p. 4). Future research might focus on the continued development of empirically-based criteria for the evaluation of statistics textbooks as well as suggestions on how to (best) objectively utilize the criteria. Including input from teachers, researchers, and students from all levels of education both nationally and internationally are also invaluable and can provide an important perspective in the development of this criteria. As the teaching of statistics continues to become more popular across disciplines and age levels, the selection of a statistics textbook will require thoughtful consideration. For this reason, evaluations of this sort will continue to be both valuable and useful to those of us involved in the wonderful world of teaching.

REFERENCES

Carver, R. H., & Nash, J. G. (2000). *Doing data analysis with SPSS 10.0*. Pacific Grove, CA: Duxbury/Thomson Learning.

Cobb, G. W. (1987). Introductory textbooks: A framework for evaluation. *Journal of the American Statistical Association*, *82*, 321-339.

Cobb, G. (1992). Teaching statistics. In L.A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3-43; MAA Notes No. 22). Washington, DC: Mathematical Association of America.

Cronk, B. C. (2002). *How to use SPSS : A step-by-step guide to analysis and interpretation* (2nd ed.). Los Angeles, CA: Pyrczak.

Garfield, J. (1997). Comment on new pedagogy and new content: The case of statistics by D. Moore. *International Statistical Review, 65*(2),137-141.

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference 11.0 update* (4th ed.). Boston, MA: Allyn and Bacon.

Giesbrecht, N. (1996). *Strategies for developing and delivering effective introductory-level statistics and methodology courses. Alberta, Canada*. (ERIC Document Reproduction Service No. ED393 668).

Goodman, T. A. (1986). Using the microcomputer to teach statistics. *Mathematics Teacher*, *79*, 210-215.

Gratz, Z. S., Volpe, G. D., & Kind, B. M. (1993). *Attitudes and achievement in introductory psychological statistics classes: Traditional versus computer supported instruction. Ellenville, NY.* (ERIC Document Reproduction Service No. 365 405)

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for windows and macintosh: Analyzing and understanding data* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Harwell, M. R., Herrick, M. L., Curtis, D., Mundfrom, D., & Gold, K. (1996). Evaluating statistics texts used in education. *Journal of Educational and Behavioral Statistics, 21*(1), 3-34.

Hoerl, R., Hahn, G., & Doganaksoy, N. (1997). Comment on new pedagogy and new content: The case of statistics by d. moore. *International Statistical Review*, *65*(2), 147-153.

Holmes, P. (2002). Some basic references for the teaching of undergraduate statistics. *Statistics Education Research Journal, 1*(2), 49-53.

Huberty, C. J., & Barton, R. M. (1990). Applied multivariate statistics textbooks [book review]. *Applied Psychological Measurement*, *14*, 95-101.

Kendrick, J. R. (2000). *Social statistics: An introduction using SPSS for windows*. Mountain View, CA: Mayfield.

Kirkpatrick, L. A., & Feeney, B. C. (2003). *A simple guide to SPSS for windows for versions 8.0, 9.0, 10.0, & 11.*0 (Rev. ed.). Belmont, CA: Wadsworth/Thomson Learning.

Lomax, R. G. (2001). *An introduction to statistical concepts for education and behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2),123-137.

Moore, D. S. (2000). *The basic practice of statistics* (2nd ed.). New York: W.H. Freeman and Company.

Norusis, M. J. (2002*). SPSS 11.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.

Pavkov, T. W., & Pierce, K. A. (2003). *Ready, set, go! A student guide to SPSS 11.0 for windows.* New York, NY: McGraw-Hill.

Scheaffer, R. L. (1997). Comment on new pedagogy and new content: The case of statistics by D. Moore. *International Statistical Review, 65*(2), 156-158.

Shannon, D. M., & Davenport, M. A. (2001). *Using SPSS to solve statistical problems: A self-instruction guide*. Upper Saddle River, NJ: Prentice-Hall.

Sweet, S. A., & Grace-Martin, K. (2003). *Data analysis with SPSS : A first course in applied statistics* (2nd ed.). Boston, MA: Allyn and Bacon.

Velleman, P. F., & Moore, D. S. (1996). Multimedia for teaching statistics: Promises and pitfalls. *The American Statistician*, *50*, 217-225.

Weinberg, S. L., & Abramowitz, S. K. (2002). *Data analysis for the behavioral sciences using SPSS.* New York, NY: Cambridge.

JAMIE D. MILLS
University of Alabama
Department of Educational Studies
P. O. Box 870231
Tuscaloosa, Alabama 35487-0231
USA

APPENDIX 1: LIMITED SUMMARY AND ANCILLARY INFORMATION

Author(s)	Audience	Syntax	Disk	Number of pages including appendix	Version(s) *	Exercises/Solutions
Carver & Nash	Introductory Intermediate Advanced	No	No, download or use SPSS program files	326	10.0 Intro to 9.0 in appendix	Selected exercise solutions only available to instructor
Cronk	Introductory Intermediate	No	No, manual input or SPSS program files	114	10.0 11.0	All exercise solutions only available to instructor
George & Mallery	Introductory Intermediate Advanced	No	No, download data files	386, no appendix	11.0	Selected exercise solutions must be downloaded
Green & Salkind	Introductory Intermediate Advanced	Yes	Yes	420	11.0 for Windows; 10.0 for Macintosh	Selected exercise solutions in appendix
Kendrick	Introductory Intermediate?	No	?, Author reports disk is included	630	9.0	All solutions within chapters; Selected in appendix --All available to instructor
Kirkpatrick & Feeney	Introductory Intermediate	Yes	No	118	8.0, 9.0, 10.0 & 11.0	No exercises
Norusis	Introductory Intermediate?	No	Yes	625	11.0	Selected exercise solutions in appendix – all available to instructor. No solutions for data analysis exercises
Pavkov & Pierce	Introductory	No	No, use SPSS program files	89	11.0	No solutions to exercises
Shannon & Davenport	Introductory Intermediate	Yes	Yes	369	9.0	All exercise solutions are appended
Sweet & Grace-Martin	Introductory Intermediate	Yes	Yes	231	11.0	No solutions to exercises
Weinberg & Abramowitz	Introductory Intermediate	No	Yes	590	10.0	All exercise solutions are provided

*Many authors have books for earlier versions. Please see the publisher for more information.

APPENDIX 2: SURVEY QUESTIONNAIRE

Dear Colleague,

I am interested in gathering some information about your teaching practices in your statistics, research, or other related methods course. Please take a few moments to answer the 10 questions below about the software textbook that you are currently using either as a supplement or primary textbook for your course. In this case, 'software textbook' refers to any book or manual that is designed to teach any aspects of the software program as well as any book that also emphasizes other objectives, such as teaching concepts.

1. What software program do you use for your course? (i.e., SPSS, SAS, Minitab, etc.)
2. For what course do you recommend or require an accompanying textbook for the software? (i.e., Statistics, Research Methods, Survey Methods, etc.)
3. What is the name of this software textbook that you recommend or require?
4. Why did you choose this particular software textbook? (i.e., cost, detailed explanation of concepts, etc.). Please explain.
5. Do you use this software textbook as a supplement to a traditional textbook for your course or is it used exclusively as a primary textbook? Please explain.
6. Do you rely on your software textbook to teach or reinforce statistics concepts that you teach in class? Please explain.
7. Do you rely on your software textbook to teach students the mechanics of the program? (i.e., how to input, print, save data and output). Please explain.
8. How much classroom time (0-100%) do you devote to teaching the software? Please explain.
9. What are some of the features of your software textbook that your students use or that you use in class that you like most? (i.e., data files on accompanying disk, exercises at end of chapter, syntax provided, data on internet, etc). Please explain.
10. Given your comments above, what level of student (undergraduate and/or graduate) and class (introductory, intermediate, and/or advanced) do you teach?

INFORMATION ON PAST IASE CONFERENCES

1. IASE SATELLITE CONFERENCE ON STATISTICS EDUCATION AND THE INTERNET
Max Planck Institute for Human Development, Berlin, Germany. August 11-12, 2003

REPORT BY LARRY WELDON AND JOACHIM ENGEL

The conference was held August 11-12, 2003, just before the 54th Biennial Meeting of the International Statistical Institute. The venue provided by the Max Planck Institute for Human Development was an ideal facility for the conference. The organizers thank the directors of the Institute, Professor Jürgen Baumert and Professor Gerd Gigerenzer, for their generous support.

The conference was organized by the International Association for Statistical Education in cooperation with the International Statistical Institute, Stochastics Section of the German Society for Mathematics Education, German Mathematical Association (special interest group on Probability and Statistics), German Statistical Society and Max-Planck Institute for Human Development.

Prof. Gerd Gigerenzer of the Max Planck Institute gave a short opening address. Seventeen invited speakers presented talks relating to our theme “Statistics Education and the Internet” – these talks were presented to the plenary sessions of 60 registrants. In addition, a two hour time slot was reserved for fifteen poster sessions also directed to this same topic. Twenty two registrants were welcomed as new IASE members.

The invited speakers papers were refereed when they were submitted early enough. Fifteen of the invited speakers completed the refereeing process in time for the conference. One invited paper was submitted by title since the author was unable to be present. A CD of the proceedings was produced and distributed to all registrants. It contained all the invited papers, abstracts of the contributed poster sessions and a list of registrants. The proceedings are available from the conference Web page at http://www.ph-ludwigsburg.de/iase/.

The **Scientific Program Committee** consisted of: Larry Weldon, Chair (Canada), Carmen Batanero (Spain), Joachim Engel (Germany), Brian Phillips (Australia) and Gilberte Schuyten (Belgium) and the **Local Arrangements Committee** consisted of Joachim Engel, Chair; Rolf Biehler, Laura Martignon and Markus Vogel.

1.1. INVITED PAPERS

Gabriela Belli. *The many faces of statistical education via internet resources*.

Rolf Biehler. *Interrelated learning and working environments for supporting the use of computer tools in introductory courses*.

Andrej Blejec. *Teaching statistics by using simulations on the Internet.*

Lea Bregar. *Teaching economic statistics in the Internet era.*

Stephen R. Clarke. *Raising interest in statistics through sporting predictions on the Internet.*

Paul Darius, E. Schrevens, H. Van der Knaap, K. Portier, G. Massonnet, L. Lievens, L. Duchateau, & O. Thas. *Using Web-based tools for teaching statistical concepts and experimentation skills*.

Tim Dunne. *Exploring threshold concepts in basic statistics using the Internet*.

Andreas Eichler. *Madin – Teaching School Mathematics with the Web.*

Juan D. Godino, Francisco Ruiz, Rafael Roa, Juan L. Pareja, & Angel M. Recio. *A didactical analysis of two internet interactive tools for teaching statistics in schools.*

Gökhan Aydinli, Wolfgang Haerdle, & Bernd Roenz. *E-learning/teaching of statistics: Student and teacher views.*

Erhard Cramer, Katharina Cramer, Petra Janzing, Udo Kamps, & Claudia Pahl. *Emilea-stat: A Web-based learning environment in applied statistics with a focus on learning and teaching in secondary schools.*

Kay Lipson, Glenda Francis, & Sue Kokonis. *Investigation of students' learning experiences with a Web-based computer simulation.*

Irena Ograjensek, & Mojca Bavda Kveder. *Student acceptance of ITT-supported teaching and internal course administration: case of business statistics.*

Brian Phillips. *Overview of Internet resources for statistics education*.

Martin Podehl. *Statistics in the classroom - societal issues*.

Jean-Claude Régnier. *Statistical education and e-learning*.

Rachel Cunliffe, Matt Regan, & Chris Wild. *Flexible learning and large numbers (a case study).*

Joseph M. Wisenbaker. *Extending the journey toward a virtual introductory statistics course*.

1.2. CONTRIBUTED PAPERS

Felicien Accrombessy. *Information and communication technology and the development of statistics teaching in Benin: Advantages and inconveniencies*.

Kazlauskiene Ausra. *Statistic's element ability subtest.*

E. L. Sanjuan, M. Isabel Parra, & I. I. Castro. *Multimedia tools for the teaching of probability and statistics.*

Joachim Engel. *Stochastic modeling and statistical thinking in technology supported environment*.

Kamel Esseghairi. *Contribution of private IT academic institutions to improve and disseminate the teaching and the use of applied statistics.*

Henryk Kolacz. *Statistics lab exams on-line system.*

Gerhard König. *Electronic access to literature in theory and practice of statistics education*.

Kang Sup Lee. *Changing statistics education models of mathematics teacher's.*

M. Isabel Parra, & Francisco Cuadros. *Statistics, probability and chaos.*

Gilbert Saporta, & Marc Bourdeau. *St@tNet: an assessment and new developments.*

Ana Jesús López Menéndez. *Economic data analysis with ADE.*

Maria A. Pannone, & Judit Jasso. *Statistics and Internet in Italy: CIRDIS Website and CIRDIS teaching materials.*

Maria Lucia Marçal Mazza Sundefeld. *A successful experience of database employment*.

Larry Weldon. *Replacing the statistics text with reader excerpts and timely Internet notes.*

2. IASE ACTIVITIES AT THE ISI-54TH SESSION
Berlin, Germany, August 10-20, 2003

IASE organised a wide and varied list of topics at the 54th Session of the International Statistical Institute for Invited Paper Meetings, both alone and in conjunction with other ISI Sections and Committees and guest societies. The Chair of IASE Programme was Gilberte Schuyten, gilberte.schuyten@rug.ac.be. Below we are listing the Invited Paper Meetings with organisers, papers and discussants. For further information see: http://www.isi-2003.de/.

2.1. IASE INVITED PAPER MEETINGS, IPM

IPM44. *Teaching probability with a modelling approach*. Organizer: Michel Henry; Discussant: Lionel Pereira Mendoza.

- Juan D. Godino, M. Jesús Cañizares, & Carmen Díaz. *Teaching probability to pre-service primary school teachers through simulation.*
- Ernesto Sanchez, & Gabriel Yañez. *Computational simulation and conditional probability problem solving.*
- Bernard Parzysz. *From frequency to probability: Some questions posed by the new French senior high school curricula.*
- Talma Leviatan. *Conceptual, computational and didactical aspects of teachers' training in probability & statistics.*

IPM45. *Statistics training for consultant or collaborator*. Organizer: Gabriella Belli; Discussant: Herb Ware.

- John Harraway, & Richard J. Barker. *The use of statistics in the workplace: A survey of research graduates in diverse disciplines.*
- Flavia Jollife. *Communication, collaboration, and consulting.*
- Elisabeth Svensson. *Statistical consulting: A matter of breaking tradition in applied research*.
- Lisbeth Cordani. *Consultancy can be of service to teaching, research, and extension activities.*

IPM46. *Research in statistics education and international cooperation*. Organizer: Lisbeth Cordani; Discussant: Carmen Batanero.

- Jorge Luis Romeu. *Juarez Lincoln Marti Project: An example of international co-operation in statistics education and research*.
- Shrikant I. Bangdiwala. *International co-operation in research: An opportunity to educate collaborators.*
- Clarice G. B. Demétrio. *International co-operation: A Brazilian example of academic exchange*.

IPM47. *Mathematics teachers teaching statistics*. Organizer: Susan Starkings; Discussant: Andrej Blejec.

- Ann-Lee Wang, & Song-How Kon. *Should simple Markov processes be taught by mathematics teachers?*
- James Nicholson, & Catherine Dranton. *Mathematics teachers teaching statistics: What are the challenges for the classroom teacher?*
- Henrik Dahl. *Mathematicians have problems teaching model assumptions in statistics.*

IPM48. *Statistics education for media reports.* Organizer: Maxine Pfannkuch; Discussants: Jane Watson & Chris Wild.

- Iddo Gal. *Functional demands of statistical literacy: ability to read press releases from statistical agencies.*
- W. Martin Podehl. *Statistics for journalists*
- J. Laurie Snell. *A course called Chance*.

IPM49. *Teaching and learning approaches aimed at developing statistical reasoning, thinking or literacy.* Organisers: Joan Garfield & Dani Ben-Zvi; Discussants: Iddo Gal.

- Jane Watson. *Statistical literacy at the school level: what should students know and do?*
- Maxine Pfannkuch, & Chris Wild. *Statistical thinking: how can we develop it?*

- Efi Paparistodemou. *Asymetric fairness and unfairness: Reinventing distribution with a computer game.*

IPM50. *Statistics Teaching in the Internet Age*. Organiser: Wolfgang Härdle; Discussants: Jung Jin Lee &Adonis Yatchew.

- Yuichi Mori, Yoshiro Yamamoto, Hiroshi Yadohisa, & Tatsuki Inoue. *Data-oriented learning systems of statistics based on analysis scenario/story (DoEStat)*.
- P. Darius, H. van der Knaap, E. Schrevens, K. Portier, G. Massonnet, L. Lievens, & S. Dufresne. *Virtual experiments and their role in teaching design and analysis of experiments.*
- E. Cramer, W. Härdle, U. Kamps, & R. Witzel. *E-stat: view, methods, applications.*

2.2. JOINT IASE WITH OTHER ISI SECTIONS/ COMMITTEES AND GUEST SOCIETIES

IPM68. *Education and assessment of literacy, numeracy and other life skills*. (Proposed by ISI.) Organiser: Denise A. Lievesley; Discussant: Nancy Gordon.

- Siobhan Carey. *Respondent variability in their approach to literacy surveys – some cross national comparisons.*
- Juan Enrique Froemel. *Recent assessments of numeracy and literacy and their proxies, in Latin America: some highlights.*
- Benedicte Terryn. *Measuring literacy in developing countries from an international perspective.*

IPM69. *Impact of developments in information systems on statistics education.* (Joint session IASE and IASC.) Organizers: Annie Morin & Albert Prat; Discussant: Yonggo Lee.

- John Maindonald. *The role of models in predictive validation (Statistics for budding data miners).*
- Jean-Hugues Chauchat. *Teaching statistical inference using many samples from a real large dataset.*
- Gilbert Ritschard. *Testing hypotheses with induction trees.*

IPM70. *Teaching Biostatistics.* (Joint IASE/International Biometrics Society.) Organisers: E. Svensson & Els Goetghebeur; Discussant: Jane L. Hutton.

- Jorge Calderón Guayquil. *Biostatistics teaching in an aquamarine setting: learning by case studies.*
- Masashi Goto. *Experiences in teaching biostatistics in the pharmaceutical industry and university.*
- Beorah Nolan. *Statlabs. Mathematical statistics through applications.*

IPM71. *Educational implications of statistical method and modelling developments developments in psychometry: Is tradition stronger than statistical relevance?* (Joint IASE/ European Mathematical Psychology Group.) Organizers: Helena Bacelar & Francesca Cristante; Discussants: James Townsend & Gilbert Saporta.

- Luc Delbeke. *The psychology of mathematics and statistics for psychologists.*
- Luca Stefanutti, & Francesca Cristante. *Empirical validation of a knowledge structure for assessment and learning of psychometrics at university level.*
- Hans-Christof Micko. *Statistical data: (empirical) facts or (theoretical) fictions?*
- Fernando Costa Nicolau, & Helena Bacelar-Nicolau. *Teaching and learning hierarchical clustering probabilistic models for categorical data.*

3. USING THE HISTORY OF STATISTICS TO IMPROVE THE TEACHING OF STATISTICS. SESSION AT THE JOIN STATISTICAL MEETINGS

San Francisco, California, USA, August 7, 2003

REPORT BY CAROL JOYCE BLUMBERG

IASE had its first invited session as an outside society on August 7, 2003 during the Joint Statistical Meetings in San Francisco, California, USA. The topic of the session was "Using the History of Statistics to Improve the Teaching of Statistics". The co-sponsors of the session were the ASA (American Statistical Association) Section on Statistical Education, the ISI (International Statistical Institute), and the ASA Section on Teaching of Statistics in the Health Sciences. The session included three paper presentations and remarks by Jeff Witmer (Oberlin College, USA) and Jay Devore (California Polytechnic State University, USA) as the discussants. The organizer and chair of the session was Carol Joyce Blumberg (Winona State University, USA). Summaries of the three presentations are given below. The complete papers from the talks and some other materials (including a bibliography of the major books written about the history of statistics) are at http://course1.winona.edu/cblumberg/jsm2003.htm.

3.1. PRESENTATIONS

Herbert A. David (Iowa State University, USA). *The history of statistics in the classroom*.

Herb David strongly believes that statistics classes at any level can be enlivened by highlighting colorful contributors to our field. Outlines of some of their research should be supplemented by historical, biographical, and anecdotal material. Laplace (1749-1827) provides a good example. Famous as a theoretical astronomer and mathematician, he is now probably best known to statisticians as originator of the central limit theorem. But apparently independently of the publication of Bayes's theorem ten years earlier, he proposed that (in modern language) posterior density $\propto$ likelihood. Thus maximizing the posterior means maximizing the likelihood, an idea that had many followers including Gauss. Laplace's book *Théorie analytique des probabilités*, was extremely influential throughout the 19th century. Laplace lived in turbulent times in France. He examined the young Napoleon for the officer exam, but lasted only six weeks in Napoleon's cabinet for bringing "the spirit of the infinitesimal into administration." However, he became Chancellor of the Senate and later was raised to Marquis by Louis XVIII. In the talk similar introductions were also given (but in less detail) about Gauss and R. A. Fisher. The talk concluded with personal comments about Fisher.

David Bellhouse (University of Western Ontario, Canada). *Statistical ideas in the classroom – Lessons from history*.

Almost any introductory statistics textbook is a compendium of the history of probability and statistics since the Middle Ages. Almost always the methods and techniques are given without the historical references. Instead, the focus is on "relevant" applications of the material presented. Generally this is a good thing. Many students are not interested in historical examples, which by their very nature are outdated - they want something more "relevant". The down side to this approach is that the textbook often becomes a cookbook with applied exercises that help you learn how to cook. When examining how the history of probability and statistics can be useful in the classroom, it is first useful to examine the styles in which the history of the subjects is written. These styles may be divided into internalist (those working within the area) and externalist (those outside the area) approaches. It is natural for teachers of probability and statistics to follow an internalist approach for classroom discussion. In order to discover what principles apply in transferring the lessons of history to the classroom, the work of William Sealy Gosset (Student) was discussed as a case study. What followed from this case study is that the most important historical lesson to convey is the motivation for an individual's work. This lesson was illustrated further in discussions of the solution to the problem of points or division of stakes and of the Fisher-Neyman dispute over their approaches to statistical inference.

Fred L. Bookstein (University of Michigan, USA). *Learning from the coercive power of numerical evidence: Three classical examples*.

This talk reviewed the way in which three all-time classics of science muster quantitative evidence in a way that leave their readers no choice except to agree. The sources for the talk were John Snow (1855) on the mode of communication of cholera, Jean Perrin (1911) on the existence of atoms, and Stanley Milgram (1974) on the social context of obedience. For many years the speaker has begun his sophomore seminar "Numbers and Reasons" with a review of these three arguments, which, taken together, encompass most of the valid ways in which statistics can persuade anybody of anything. The talk sketched the pedagogy involved and the way in which these themes can shape additional teaching in elementary courses about "where numbers come from".

4. SRTL-3: THE THIRD INTERNATIONAL RESEARCH FORUM ON STATISTICAL REASONING, THINKING, AND LITERACY

University of Nebraska, Lincoln, Nebraska, USA, July 23-28, 2003

REPORT BY DANI BEN-ZVI AND JOAN B. GARFIELD

The third research forum in a series of international research forums on statistical reasoning, thinking and literacy (SRTL) took place this summer at the Teachers College Institute of the University of Lincoln-Nebraska. This particular gathering of researchers has played an important role in advancing our understanding of the richness and depth of reasoning about variability, a key focus of statistics education. The forum was sponsored by Teachers College Institute of the University of Lincoln-Nebraska (UNL), Department of Educational Psychology at the University of Nebraska-Lincoln (UNL), Faculty of the Quantitative and Qualitative Methods in Education (QQME) Program at UNL, the IASE Statistical Education Research Group (IASE-SERG), the American Statistical Association (ASA) Section on Statistics Education, and Vanderbilt University.

Eighteen researchers in statistics education from six countries shared their work, discussed important issues, and initiated collaborative projects in a stimulating and enriching environment. Sessions were held in an informal style, with a high level of interaction. With emphasis on reasoning about variability, a wide range of research projects were presented spanning learners of all ages, as well as teachers (see abstracts of studies below). These demonstrated an interesting diversity in research methods, theoretical approaches and points of view. As a result of the success of this gathering, plans are already underway for the next gathering (SRTL-4) in 2005.

The research forum proved to be very productive in many ways. Several types of scientific publications will be produced including a CD-ROM of the proceedings, papers in refereed journals, and a special issue of Statistics Education Research Journal (SERJ) on reasoning about variability. An additional product of the meeting will be a new SRTL Website that will include a variety of resources (srtl.haifa.ac.il/). These will all serve as a rich resource for statistics educators and researchers.

Bill Mickelson (UNL) took care of all aspects of local planning and organizing prior to the SRTL-3 gathering. During the research forum, he was very resourceful in attending to all needs of the program and the participants regardless of this being a time-consuming and labour intensive task. Beyond the scientific program, participants took part in a variety of social events and local excursions that helped to build the sense of a community amongst the researchers and enjoy the beauty of Lincoln and its immediate surrounds. Chris Reading was very also extremely helpful, building on her experience as the local organiser of SRTL 2 in 2001. For further information please contact the SRTL Co-chairs Joan Garfield (jbg@umn.edu) and Dani Ben-Zvi (dbenzvi@univ.haifa.ac.il).

4.1. THE SCIENTIFIC PROGRAM

The focus of this gathering on reasoning about variability emerged from the previous two conferences (SRTL-1 in 1999, SRTL-2 in 2001). Variability stands in the heart of statistical theory and practice. “Variation is the reason why people have had to develop sophisticated statistical methods to filter out any messages in data from the surrounding noise” (Wild & Pfannkuch, 1999, Statistical thinking in empirical enquiry, *International Statistical Review*, *67*(3) p. 236). In this gathering we aimed at investigating some the following questions:

What is variability in data? Is there a distinction between variability and variation? Why is variability important? When and how do children begin to develop the preliminary idea of variability? What are the simplest forms that children can understand? How does reasoning about variability develop? What are instructional tasks and technological tools that promote the understanding of variability? What are the common misconceptions regarding variability? What are the difficulties that people encounter when dealing with variability in data? What does correct reasoning about variability look like? What are ways to assess understanding of variability? How does an understanding of variability connect and effect understanding of other statistical concepts and types of reasoning? What are useful methodologies for studying the understanding of variability? What type of understanding of variability is sufficient for a statistically literate person?

The program began with an overview talk by Robert Gould (rgould@stat.ucla.edu), titled: “Variability: One statistician’s view”. The ten presentations of SRTL-3 were thematically grouped into clusters. A cluster included two one-hour research presentations to the entire group, small group discussions, small group reports and finally a whole group discussion. Most presenters showed a small subset of video segments of their research. In addition, optional time was devoted to viewing and discussing the research video-tapes.

4.2. SRTL-3 STUDIES PRESENTED

Chris Reading. *Student perceptions of variation in a real world context.*

This research qualitatively analysed responses of students (aged 13 to 17) in an inference task on weather with respect to the descriptions of variation that were incorporated. A Data Description hierarchy is used to code the better responses and is extended to accommodate a range of less statistically sophisticated responses identified. The SOLO Taxonomy is used as a framework for the hierarchy. Two cycles of U-M-R levels, one for more qualitatively descriptions and the other for more quantitative descriptions, are identified in the responses.

Jane M. Watson & Ben A. Kelly. *Developing intuitions about variation: the weather.*

This study uses the weather context to explore students’ development of intuitive ideas of variation from the third to the ninth grade. Three aspects of understanding these intuitions associated with variation are explored in individual video taped interviews with 66 students: explanations, suggestions of data, and graphing. The development of these three aspects across grades is explored, as well as the associations among them. Fifty-eight of the students also answered a general question on the definition of variation and variable and these responses are discussed and compared with responses to the weather task.

Robert C. delMas & Yan Liu. *Exploring students’ understanding of statistical variation.*

The main purpose of this study is to gain a better picture of the different ways that students look at the standard deviation as this concept develops at the beginning of an introductory statistics course. Thirteen students registered in an introductory statistics course are interviewed while interacting with a computer program, which was designed for this study. Several strategies and rules that students use to handle standard deviation tasks are identified. While some students appear to start the interview with a fairly sophisticated understanding of factors that affect the standard deviation and how these factors work together, most students have a very simple, rule oriented approach.

Avital Lann & Ruma Falk. *What are the clues for intuitive assessment of variability?*

The purpose of this study is to determine the data features that influence people's intuitive judgment of the extent of dispersion of a set of numerical data. A questionnaire comprising eight items, in each of which an intuitive comparison between the extent of dispersion of two sets of numbers is required, was administered to 354 statistically naive first-year students. A procedure was designed to determine each student's diagnosed intuitive measure of dispersion. Out of the four dispersion measures, the range is the most "popular", then the variance, MAD, and finally the IQR.

Dani Ben-Zvi. *The emergence of reasoning about variability in comparing distributions: A case study of two seventh grade students.*

The focus in this paper is on the emergence of beginners' reasoning about variation in a comparing groups situation during their encounters with Exploratory Data Analysis (EDA) curriculum in a technological environment. This study concentrates on the qualitative analysis of the ways by which two seventh grade students started to develop views (and tools to support them) of variability in comparing groups using various numerical, tabular and graphical statistical representations.

Arthur Bakker. *Reasoning about shape as a pattern in variability.*

The main question of this research was how to promote coherent reasoning about variability, sampling, data, and distribution in a way that is meaningful for students with little statistical background. Two instructional activities that support such reasoning are presented in this paper: growing a sample and reasoning about shape.

James K. Hammerman & Andee Rubin. *Reasoning in the presence of variability.*

This study focus on the ways that people work with the variability of data to make it more manageable and comprehensible, especially when using software tools that make it easier to view and manipulate data. The data for this paper come from VISOR – a teacher professional development and research project studying how people learn about data analysis and statistics and how computer visualization tools can enhance that learning. Several techniques that people use to reduce the need to attend simultaneously to signal, noise, and sample size are described.

Katie Makar & Jere Confrey. *Chunks, clumps, and spread out – secondary preservice teachers' informal notions of variation and distribution.*

In this paper, the authors examine how prospective secondary mathematics and science teachers use their own words to compare the performance of students on state exam. Through a set of interviews conducted at the beginning and end of a one-semester university course in assessment, equity and data analysis, their descriptions of the data distributions under scrutiny are analyzed and categorized. Their language is often informal, but reflect a strong emergent intuition about variation and distribution in data.

Maria Meletiou-Mavrotheris & Carl Lee. *Studying the evolution of students' conceptions of variation using the transformative and conjecture-driven research design.*

The paper describes how the transformative and conjecture-driven research design, a research model that utilizes both theory and common core classroom conditions, was employed in a study examining introductory statistics students' understanding of the concept of variation. It describes how the approach was linked to classroom practice and was employed in terms of research design, data collection, and data analysis. The rich insights into the evolution of students' thinking about variation that have originated from this research are then discussed. Implications for research and instruction follow.

William T. Mickelson & Ruth M. Heaton. *Purposeful statistical investigation merged with k-6 content: Variability, learning and teacher knowledge use in teaching.*

This study explores the role variability plays in teaching and learning. The paper analyses what teachers need to know about variability and be able to do with variability in data so that purposeful investigations into topics of the curriculum can be successful in teaching both statistical concepts and process and the important ideas associated with content. The findings of this study point toward the situated nature of knowledge about variability needed for and used in teaching and leads to significant implications for the growth of teachers' statistical knowledge.

FORTHCOMING IASE CONFERENCES

1. IASE 2004 RESEARCH ROUND TABLE ON CURRICULAR DEVELOPMENT IN STATISTICS EDUCATION
Lund, Sweden, June 28 - July 3, 2004

The International Association for Statistical Education (IASE) and the International Statistical Institute (ISI) are organizing the 2004 Roundtable on Curricular Development in Statistics Education, which will be held at Lund Institute of Technology at Lund University in Lund, Sweden from 28 June to 3 July 2004. The Roundtable will bring together a small number of experts, representing as many different countries as possible, to discuss one another's views and approaches to curriculum for teaching statistics. The Roundtable Conference will provide opportunities for developing better mutual understanding of common problems and for making recommendations concerning the statistics curriculum. A main outcome of the Roundtable will be a monograph containing a set of papers, which have been prepared for and discussed during the conference. The monograph will present a global overview of the conference that can serve as starting point for further research on issues related to the statistics curriculum.

The need for processing the increasing amount of data people receive in the course of their work and lives has made it imperative that students leave elementary and secondary schools prepared to make reasoned decisions based on sound statistical thinking. Countries and communities have approached this problem in different ways. The Roundtable will provide the opportunity for sharing what works and to highlight the challenges and potential solutions researchers and teachers have faced as they design and implement curricula to produce statistically literate citizens. The Roundtable will be held immediately prior to the Tenth International Congress on Mathematical Education to be held in Copenhagen, Denmark in 2004, July 4-11.

The IASE Scientific Program Committee will prepare the program and schedule for the Roundtable. The Committee has agreed on a list of topics that will form the basis of the discussions and invites those interested to send in a three-page summary of their proposed paper. The major topics to be addressed at the primary, secondary, tertiary, or inservice levels are: Relationship between curriculum and assessment; Role of research in shaping curriculum; Impact of technology on the statistic and probability curriculum; Innovative curricular practices; Teacher preparation and Statistical literacy.

- Theoretical papers should include; a) the statement of the problem, b) background or appropriate previous work, c) discussion of main arguments, d) implications for curricular development, e) references.
- Descriptions of experimental research should include; a) the statement of the problem b) background or appropriate previous work; c) methodology, data analysis and discussion of main results; d) implications for curricular development; e) references.
- Descriptions of curriculum innovations should include; a) focus and philosophy of the curriculum, b) background and development process, c) description, d) pilot and implementation results, e) sources and references.

Lena Zetterqvist (lena@maths.lth.se) and Ulla Holt will be local organisers. Those interested can contact Gail Burrill, Division of Science and Mathematics Education, College of Natural Science, Michigan State University, 116 North Kedzie, East Lansing MI 48824, USA, Email: (burrill@msu.edu). For more information, see http://hobbes.lite.msu.edu/~IASE_2004_Roundtable/.

2. STATISTICAL ACTIVITIES AT THE 10TH INTERNATIONAL CONGRESS ON MATHEMATICS EDUCATION
Copenhagen, Denmark, July 4-11, 2004

As a part of the 10th International Congress on Mathematical Education to be held in Copenhagen, Denmark July 4-11, a set of sessions have been set aside to address issues related to research and development in the teaching and learning of probability and statistics. Jun Li of the Department of Mathematics at East China Normal University and Joe Wisenbaker of the Department of Educational Psychology at the University of Georgia, are co-chairs of Topic Study Group 11, Research and development in the teaching and learning of probability and statistics. Team Members are Dani Ben-Zvi, Manfred Borovcnik, and Maxine Pfannkuch. They all encourage submissions across a wide range of issues including but not limited to:

- The use of technology to enhance student learning,
- Efforts to understand how students learn about statistics and probability,
- Developing teachers' statistical knowledge,
- Distance education,
- Assessment strategies as a means of promoting learning,
- Efforts to introduce younger students to statistics and probability, and
- Developing statistical reasoning, thinking and literacy.

Initial proposals for papers should be submitted to either of the co-chairs no later than November 30. Links to more details concerning submission requirements can be found at http://www.icme-organisers.dk/tsg11/.

The ICME-10 venue will be the Technical University of Denmark, located in a northern suburb of Copenhagen. Chair International Programme Committee: Mogens Niss (ICME10-IPC@ruc.dk). Chair Local Organising Committee Morten Blomhøj (ICME10-LOC@ruc.dk). More detail about the overall conference including the full list of topics included in the program, advance registration, housing, etc. can be found at http://www.icme-10.dk/.

3. IASE ACTIVITIES AT THE 55TH SESSION OF THE INTERNATIONAL STATISTICAL INSTITUTE
Sydney, Australia, April 5-12, 2005

Chris Wild is the IASE representative at the ISI Programme Co-ordinating Committee for ISI-55th Session, to be held in Sydney, Australia, April 5-12, 2005. The sessions approved for ISI 55 in Sydney that were sponsored or co-sponsored by IASE are as follows (titles may change slightly). More information from Chris Wild at c.wild@auckland.ac.nz.

- Reasoning about Variation.
- The use of Simulation in Statistics Education
- Teaching Statistics Online
- Statistics for Life: What are the Statistical Ideas or Skills that Matter most and why?
- Research in Statistical Education
- Teaching Bayesian Statistics
- Challenges in the Teaching of Survey Sampling
- Using History of Statistics to Enhance the Teaching of Statistics
- Promotion of Statistical Literacy among Students
- Quality Assurance in Statistics Education
- Educating the Media on how best to Report Statistics
- Ethical Standards in Statistics Education

4. ICOTS-7: WORKING COOPERATIVELY IN STATISTICS EDUCATION, SALVADOR (BAHIA), BRAZIL, JULY 2-7, 2006

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

The major aim of ICOTS-7 is to provide the opportunity for people from around the world who are involved in statistics education to exchange ideas and experiences, to discuss the latest developments in teaching statistics and to expand their network of statistical educators. The conference theme emphasises the idea of *cooperation,* which is natural and beneficial for those involved in the different aspects of statistics education at all levels. Some examples are given below.

- *Cooperative learning in statistics education*. Recent trends in educational psychology emphasise the role of student activity and social interaction in learning. These developments are particularly important in the case of statistics where students are taking a more active role in working on cooperative projects and studies.
- *Cooperation between statistics teachers and researchers*. Real life applications generated by working with a researcher in another area help motivate the teaching of statistics. The subject is more enjoyable for students when a teacher can call on such real applications. At the same time, teachers are an essential part of a research team in statistics education, since they collaborate both in collecting data from the students and in helping with the design and evaluation of action-research programs.
- *Cooperation between statistical agencies and statistics educators*. Statistical agencies need the cooperation of the population at large when collecting their data. They are also interested in improving the statistical literacy of their citizens. Consequently, the agencies are communicating statistical ideas to their populace as well as providing official data for research on different topics, including teaching. Statistical offices and educators collaborate in the development of teaching resources based on official data and set up workshops and conferences on the teaching of statistics.
- *Interdisciplinary cooperation for research*. Interdisciplinary research is natural both in applied statistics and statistics education. Many central statistical concepts and procedures arose from research problems in other subjects. At the same time the researcher, whatever subject he or she is working in, benefits by having problems actually solved. Statistics education is based on many different disciplines, such as psychology, pedagogy, statistics and sociology, which all contribute in their own unique way to the study and solution of teaching problems.
- *International cooperation in statistics education*. Global communication and increasing interest and respect for complementarity in education are leading to an increasing number of successful international research or educational programs at different levels: e.g., Large scale statistical literacy comparative studies; Regional, National or International funded projects; International statistical education centres; International training programs or conferences in statistics education.
- *Globalization and diversity in statistics education*. Cooperation requires both global and local approaches to research and teaching. There is a contrast and a complementarity of global and local approaches in statistics education; e.g., large sample, quantitative studies versus qualitative and ethnographic research; the need to recognise global tendencies, and at the

same time being sensitive to specific difficulties or talents of special and gifted students, minorities, etc.

The Conference will include keynote speakers, invited speakers, contributed papers, workshops and forums, demonstration lessons, roundtable sessions, poster sessions, book and software displays. People interested in organising a session or in presenting a paper are encouraged to contact the appropriate Topic Convenor. More information is available from the ICOTS7 Web page at http://www.maths.otago.ac.nz/icots7/ and from Carmen Batanero (batanero@ugr.es).

4.1. TOPICS AND TOPIC CONVENORS

Topic 1. *Working cooperatively in statistics education.* Lisbeth Cordani, lisbeth@maua.br and Mike Shaughnessy, mike@mth.pdx.edu

Topic 2. *Statistics Education at the School Level*. Dani Ben-Zvi, benzvi@univ.haifa.ac.il and Lionel Pereira, lpereira@nie.edu.sg

Topic 3. *Statistics Education at the Post Secondary Level*. Martha Aliaga, martha@amstat.org and Elisabeth Svensson, elisabeth.svensson@esi.oru.se

Topic 4. *Statistics Education/Training and the Workplace*. Pedro Silva, pedrosilva@ibge.gov.br and Pilar Martín, pilar.guzman@uam.es

Topic 5. *Statistics Education and the Wider Society*. Brian Phillips, BPhillips@groupwise.swin.edu.au and Phillips Boland, Philip.J.Boland@ucd.ie

Topic 6. *Research in Statistics Education*. Chris Reading, creading@metz.une.edu.au and Maxine Pfannkuch, pfannkuc@scitec.auckland.ac.nz

Topic 7. *Technology in Statistics Education.* Andrej Blejec, andrej.blejec@uni-lj.si and Cliff Konold, konold@srri.umass.edu

Topic 8. *Other Determinants and Developments in Statistics Education*. Theodore Chadjipadelis, chadji@polsci.auth.gr and Beverley Carlson, bcarlson@eclac.cl

Topic 9. *An International Perspective on Statistics Education.* Delia North, delian@icon.co.za and Ana Silvia Haedo, haedo@qb.fcen.uba.ar

Topic 10. *Contributed Papers*. Joachim Engel, Engel_Joachim@ph-ludwigsburg.de and Alan McLean, alan.mclean@buseco.monash.edu.au

Topic 11. *Posters*. Celi Espasandín López, celilopes@directnet.com.br

4.2. LOCAL ORGANISERS

Pedro Alberto Morettin, (Chair; pam@ime.usp.br), Lisbeth K. Cordani (lisbeth@maua.br), Clélia Maria C. Toloi (clelia@ime.usp.br), Wilton de Oliveira Bussab (bussab@fgvsp.br), Pedro Silva (pedrosilva@ibge.gov.br).

4.3. IPC EXECUTIVE

Carmen Batanero (Chair, batanero@ugr.es), Susan Starkings (Programme Chair, starkisa@vax.sbu.ac.uk), Allan Rossman and Beth Chance (Editors of Proceedings; arossman@calpoly.edu; bchance@calpoly.edu), John Harraway (Scientific Secretary: jharraway@maths.otago.ac.nz), Lisbeth Cordani (Local organisers representative; lisbeth@maua.br).

OTHER FORTHCOMING CONFERENCES

1. OZCOTS-5 AUSTRALASIAN STATISTICAL EDUCATION
Swinburne University of Technology, Australia, December 1-2, 2003

In December 1998 the first OZCOTS was held following the Fifth International Conference on Teaching Statistics, ICOTS-5. This gave people the opportunity to hear Australians who presented at ICOTS-5 and to retain links formed in Singapore. This year the fifth meeting in this very successful series is hosted by the School of Mathematical Sciences, Swinburne University of Technology. It is planned to continue to give the opportunity for all interested to learn about the latest in Statistical Education. More information from Brian Phillips, at bphillips@swin.edu.au. Web page: http://www.swin.edu.au/maths/iase/ozcots5.html.

2. ESPACE MATHÉMATIQUE FRANCOPHONE
Tozeur, Tunisia, December 19-23, 2003

This international colloquium is organised jointly by the Tunisian Commission for the Teaching of Mathematics, the Tunisian Association of Mathematical Sciences, and with the co-operation of the French Commission for the Teaching of Mathematics. It will give the international community of French speaking specialists the opportunity to address the main problems concerning the teaching of Mathematics in our societies, mainly at the elementary and high school levels, as well as in teacher training. Web page: http://www.mathinfo.u-picardie.fr/EMF2003.

3. MATHEMATICS EDUCATION - THEN, NOW, AND THE FUTURE
Oklahoma City, USA, February 19-21, 2004

The Research Council on Mathematics Learning seeks to stimulate, generate, coordinate, and disseminate research efforts designed to understand and/or influence factors that affect mathematics learning. The conference is being held in the metropolitan Oklahoma City area. Organisers are especially interested in "works in progress" and feel this conference is a wonderful opportunity for pre-tenured faculty and graduate students to present their ideas and receive feedback from some of the leading mathematics educators in the country. More information from Jayne Fleener, at fleener@ou.edu. Web page: http://www.unlv.edu/RCML/conference2004.html.

4. ATM'S 2004 EASTER CONFERENCE
Loughborough, United Kindgdom, April 5-8, 2004

ATM has been involved in the business of the professional development of mathematics educators for over fifty years. The ATM Easter Conference provides opportunities for delegates involved with mathematics education related to any age to; share in successful practices, develop their own mathematics, tap into recent pedagogical and curriculum research, learn about different international developments and practice, access and evaluate recently produced resources, find out about and experience the most recent information and communication technological developments, join in new research networks and projects, learn about the most recent national developments in mathematics education and take new and renew contacts with the professional network of mathematics educators. Web page: http://www.atm.org.uk/conferences/conference2004.html.

5. SYMPOSIUM ON BIOMEDICAL INFORMATICS AND BIOMEDICAL STATISTICS EDUCATION
Prague, April 12-15, 2004

The symposium will be held at the occasion of the 10th anniversary of the European Centre of Medical Informatics, Statistics and Epidemiology (EuroMISE Centre) of Charles University and Academy of Sciences of the Czech Republic. More information from L. Semeráková at semerakova@euromise2004.org. Web page: http://www.euromise2004.org/about/symposium1.html.

6. ANNUAL CONFERENCE OF AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (AERA)
San Diego, California, USA, April 12-16, 2004

AERA Meetings Department, American Educational Research Association, 1230 17th Street NW, Washington, DC 20036, Email: 2004annualmtg@aera.net. Web page: http://www.aera.net/meeting/

7. HAWAII INTERNATIONAL CONFERENCE ON STATISTICS, MATHEMATICS AND RELATED FIELDS
Honolulu, Hawaii, USA, June 9-12, 2004

The 2004 Hawaii International Conference on Statistics, Mathematics and Related Fields will be held from June 9 (Wednesday) to June 12 (Saturday), 2004 at the Sheraton Waikiki Hotel in Honolulu, Hawaii. The conference will provide many opportunities for academicians and professionals from statistics and/or mathematics related fields to interact with members inside and outside their own particular disciplines. Cross-disciplinary submissions with other fields are welcome. More information from statistics@bogus.example.com. Web page: http://www.hicstatistics.org.

8. HISTORY AND PEDAGOGY OF MATHEMATICS 204 SATELLITE CONFERENCE OF ICME-10
Uppsala, Sweden, July 12-17, 2004

The HPM satellite conferences have taken place every 4 years since 1984. They provide a unique occasion to attend lectures, workshops, research reports from all over the world about the use of history in mathematics education and history of mathematics. The participants to the HPM meetings are researchers in history and in mathematics education, and teachers who have experimented with the use of history in their teaching. More information from Fulvia Furinghetti, furinghe@dima.unige.it. Web page: http://www.mathedu-jp.org/hpm/index.htm.

9. PME-28
Bergen, Norway, July 14-18, 2004

The 28th Annual Meeting of the International Group for the Psychology of Mathematics Education, will take place in Bergen University College, July 14 to July 18 2004. The theme of the conference is “Inclusion and Diversity”. Conference Secretariat Bergen University College Faculty of Education Landåssvingen 15 N-5096 Bergen, Norway. Email: secretariat@pme28.org. Web page: http://www.pme28.org/.

10. JOINT STATISTICAL MEETINGS
Toronto, Canada, August 8-12, 2004

This meeting is to be held at the Metro Toronto Convention Centre and Royal York Hotel, Toronto, Canada, August 8-12. It is sponsored by the American Statistical Association, ENAR, WNAR, IMS, and SSC. More information from the Americal Statistical Association, at meetings@amstat.org. Web page: http://www.amstat.org/meetings/jsm/2004/.

STATISTICS EDUCATION RESEARCH JOURNAL REFEREES
DECEMBER 2002 – NOVEMBER 2003

Acknowledgement: We are grateful to the following IASE members and researchers who acted as referees for the Statistics Education Research Journal in the period from December 2002 to November 2003:

Martha Aliaga, USA
Richard Alldredge, USA
Livia d'Andrea, USA
Kant Bandgdiwalla, USA
Gabriella Belli, USA
Carolyn Boyle, USA
Carmen Capilla, Spain
Katarina Cobanobic, Slovenia
Antonio Estepa, Spain
Assumpta Estrada, Spain
Celi Aparecida Espasandín, Brazil
Brian Greer, USA
David Griffiths, Australia
Michel Henry, France
Sterling Hilton. USA
Yamamoto Kentaro, USA
Irene Mauricio-Cazorla, Brazil
Verena Nolan, South Africa
Andrés Nortes-Checa, Spain
Lionel Pereira-Mendoza, Singapore
Dave Pratt, UK
Victor Polaki, Lesotho
Jorge Luis Romeu, USA
Ernesto Sánchez, México
David Saville, New Zealand
Candance Schau, USA
Dalene Stangl, USA
Nigel Smeeton, UK
Susan Starkings, UK
Dirk Tempelmaar, Netherlands
Joe Wisenbaker, USA