Statistics Education Research Journal

Volume 2 Number 1 May 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;
- improving the teaching of statistics at all educational levels.

It encourages the submission of quality papers, including research reports, theoretical or methodological analyses, literature surveys, thematic bibliographies, summaries of research papers and dissertations. Contributions in English are recommended. Contributions in French and Spanish will also be accepted. All the papers will be refereed.

ISSN: 1570-1824

Submissions

Manuscripts should be sent to Flavia Jolliffe (F.Jolliffe@kent.ac.uk), by e-mail, as an attached document, in rtf format. Two files are required. In one of them the author's name and affiliation should be removed to ensure anonymity in the reviewing process. Manuscripts should conform to the style specified at the Journal web page: http://fehps.une.edu.au/serj

TABLE OF CONTENTS

EDITORIAL

Since the previous issue we have been working in various ways to improve SERJ and to expand its impact on statistics education research. Different extracting services, associations, and statistics education web server masters were contacted with a request to help us in this task. We received a positive response from many of them.

SERJ is now extracted in the Mathdi Database and the Current Index to Statistics. A link to the SERJ web page is now included in the Journal of Statistics Education (www.amstat.org/publications/jse) and JSE is also including summaries of our papers. A link to SERJ is now contained on 45 other web pages, including those of the International Statistical Institute, the International Association for Statistical Computing, the American, Australian, Canadian and New Zealand Statistical Associations, the American Education Research Association, CIRDIS and the Royal Statistical Society Statistics Education Centre, and we plan to continue contacting webmasters of other relevant web pages. Please let us know of any suggestions you have as to whom we might contact.

The impact of a journal depends heavily on receiving good papers and on being quoted. We are pleased to let you know that the flow of papers is increasing quickly and at this moment we have already accepted enough papers to publish the next issue, Volume 2 Number 2, later in 2003. In addition we have a number of papers currently undergoing refereeing. We thank you for your confidence and we expect this support to continue in the future. In the same way we encourage you to quote SERJ and to include a link in your Department or personal web pages. Please let us know of any such links.

We have two papers written by associate editors, Iddo Gal and Joan Garfield, in this issue. Both papers were submitted before Iddo and Joan were invited to join the editorial board, but the fact that they submitted the papers and agreed to join the board is an indication of their commitment to statistics education and to SERJ. In his paper Iddo expands his ideas about statistical literacy and presents an exploratory study in an innovative area. Joan gives an analysis of her Statistical Reasoning Assessment, an instrument which has proved to be useful in different research settings, and at the same time she presents data from a cross-cultural study. We also have a paper by Peter Petocz and Anna Reid in which they report on an empirical research project looking at relationships between students' conceptions of learning statistics and the teaching that they receive. For the first time we are also including a paper in Spanish. This paper deals with how University students solve correlation and regression problems, an under-researched topic. These papers are complemented by our usual sections. Here we should like to draw your attention to summaries of papers presented at CERME 3 (Congress of European Research in Mathematics Education), where, for the first time a stochastic group, with 29 participants from Europe and other countries, was run.

The editors and associate editors have held long email discussions about the formatting of SERJ and we are trying a different format, with wider margins, a font change from Arial Narrow to Times New Roman, and some space after paragraphs, in this issue. We hope you like the change and would be interested to receive your comments and feedback.

We are delighted to announce that Iddo Gal has agreed to take over from Carmen as editor from the 1 December 2003, in spite of his many other commitments. Flavia will continue as editor for one more year, and after that the plan is that editors will each serve for two years. Carmen will continue to serve on the editorial board as an associate editor for a while.

Carmen Batanero and Flavia Jolliffe

EXPANDING CONCEPTIONS OF STATISTICAL LITERACY: AN ANALYSIS OF PRODUCTS FROM STATISTICS AGENCIES

IDDO GAL
University of Haifa, Israel
iddo@research.haifa.ac.il

SUMMARY

This paper reports the results of an exploratory study of the characteristics of key information products released by statistics agencies. Such products are central to debates and decisions in the public arena, but have received little attention in the literature on statistical literacy, statistics education, or adult numeracy. Based on a qualitative analysis of Internet-based products of six national and international statistics agencies, the paper sketches the characteristics of five product types (Indicators, Press releases, Executive summaries, Reports, and Aggregate data) and of the environment in which they are found. The paper discusses implications for the specification of the skills needed for accessing, filtering, comprehending, and critically evaluating information in these products. Directions for future research and educational practice are outlined.

Keywords: Statistics education research; Statistical literacy; Official statistics; Educational technology; Adult numeracy; Mathematics education

1. INTRODUCTION

Increasing attention has been given over the last decade by the statistics and mathematics education communities to the development of statistical literacy and numeracy skills of all citizens (Gal, 2000). This trend has most recently been exemplified in the overall theme of ICOTS-6, the 6th International Congress on Teaching Statistics (Cape Town, South Africa), "Developing a Statistically Literate Society". Multiple paper sessions and presentations in this conference directly or indirectly addressed issues related to this theme (Phillips, 2002).

The term "statistical literacy" has not yet gained an agreed-upon meaning among educators and professionals, and some use it without an explicit definition (Cerrito, 1999). The view of statistical literacy that guides this paper stems from the assumption that most adults will be consumers, rather than producers, of statistical information. Diverse but related conceptions of statistical literacy have followed from this assumption. Wallman (1993) argued that statistical literacy is the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions. According to Gal (2002a), statistical literacy refers to people's ability to interpret, critically evaluate, and when relevant express their opinions regarding statistical information, data-related arguments, or stochastic phenomena. Lehohla (2002) views statistical literacy as the ability to read and understand quantitative information such as indices and indicators.

While statistical literacy appears as a key competency for many adults and for various professionals and officials, discussions of the kinds of information to which students and

Statistics Education Research Journal 2(1), 3-21, http://fehps.une.edu.au/serj

adults have to be able to apply this competency are lacking. Many educators focus on developing students' ability to be aware of or not be fooled by "misleading" statistics or "biased" reports in the *media* (Crossen, 1994; Cerrito, 1999; Moreno, 2002). Very few sources focus on the ability to read and critically interpret information from other sources (Frankenstein, 1990; Gelman, Nolan, Men, Warmerdam, & Bautista, 1998). This paper aims to add to an emerging dialogue on the promotion of statistical literacy, by examining the nature of the products made available to the public by *statistics agencies* and by reflecting on the skills demands of these products.

1.1. CLIENTS AND PRODUCTS OF STATISTICS AGENCIES

The term *"statistics agencies"* is used here to encompass three types of statistics-producing organizations operating in the public sphere and funded by governments:

1. "*National/Central agencies"*: National statistical offices and organizations responsible for conducting a nation's *census* and associated surveys and reporting their results (e.g., the United States Bureau of the Census), or for producing all national social and economic statistics, including the census (e.g., Statistics Sweden, Italian National Statistical Institute-ISTAT).
2. *"National thematic agencies":* National organizations assigned by their governments to collect and report official statistics in designated areas not covered by the agency conducting the census or by other national agencies. Examples are the U.S. National Center for Education Statistics, or the Australian Institute of Health and Welfare.
3. *"International agencies":* These are international organizations established by member nations in order to contribute to social, human, and economic progress. One of their primary missions is the reporting of comparative statistics collected by member nations, but they also initiate or conduct special comparative studies. Examples are: UNESCO Institute for Statistics, World Bank, Organization for Economic Cooperation and Development [OECD]. This category also includes ad-hoc multi-national research projects not associated with specific countries, such as the Third International Mathematics and Science Study [TIMSS].

These three types of organizations are viewed as *official* statistics agencies since they focus primarily (though not exclusively) on statistical and research work aimed at informing policy makers, and are funded by tax-based budgets. Of course, many other organizations which release statistics to the public do exist, but fall outside this definition because of their funding sources or the issue-specific nature of their statistical work. Examples are Non Governmental Organizations (NGOs) such as Amnesty International, non-profit research institutes and advocacy groups, and industrial companies or survey organizations such as Gallup.

Murray and Gal (2002) analyzed products from Statistics Canada, a central national agency responsible for the census and other official statistical work in Canada, and grouped them under five key categories: (1) *Indicators*, (2) *Press releases*, (3) *Executive summaries and Highlights*, (4) *Reports*, and (5) *Aggregate data*. (They also listed other products, such as raw data files or technical documentation that are mostly of interest to specialists rather than to the general public). Murray and Gal (2002) argued that information products from statistics agencies serve multiple clients, but are especially geared for policy-makers and politicians and designed to inform their decisions and policy setting. They suggested that such products are also of major interest to the media, whose job is to transform official statistical reports and findings into simpler messages that reach wide audiences who otherwise may be unaware of official statistics. In addition, official statistics are of interest to NGOs, business managers, and researchers and consultants. Finally, Murray and Gal claimed that understanding of information products provided by statistics agencies is essential for citizens at large and to interest and advocacy groups who desire to participate in public debate and influence the direction of decisions on local and national issues (European Commission, 1996).

The Murray and Gal (2002) work has provided a preliminary map of key information products from official statistics agencies, but little is known about their actual features. Such products overall have received little attention from both educators and researchers, although they have an important role in the information fabric of modern societies, and are the basis for many media messages that are of interest to the general public and to various officials and interest groups (European Commission, 1996; Podehl, 2002). With this in mind, this paper reports the results of an exploratory study designed to analyze the characteristics of key products of statistics agencies and of the environment in which users have to find these products. Such an analysis is an essential step toward expanding the understanding of the statistical tasks facing citizens; it can help to clarify the skills that should be considered part of statistical literacy and to identify targets for research and education efforts.

2. METHOD

2.1. APPROACH

Given the lack of prior research regarding products of statistics agencies, the approach taken was to use a multiple-case-study method as a basis for generating qualitative descriptions of the characteristics of key products of statistics agencies. Such an exploratory approach is an accepted way to aid the formulation of research directions and questions in an uncharted area. The study focuses on analysis of information appearing in *Internet sites* of statistics agencies since they provide clients, including researchers and educators, with access to all current key products (printed as well as interactive products not available in print), and since the Internet appears to be a central distribution medium for products of statistics agencies.

2.2. SAMPLE

This study used an intentional sample of six Internet web sites of statistics agencies, two from each of the three main categories listed above. Only agencies operating sites in English were considered, due to the need to compare both national and international agencies. The sample was designed to include agencies operating in countries with different types of governments, economies, and demographics, as well as international agencies with diverse missions. The chosen agencies created products encompassing a very wide range of themes and issues of interest to citizens and policy-makers in both developing and developed or knowledge-based nations.

The resulting sample included the agencies listed in Table 1 (acronyms appearing after the full agency name are used throughout this paper).

Table 1. Agencies Included in the Sample

Agencies and Acronyms	URL
National/Central agencies:	
a. Statistics South Africa (Statistics-SA)	www.statssa.gov.za
b. UK Statistics (Statistics-UK)	www.statistics.gov.uk
Thematic agencies:	
a. United States National Center for Education Statistics (US-NCES)	www.nces.ed.gov
b. Australian Institute of Health and Welfare (Australian-IHW)	www.aihw.gov.au
International agencies:	
a. UNESCO Institute for Statistics (UNESCO-IS)	www.unesco.org
b. Organisation for Economic Cooperation and Development (OECD)	www.oecd.org

2.3. PROCEDURE

All Internet sites were accessed in July 2002 and analyzed for the presence of the five product categories described in Table 2. A site was considered as carrying a product if at least three instances of this product type were found. An overall impression of the characteristics of the product category was reached through content analysis of at least three instances/items in each product category on each site.

Table 2. Product Descriptions

Category	Description
Indicators	Summary statistics that reflect the condition of key aspects of social, human, or economic importance. Examples: Gross National Product [GNP], Per-capita income, Pupil-teacher ratio, Infant mortality.
Press releases	Short summaries written to inform reporters and media personnel about changes in key indicators, about key findings from new or recent studies, etc.
Executive summaries and highlights	Summaries of studies, describing key findings and conclusions, without technical details. Designed primarily for policy makers.
Reports	Lengthier publications than Executive Summaries. Contain longer discussions, elaborate statistical tables and displays, and information about methodological and technical aspects of studies
Aggregate data	Detailed numerical information about indicators or other key variables that users can access directly, i.e., not in a Report.

3. RESULTS AND ANALYSIS

Overall, the product types listed in Table 2 were found on all six sites, yet ranged widely in their characteristics. Below is a sketch of the environments encountered when accessing the six sites, followed by an outline of the key features of the five product types. Subheaders in each section below sketch key findings in each area. (Note: exceptions to key findings were found in some cases, as could be expected given the breadth and complexity of the information carried by six different official agencies; to facilitate the presentation of overall trends, such exceptions are not always noted below).

3.1. OVERALL STRUCTURE AND USABILITY OF WEB SITES

This section sketches the characteristics of the Internet web sites of the six agencies analyzed. Attention was placed on how each site is organized, how descriptive and technical information is presented to the user, and how easy or difficult it is to use the interfaces or facilities provided on each website for finding needed materials, to access different products, or to perform various operations. Given that this exploratory study was designed to analyze five types of information products of statistics agencies with an eye towards educational or institutional implications, the analysis of ergonomic characteristics and of features that may affect the overall usability of the six websites was informal and qualitative in nature.

Agencies carry a large number of items on different topics

Most sites carried many hundreds and often thousands of separate information items (i.e., specific instances of a product, such as reports, executive summaries, or stand-alone tables). To illustrate the range of items carried, Statistics South Africa, one of the smaller agencies reviewed, carries over 70 press releases, over 60 reports, over 100 documents of various types defined as “publications” or “papers”, and hundreds of other types of stand alone tables and time series. Australian-IHW noted that it adds around 80 new reports a year. Most other agencies carried many more items. The OECD site lists 20 different product categories and enables access to thousands of different items.

Sites differ in their structure and organization

The homepage (the first screen the user encounters, from which various items can be located) of some sites was organized according to type of product sought by the user, while in others items were organized around areas or themes. For example, Statistics-SA lists on its homepage categories such as “Reports” or “Working Papers”. When one is chosen, the user can further select from a list of sub-categories or of products sorted alphabetically within sub-categories, requiring the user to skim long listings to find if a document on a specific topic exists. In contrast, upon entering the OECD site, the user encounters a list of over *30* distinct themes (e.g., ageing society, energy, health, education, corruption, taxation, etc.). Upon choosing any of them, a list of sub-themes is displayed and associated documents shown for each. As a result, users have to be able to use different search logics when searching for information in different sites.

Help and search options are cumbersome

Most sites offer help screens with tips on how to use the site, or a site map with an overview of the parts of the site, but the sheer volume of options in all sites requires that users adopt various strategies to locate needed information. To find a document on a specific topic, users have to scan multiple and sometimes long listings, or use the site’s search engine with the hope of reducing the labor involved in pinpointing potentially useful documents through list scanning. Yet, search engines follow a different logic in each site. Users have to generate alternate search terms or keywords when searches fail to provide adequate hits, or, when searches yield many hits, navigate through the resulting lists or try other keywords. Hence, users have to be able to combine different search facilities: browse product lists on screen, use a search function which is offered by the site, or use the search function of the browser with which they view the site.

Glossaries are formal and demanding

A search for a glossary, i.e., dictionary explaining the meaning of terms or computation of various variables, is a demanding process in most sites. A clearly marked and comprehensive central glossary could not be easily found on any of the sites reviewed. Requesting a search engine on any site to locate a "glossary" usually yielded multiple hits referring to glossaries developed independently by specific authors within some documents, forcing the user to open these documents in search of the desired definition. Such embedded glossaries are not available to users who do not reach into their host reports. Explanations in glossaries, when they could be found, were often formal and verbose, and some assumed prior technical or statistical knowledge (as illustrated by the example in section 3.2 below).

Documents may be created by multiple sources

Some agencies offered products created not only internally but also by outside sources, such as other statistics agencies, government departments, or collaborating institutions in member nations. For example, some of the products carried by Statistics-UK originated from collaborating agencies in Scotland and Wales, and Australian-IHW carried products created by departments in different Australian states. The existence of multiple source organizations can help explain the diversity in style and internal organization found in some of the product categories described below.

Access to information requires familiarity with various computer programs

All sites store many texts of interest in Acrobat Reader (PDF) or in Microsoft Word or Excel file formats, or in these formats inside compressed archives (ZIP files). Documents a user chooses to view are sometimes opened automatically by a plug-in module of the browser (without user intervention). However, in many cases users are asked to open or save the file, or may only be given the option to download (and save) the file, in which case they need to know how to locate the saved file on their computer and open it on their own in separate steps.

3.2. INDICATORS

Indicators are statistics that reflect the condition of key variables and phenomena in areas such as economic activity, health, education, or the environment. Some indicators are generated and reported on a recurring basis, such as every week or month (e.g., *Consumer Price Index, Gross Domestic Product*), and are of interest to local clients who need to know of changes in these measures as a basis for their ongoing activity. Others are derived on a yearly or ad-hoc basis (e.g., *Infant Mortality, Number of teachers, Transport-related Air Emissions*). All six sites presented information about many types of indicators, including historical and time series information. The following three interrelated issues emerged from the analysis:

Indicators are reported with varying levels of explanatory support

Agencies enable direct access to recurring indicators, though with different levels of ease. For example, Statistics-SA has a "Key indicators" option on the opening homepage, while users of Statistics-UK have to follow a longer route, by choosing "latest figures" but then browse a long list of releases to find those carrying indicators information. In both agencies, accompanying and sometimes lengthy notes explain changes in indicator values from prior months and discuss points that should be considered when interpreting the figures. Yet, the nature of the indicators themselves is not explained, probably on assumption they are standard entities with which all key clients are already familiar. In addition, in all six sites multiple lengthy reports are available that discuss the status of different indicators (e.g., changes in the status of the health system as reflected by indicators such as mortality, staffing levels, or expenditures).

Indicators vary in concreteness and transparency

Some indicators carry straightforward names (*Number of hospital beds*, *Adult literacy*) that make intuitive sense even for people without advanced levels of education or who do not know how the indicators are actually measured or calculated. Other indicators appear more abstract (*Rate of primary school enrollment, Life expectancy*) but still use common terms. Yet, some indicators carry complex names that make no intuitive sense and hence cannot be understood by users lacking specialized training. For example, *Apparent Intake Rates* (UNESCO-IS), or *Purchase Power Parities* (OECD). Possibly their meaning could be understood by consulting additional explanatory material, but as noted earlier most sites do not offer easy access to glossaries.

Understanding indicators requires familiarity with mathematical and statistical terms

A vague sense for the issue(s) addressed by an indicator may sometimes be sufficient to allow a user to understand the general meaning of indicator values. However, the names of some indicators require familiarity with mathematical or statistical terms. For example, knowing what “ratio” or “average” mean is a prerequisite for understanding *Pupil-teacher ratio* (an education indicator) or *Average length of stay* (a healthcare indicator). Full understanding of an indicator’s meaning, however, requires going beyond its name and coming to grips with how it is measured or calculated (especially if the indicator is a composite of other variables, as with *Gross national product*). For example, on UNESCO-IS, when users request additional information regarding the meaning of *School-life expectancy*, one of the 16 key indicators reported, the following glossary entry appears:

> *“The total number of years of schooling which a child of a certain age can expect to receive in the future, assuming that the probability of his being enrolled in school at any particular age is equal to the current enrolment ratio for that age”.*
>
> And the calculation method is defined as follows:
>
> *“For a child of a certain age, school life expectancy is calculated as the sum of the age specific enrolment ratios for the reference age-range a to n, divided by 100.”*

Complex glossary descriptions such as the above, which were also found on other sites, may reflect an attempt to be technically accurate and inform users of “best practices” of official statisticians. The price is that their understanding requires users to be familiar with school-based terms such as “total”, “probability” or cope with more complex mathematical and statistical terms. In addition, users have to grapple with the linguistic complexity created by composite phrases (e.g., “*age specific enrolment ratios*”, “*reference age-range*”).

3.3. PRESS RELEASES

The term “press release” refers to short summaries issued by statistics agencies to inform media personnel about changes in key indicators, about key findings from recent studies, or about new statistical data or results that may be of interest to the general public. As noted earlier, press releases are expected to be used by journalists or reporters as the basis for messages suitable for their own particular audiences.

Releases have a relatively uniform structure within agencies, less so across agencies

Press releases were found on all sites, and had a rather uniform style and format within most agencies, but a more pronounced variability in format and style across agencies. For example, press releases from OECD and US-NCES were normally 1-2 pages in length, and comprised of separate, relatively short paragraphs written in a journalistic style, i.e., in the way reporters may wish to write results and conclusions so as to pique reader's interest. Releases by Statistics-SA employed a journalistic style in discussing main points, but tended to make fewer interpretive statements and presented more raw statistics (absolute numbers and percentages) than OECD or US-NCES.

In contrast, "press releases" from Statistics-UK were very diverse in content and format. Some were only a short notice about the release of a new report, without any summary of results. Some also contained a link through which the user could access a table of contents of a large report and there review a "summary" or "highlights" section. Others present data tables (e.g., total numbers and causes of death last month in the UK), followed by brief background comments about the data collection process, but without a discussion of the implications of the data. Only some Statistics-UK releases were similar to those from OECD or US-NCES, but with a different internal structure and a more formal, technical style; these start with a single opening page containing a lead paragraph, then a few separate sentences presenting key points, and 1-2 simple graphs or a table. Additional pages either had text paragraphs discussing other findings, or multiple tables with supporting data but with little interpretive text.

Press releases vary in the amount of interpretation they offer the user

Press releases vary somewhat within agencies but more so across agencies in the extent to which they list raw statistical findings and "let the numbers speak for themselves", or on the other hand surround the findings with interpretations and commentary about their meaning, implications, or limitations. The diversity found in the amount of interpretive text may stem from different factors, such as differences across countries or agencies in expectations regarding the media's role in society, in assumptions regarding the ability of journalists to understand statistical information, or in the philosophy of agency administrators regarding what is proper to include in a press release.

When raw data or findings are presented to users without much interpretation or explanatory text, it may be argued that users benefit because they gain access to "objective" data or to "untainted" reporting. On the other hand, the lack of interpretations and explanations about the import of the findings or the overall meaning of separate findings puts more burden on the users. In such cases they need to interpret on their own factual statements and read messages carefully to notice patterns or links between findings presented separately. Thus, the amount of interpretation or commentary that agencies offer in press releases has implications for the type and level of background knowledge and skills that users need to possess.

Press releases use a wide range of statistical and mathematical terms

Along with diversity in the amount of interpretive commentary, press releases also varied in the range of statistical and mathematical terms and ideas they contained. A partial quote from the opening paragraphs of an US-NCES release about the National Assessment of Educational Progress' Geography 2001 survey illustrates the use of a journalistic style but also the use of multiple statistical and mathematical concepts and ideas:

> *A new report released today by ...National Center for Education Statistics shows that average scores of the nation's fourth and eighth graders, while low, have improved from 1994. Lower-performing students at grades four and eight showed an increase in average scale scores, whereas no overall changes were seen for 12th graders… the improvements for fourth-and eighth-graders were seen among students scoring in the tenth and 25th percentiles of performance.*

Clients seeking concise, non-technical information in press releases have to be prepared for the range of approaches to writing of press releases in different agencies. Clients may need to be able to read, interpret, and critically evaluate not only highly abstracted text, but also understand non-technical but also more advanced statistical or mathematical terms, and be able to elicit additional information from tables or graphical displays provided as part of the press release.

3.4. EXECUTIVE SUMMARIES AND HIGHLIGHTS

According to Murray and Gal (2002), policy makers and officials are the primary clients of statistics agencies. Given that these individuals do not have the time or need to deal with many details, they seek concise documents that provide an overall picture of key findings, or a summary of the main conclusions from a given study or activity of a statistics agency. Short summary publications that can satisfy the information needs of such key users were thus expected to exist as a distinct product category in all agencies examined in this study.

Executive summaries vary in length and location

The analysis provided mixed support for the presence of short publications aimed at policy makers or officials as a separate category from longer publications. In all sites it was possible to identify some brief stand-alone summary documents, often called Executive summary or Highlights, usually as brief as two pages but sometimes ten pages or more in length. Executive summaries or Highlights were more commonly found, however, as an integral opening section in a long Report.

Searching for executive summaries is complex and affected by site organization

As noted in 3.1, some agencies organize their homepages in terms of key products, others by topics. Users who are interested in reading a summary on a topic, not a full report, face a challenge: it is not obvious how to find an executive summary on a given topic where agencies place summaries inside reports, not as stand alone documents. Search engines do not enable identification of documents that contain an executive summary or highlights section, leaving users no choice but to download or open lengthy reports in order to locate executive summaries on topics of interest. An interesting example for the flexibility required of users was found in Australian-IHW, which presents to users two separate lists titled "Welfare" and "Health", each of which contains dozens of items. Almost none of the documents listed under "Welfare" were stand-alone Executive Summary or Highlights (though some reports opened with an Executive summary), whereas many documents listed under "Health" were marked as "Highlights". Thus, products in the two key theme areas of this agency were being shaped by different principles, requiring different reading and search strategies in the same site.

Executive summaries can demand substantial statistical and mathematical knowledge

As indicated above, Executive summaries varied in length. The longer ones included extended text passages, as well as graphs, charts, and tables to organize key results or show important trends discussed in detail in an accompanying Report. The range of statistical terms and concepts addressed in these summaries is very wide and covers the full spectrum of topics included in introductory as well as more advanced statistics textbooks or courses. Table 3 includes excerpts from executive summaries from three of the six agencies, chosen to illustrate the range of statistical concepts and ideas that users can expect to encounter in texts.

Table 3: Excerpts from Executive Summaries from three Agencies

#1: On a worldwide scale, the total number of primary school pupils of any age rose from just under 600 million to over 680 million between 1990 and 1998. The percentage of children of primary age in school edged upwards, from 80 per cent to 84 per cent. Since the beginning of the decade, primary enrolments have increased by an average of 10 million each year, almost twice that recorded in the 1980s.

#2: Latin America and the Caribbean and East Asia and the Pacific appear to be the only less developed regions with the capacities to provide education for all primary school-age children. Although their net enrolment ratios are still below 100 per cent, their gross enrolment ratios were 100 per cent and above throughout the decade.

#3: The disparities within regions are immense, however, with a difference of over 70 percentage points between the highest and lowest enrolment rates... The greatest variations can be seen in sub-Saharan Africa with gross enrolment ratios lower than 1 per cent in Congo to 111 per cent in the Seychelles.

#4: The number of pupils per teacher varies enormously on a global scale, from a low of 9:1 to a high of 72:1. In 1998, 75 per cent of the countries reporting had pupil/teacher ratios below 37:1...The highest ratios are found in Central and Western Africa, where the average (median) pupil/teacher ratio rose from an already high 50:1 in 1990 to 52:1 in 1998.

#5: For explanatory aspects of poverty, for example educational attainment and access to services, the IES data were merged with data from the 1995... household survey (OHS)... A series of regression analyses was carried out, using annual household expenditure as the dependent variable, and the poverty-related variables common to the OHS and the census as the explanatory variables, to impute expenditure values for each household....Two...indices – the household infrastructure index and the household circumstances index – were constructed to measure the extent of under-development in different parts of [the country], using both the data from Census '96, and the imputed expenditure values described above.

#6: (Table subtitle) Relationship between average performance across combined reading, mathematical and scientific literacy scales and cumulative expenditure on educational institutions up to age 15.

#7: (Explanation on how to read a complex table with over 30 country names on both the X and Y axis): The figure below summarises the performance of countries on the reading literacy scale. It also indicates whether countries perform significantly higher or lower than the comparison countries as well as the estimated rank order position of each country. For example, Finland, with all triangles pointed up performed significantly better than all other countries while Canada performed significantly lower than Finland shown by a triangle pointed down, similarly to New Zealand, Australia, Ireland and Japan shown by a circle and significantly better than all other countries shown by a triangle pointed up.

#8: Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

#9: The fact that [the survey] measures proficiency not at the aggregate country level, but at the level of individual students, makes it possible to also examine variation in student performance within countries. Such variation may result from the socio-economic backgrounds of students and schools, from the human and financial resources available to schools, from curricular differences, from selection policies and practices and from the way in which teaching is organised and delivered.

The excerpts in Table 3 show that Executive summaries make reference to all key types of standard summary descriptive and inferential statistics, such as totals (#1), percents (#1-#3), percents below 1% or larger than 100% (#3), range or high-low differences (#3), ratios (#4), rank ordering (#8), average (mean and median; #1, #4, #6, #8), correlation and regression (#5, #6), or confidence intervals (#8). In addition, phrases also relate, directly or indirectly, to core "big ideas" in statistics, such as variation and causes of variation (#5, #9), sampling (#5, #8), data reduction and aggregation (#5, #6), significance of difference (#7), and prediction (#5, #6). The many types of tables, charts and graphs that appear in summary documents cannot be illustrated here due to space limitations, but excerpt #7 illustrates that such displays may take non-standard or complex forms that force their creators to add rather dense explanations to make sure users can read them.

Often, several types of statistical ideas are combined in single text passages in ways that increase their complexity. For example, the single sentence in #4: "The highest ratios are found in Central and Western Africa, where the average (median) pupil/teacher ratio rose from an already high 50:1 in 1990 to 52:1 in 1998" requires that the reader can make sense of composite phrases that are seldom found in standard textbooks, such as "highest ratio" or "Average (median)". Texts may also refer to data management or data transformations, such as in the case of discussion of merging of datasets or imputation (#5). In addition, indices and composite variables of various degrees of complexity (#4, #5, #6) may be central to the text, requiring that users can make sense of second-order variables, i.e., that represent relationships derived from other relationships and ratios.

3.5. REPORTS

In addition to brief executive summaries described earlier, statistical agencies carry an assortment of long documents termed "reports". These publications include a detailed narrative text that discusses the background, methodology, findings, conclusions, and implications of a specific study or project. Some reports summarize an ongoing statistical activity, such as an analysis of trends or changes in certain indicators. In addition, reports also present numerous tables with detailed statistical information and graphical displays, and provide technical details that enable users to understand the limitations of the work or findings. Murray and Gal (2002) claimed that such publications are not expected to be consumed by decision makers and officials, given that these clients seek encapsulated information. Rather, reports are intended to inform other user groups, such as the support staffs who advise policy makers, or analysts and researchers who need detailed information.

Naming conventions and content vary

Multiple labels are used to describe reports across agencies. For example, Statistics-SA lists on its homepage, in addition to "Reports", "Publications", and "Working papers", users also subcategories such as "Discussion papers", "Occasional papers", and "Working papers", all of which may qualify as a report in some way. OECD lists over 20 documents types, such as "Case studies", "Country surveys", "Reviews", "Reports", "Policy Briefs", "Working papers", and others. A similar if somewhat narrower assortment can be found in other sites. Yet, no general explanation is provided on any of the sites as to the distinguishing features of overlapping entities. Document lists that appear on screen often present only document names, without additional explanations about the nature of the document. Users thus need to open documents and browse their introductory sections to figure out their actual content. With long reports, users may also have to be able to activate the "search" function of the program (e.g., Word, Acrobat Reader) they used to view the particular document in order to locate text of interest.

Reports summarizing statistical studies vary in scope, depth and length

Some reports reached upwards of 200 or 300 pages, and on occasion could not be viewed on-screen or downloaded and could only be purchased in a hardcopy or book format. A unique type of a report that appeared in all sites was an "annual statistical yearbook", a compendium of large amounts of statistical information covering the full scope of issues or areas under the jurisdiction of the agency. The information in such yearbooks (sometime carrying a different name, such as "The condition of education") often appeared with little interpretation of the data, though with some cautionary comments about data quality (see excerpt #8 in Table 3), and sometimes with a glossary or other texts explaining the meaning of some of the entities discussed in tables or graphs.

Report demand diverse and substantial statistical and mathematical knowledge

Since reports are normally longer and more comprehensive than executive summaries that summarize them (see 3.5 above and the excerpts in Table 3), reports make reference to a very wide range of ideas and concepts from descriptive and inferential statistics, combined with descriptions of data management techniques and computational procedures, such as those related to the creation of indices. However, reports do not necessarily explain the meaning or derivation of indicators and variables discussed or of various technical issues. For example, basic concepts in official statistics, such as *incidence* and *prevalence*, which are at the heart of understanding phenomena and trends related to health and well-being, were often used freely in reports. Likewise, technical comments in some reports refer to issues that are not addressed in introductory statistics courses, such as weighting of cases, dealing with missing data, standardization of values, and complex sampling schemes. Further, reports extensively use tables of varying degrees of density, including multi-page tables or tables with "nested" subdivisions, i.e., representing multiple variables by subgroupings on both the horizontal and vertical axes; see Mosenthal & Kirsch (1998).

3.6. AGGREGATE DATA

Murray and Gal (2002) argue that clients such as managers in the public and private sectors, or members of advocacy groups, need to have information on "local" issues they care about, and hence see data about subgroups, not the total population. All agencies indeed enable users to retrieve numerical information organized according to certain breakdowns, but data could be accessed in varied ways.

Users can view static tables that show statistical data for fixed subgroupings

In some agencies, users can choose tables from a long list of available titles; the row-and column summary data (e.g., counts, percents) that appear on screen seem to be a page from a published report or a yearbook. In other cases users do not see a long list of titles of tables, but rather encounter an interactive interface through which they define the table to be retrieved. For example, UNESCO-IS users can see data about specific geographical areas by picking an indicator (e.g., *Adult literacy*) and then choosing from lists of "countries" (or "regions") and "years". This form of retrieval just simplifies the interface – the tables that appear are pre-fabricated, just like above. In some cases users choose the data source, not only the type of aggregation. Users of Statistics-UK, for example, can access over 4000 different datasets pooled from various statistical agencies in the British Commonwealth via a facility called StatBase. Individual tables retrieved in this way allow users to reach information that otherwise they will be challenged to find in printed reports, though most data tables again seem to be pre-fabricated, i.e., as if they are tables from a former report.

Some agencies enable users to interactively construct new tables and displays

In addition to the ability to retrieve pre-fabricated tables with aggregate statistics for subgroups, several agencies allow users to customize displays to their specific needs, and create tables at different levels of sophistication. US-NCES users, for example, can build tables with school-related statistics by specifying table *rows* (e.g., the level of aggregation: state, county, district, school), *data category* (i.e., variables to be analyzed in the table cells, such as “enrollment by ethnicity”), and *column* (e.g., grade levels). Australian-IHW enables users to use a product called Data Cubes to compare multiple subgroups both through tables, graphs and charts; resulting displays can span a range of levels of complexity that sometimes exceed the capabilities of basic statistical packages. The implication is that agencies differ in how much flexibility they allow users, and increasing levels of interactivity seem to require more statistical and technical sophistication from users.

Agencies allow users to download data files for independent analysis

All agencies enable users to download complete datasets of aggregate statistics or time series that can be analyzed either with a spreadsheet program or various statistical packages. Given that these sub-products are meant to enable users to actively perform a statistical analysis of datasets, and that some of them can be accessed only for a fee, a review of these products goes beyond the scope of the present paper.

4. DISCUSSION

Statistics agencies are a key source for official statistical information about a broad spectrum of issues, and their products are valuable to politicians and decision makers, managers in the public and private sectors, advocacy groups, and citizens in general. Such products, however, have received little attention from both educators and researchers, although they have an important role in the information fabric of modern societies (European Commission, 1996), and enable the media to generate messages to the public (Podehl, 2002).

Given the lack of prior research in this area, this exploratory study was designed to describe the characteristics of five key products: Indicators, Press releases, Executive summaries, Reports, and Aggregate data, created by agencies of three types: National/central, Thematic, and International. All five product categories were indeed found on all six sites examined, usually with dozens and sometimes hundreds of instances in each category. This finding confirms Murray and Gal’s (2002) initial observation that statistical agencies publish product variants based on the same broad analyses, and incorporate multiple rhetorical levels to make sure that publications respond to the questions and concerns user groups may have.

As illustrated in the Results section, there is some overlap between the five product categories, for example when executive summaries are part of longer reports, when aggregate tables pertain to indicator data, or when press releases discuss details of an executive summary. Nonetheless, despite the existence of hybrid cases and variability within and across agencies, the five product categories appear distinct and each has specific characteristics. Each product category hence merits separate attention from researchers and educators.

The results show that Internet sites of statistics agencies vary in their organization and ease of use, and that the products created by such agencies range widely in their characteristics and content. The results of this study should be viewed with caution, however, as the purposeful sample used here did not include non-English speaking agencies, nor did it cover all existing agencies within some countries. It was not possible, due to lack of prior knowledge of the organization of products on each site, to construct a probability sample of items (i.e., product instances) within each agency. That said, agencies addressing a wide range of themes were selected, and multiple items from different topics in all product categories were chosen within each agency to increase representativeness. With these caveats in mind, conclusions are discussed below with regard to the skills needed for statistical literacy and to future research and educational practices.

4.1. SKILLS DEMANDS

A model proposed by Gal (2002a) assumed that statistically literate behavior requires the joint activation of five knowledge bases (literacy, statistical, mathematical, world, and critical) and of a dispositional component (involving critical stance, and beliefs and attitudes). This study highlights the importance and multi-faceted nature of the "literacy" aspect of statistical literacy, both in terms of general literacy as well as in terms of "information technology and computer literacy" (Dede, 2000). Equally importantly, the breadth and depth of issues addressed in the five product categories implies that clients of statistics agencies need to possess and activate diverse statistical and mathematical knowledge.

Literacy-related knowledge and skills

Statistical literacy was broadly defined as the ability to interpret, critically evaluate, and express one's opinions about statistical information and data-based messages. This study shows that statistical literacy can also involve the ability to *access, define, locate, extract, and filter* needed information in a complex array of information products. Clients must be familiar with the expected content of typical products, and recognize the differences between press releases, executive summaries, full reports, or indicator and aggregate data. They have to be able to plan and combine flexible search methods and adjust search strategies and keywords to sometimes inconsistent product organization and naming conventions. When reaching potentially useful documents, users have to be aware of the mediating role and basic features of computer programs used to view or open relevant files. When viewing a document, users have to employ different reading strategies given the diversity found within and across sites in document content, style, and terminology. Users may have to fit their reading strategy to the nature of the document examined, such as by skimming large reports to identify needed segments, or by activating critical reading skills when reading interpretive statements in press releases (Thistlewaite, 1990). Users also need to possess adequate *document literacy* (Mosenthal & Kirsch, 1998) and *graph interpretation skills* (Bright & Friel, 1998) to be able to analyze complex tables, graphs and charts found in all product types.

Even if users have adequate levels of *both* general literacy and information technology skills, they could be helped by possessing complementary knowledge of the specific characteristics and organization of the products of statistics agencies and the environment within which they can be found. For example, clients need to be aware of the varied terminology used by statistics agencies to name product categories and subtypes, or specific indicators and variables and of differences between agencies in this regard, but at the same time be aware of *common* terms or concepts used across agencies. Clients also need to possess knowledge and skills that enable them to effectively navigate, search, use help systems and glossaries, and browse diverse types of lists and documents. Finally, clients need to be able to comfortably switch between interfaces and naming conventions both within and across agencies, and activate site-specific programs or tools for retrieving tables or constructing customized displays.

Statistics-related knowledge and skills

Documents such as press releases, executive summaries, and reports were found to cover a broad spectrum of statistical topics, and carry a wide range of tables, graphs and charts at different levels of complexity. Press releases presented some surprises: they could be expected to require lower level of statistical sophistication from readers, as they are supposed to consist of non-technical narratives that are within the grasp of politicians and the general public (Podehl, 2002). Yet, some press releases employ statistical concepts and quantitative statements and arguments that appear quite demanding or present interpretation challenges, even though they are written in a relatively informal language.

As the various examples in the Results section illustrated, statistical terms and findings in executive summaries and full reports are embedded in rich and sometimes complex text passages. Readers have to consider explanatory variables and contextual factors, and grasp comparisons and trends across time and subgroups, taking into account the context of the study. Such products also present users with "qualifying" information, i.e., regarding the limitations imposed by data-gathering processes on the quality of the data and conclusions, or the relative confidence that can be placed in inferences or conclusions, for example due to characteristics of the samples or instruments. This means that users have to recognize the significance of such qualifying information and use it to inform their interpretation of what is being read (Utts, 1996). Qualifying information or details about a study's context and methods, however, are not always available or easy to find, especially when viewing stand-alone tables of aggregate data or displays of indicators, which often appear without much supporting narrative text. Indeed, many publications do not fully explain the meaning or computation of standard indicators or variables, probably on the assumption that the primary target audiences of statistics agencies, policy makers and officials, are familiar with the standard tools and concepts of the trade.

Active aspects of statistical literacy

A view of adults in general as "passive" consumers of statistical messages (Gal, 2002a) may be insufficient once information products from statistics agencies are considered as a legitimate part of the information landscape that all citizens may have to deal with. This study shows that "consumption" of such products requires clients to take a more active role than traditionally implied with regard to the consumption of media-based statistical messages. Of course, users have had to be able to define their information needs and the level of detail they desire even with traditional printed products, and have had to activate critical reading skills when reading any type of printed publication. Yet, the possibilities afforded by the Internet for quick and instant interaction with multiple types of products and items present new opportunities but also demand additional, more active involvement than before from clients of statistics agencies.

After defining their information needs and retrieving and reading documents deemed relevant (e.g., a press release, an executive summary) users may realize that more detailed information is needed. This may result in cycles of search, retrieval and analysis of additional items, or a decision to generate views of data for specific subgroups of the user's choice, through interaction with web-based facilities for accessing indicators or aggregate data. These stages require that users can re-examine their information needs and adjust or extend their actions, in light of their evaluation of the quality of the information they read so far. The notion of critical evaluation, though, implies a form of action, not just passive interpretation or understanding. As Gal (2002a; 2002b) argues, for any action to occur (internally or overtly), certain dispositions need to be in place. A statistically literate or numerate person has to possess positive attitudes and beliefs, such as a belief in the legitimacy of questioning information from official sources, before he or she will be willing to invest energy in obtaining the information needed or feel comfortable to activate critical questions to evaluate this information.

Future skill demands

Products of statistics agencies and the information space within which users have to find such products (i.e., the website with all its elements and hierarchy of webpages, screens, and items that can be viewed or used) are not static entities that will remain unchanged for years to come. It can be assumed that cumulative wisdom is developing among website designers in general and within statistics agencies as well, in light of feedback and suggestions from various users and professionals. Such knowledge, together with improvements in Internet technology, may enable statistics agencies to improve upon some of the design problems noted in this study, such as by enhancing the features of online help systems and search engines, by simplifying and expanding glossaries, or by redesigning the structure of websites and reorganizing document groupings. However, core skills demands associated with comprehension of information in products of statistics agencies cannot be expected to lessen, given the ever-present need to read and comprehend texts that combine statistical terms and tabular and graphical displays in diverse and sometimes complex documents.

While some aspects of Internet sites of statistics agencies may become simpler to use, it can be conjectured that such sites will gradually evolve into more complex entities. Agencies will have to continue and make accessible existing "legacy" products, some of which were shown in this study to be problematic, but will continue to add each year many new items in all products categories, given the accumulation of new data and findings. Also, to better serve the broadening information needs of a spectrum of users whose average educational level is rising, agencies may offer users more services or product subtypes, including the ability to interactively tailor data views or aggregations to users' specific needs. Thus, even if agencies improve retrieval processes and re-organize sites and products, future users may need to possess broader and more advanced skills in order to extract maximum value from products of statistics agencies.

Summary

The picture emerging from the above analysis is that the statistical (and mathematical) knowledge base needed to comprehend and critically evaluate various products can be quite substantial. Clients of statistics agencies will often need relatively *formal* knowledge of the meaning of concepts and underlying statistical ideas. Some of the needed knowledge falls outside the material normally included in introductory-level statistics instruction, and will not be familiar to all those students or adults whose exposure to statistical topics is based only on seeing statistics in use in the media, or on exposure to procedural or computational aspects of statistics (Cobb, 1992). It follows that in conceptualizing the skills needed for statistically literate behavior there may be a need to think of subtypes, depending on the nature of the real-world "target stimuli" or "target tasks" that adults face. The present study points to the need to consider "official-statistics literacy" as a subtype that places unique demands on strong integration of relatively formal statistical and mathematical knowledge with both general literacy and information technology skills, coupled with specific knowledge regarding the types of products, document structures, terminology, and interfaces that comprise the information space presented by statistic agencies to their clients.

5. IMPLICATIONS FOR RESEARCH AND EDUCATION

The use of a qualitative approach is well-accepted as a first step in exploratory research. The analysis of the six case studies selected in this study, while of limited generalizability, can inform further research and contribute to clarification of educational practices that can promote the statistical literacy of students and adults from all walks of life.

Product characteristics

This study sketched in broad strokes the characteristics of key categories of product from statistics agencies, but further inquiries are needed to specify and quantify the characteristics of such products in a broader range of agencies. For example, it would be useful to examine the relative frequency of specific mathematical and statistical entities (e.g., percents, average) or of different types of tables, graphs and charts that appear in specific product categories, and compare to the frequency of such entities in regular media articles that refer to statistics issues (Joram, Resnick, & Gabriele, 1995). Similarly, it would be important to identify and describe linguistic structures and writing styles that more commonly appear in different product categories, to better understand the shift in skills required to comprehend each one (e.g., press releases as opposed to more formal executive summaries). Research along these lines can clarify the differences between the separate but related target stimuli on which educational efforts should focus when attempting to develop different subtypes of statistics literacy.

Knowledge and skill levels

Little information exists about the extent to which adults in general, members of specific occupational groups, or students in academic institutions or in high-schools, possess the required skills (in literacy, information technology, statistical, and mathematical areas) described above as necessary for accessing, filtering, comprehending and critically evaluating the information in products from statistics agencies. National and international comparative studies aimed at school students (e.g., TIMSS and PISA, see www.oecd.org) and adults (e.g., IALS and ALL, see www.ets.org/all) assess only a subset of the essential skills described earlier, and thus can provide only partial information in this regard. Studies are needed that will examine students' and adults' actual ability to effectively engage the interactive environment found on Internet sites of statistics agencies. Through a combination of qualitative and quantitative techniques it should be possible to analyze interpretation and thinking processes of clients who possess different skill levels, and identify gaps in comprehension, performance, or ability to critically analyze actual texts or displays in different product categories. Such studies have to be designed with the understanding that knowledge bases underlying statistical literacy, and hence manifested performance on functional assessment tasks, can exist on multiple levels, from informal and rudimentary to more formal and advanced (Gal, 2002a; Watson, 1997; Watson & Moritz, 2000). Research on the above areas can contribute to the specification and understanding of the skills and abilities that underlie statistically literate behavior, and inform educational activities.

Educational activities

An interest in the ability of all segments of the population to act as informed and empowered citizens requires attention not only to people's ability to interpret and critically evaluate statistics in the media, but also to their ability to access, locate, filter, comprehend, and evaluate information products from statistics agencies. The skills involved in the latter, however, seem broader, more formal, and at times more advanced than those required for interpreting and evaluating statistical messages in newspapers or advertisements. Clients of statistics agencies have to *activate in integration* a network of skills and knowledge bases, but these are often *separated* in traditional modes of instruction which are organized around distinct subject areas, such as statistics, mathematics, language arts, or computer skills. To develop "official-statistics literacy", the facet of statistical literacy that pertains to official statistical information, it will be necessary to create, in both teaching and assessment, opportunities for students to apply skills in the context of realistic, socially meaningful, and motivating tasks. This requires that educators employ authentic products from statistic agencies, and implement teaching methods that focus on the actual skill demands imposed by products of statistics agencies.

6. REFERENCES

Bright, G. W. & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on Statistics: Learning, Teaching, and Assessment in Grades K-12* (pp. 63-88). Mahwah, NJ: Lawrence Erlbaum.

Cerrito, P. B. (1999). Teaching statistical literacy. *College Teaching,* 47(1), 1-7.

Cobb, G. W. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the Call for Change: Suggestions for Curricular Action* (pp. 3-43). Washington, D.C.: Mathematical Association of America.

Cohn, V. & Cope, L. (2001). *News and Numbers : A Guide to Reporting Statistical Claims and Controversies in Health and Other Fields*. Iowa City: Iowa State University Press.

Crossen, C. (1994). *Tainted Truth: The Manipulation of Fact in America*. New York: Simon & Schuster.

Dede, C. (2000). Emerging influences of information technology on school curriculum. *Journal of Curriculum Studies,* 32(2), 281-303.

European Commission. (1996). *White paper on education and training: Teaching and learning-towards the learning society*. Luxembourg: Office for official publications of the European Commission.

Frankenstein, M. (1990). Incorporating race, gender, and class issues into a critical mathematical literacy curriculum. *Journal of Negro Education,* 59(3), 336-347.

Franklin, B. (Ed.) (1999). *Social Policy, the Media and Misrepresentation*. London: Routledge.

Gal, I. (2000). The numeracy challenge. In I. Gal (Ed.), *Adult Numeracy Development: Theory, Research, Practice* (pp. 1-25). Cresskill, NJ: Hampton Press.

Gal, I. (2002a). Adult statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.

Gal, I. (2002b). Dispositional aspects of coping with interpretive numeracy tasks. *Literacy and Numeracy Studies,* 12(2), 47-61.

Gelman, A., Nolan, D., Men, A., Warmerdam, S. & Bautista, M. (1998). Student projects on statistical literacy and the media. *The American Statistician,* 52(2), 160-166.

Joram, E., Resnick, L. & Gabriele, A. J. (1995). Numeracy as a cultural practice: An examination of numbers in magazines for children, teenagers, and adults. *Journal for Research in Mathematics Education*, 26(4), 346-361.

Lehohla, P. (2002). Promoting statistical literacy: A South African perspective. In B. Phillips, (Ed.), *Proceedings of the Sixth International Conferences on Teaching Statistics*. Voorburg, the Netherlands: International Statistical Institute. CD ROM.

Moreno, J. L. (2002). Toward a statistically literate citizenry: What statistics everyone should know. In B. Phillips, (Ed). *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, the Netherlands: International Statistical Institute. CD ROM.

Mosenthal, P. B. & Kirsch, I. S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. *Journal of Adolescent and Adult Literacy,* 41(8), 638-657.

Murray, S. & Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. In B. Phillips, (Ed). *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, the Netherlands: International Statistical Institute. CD ROM.

Phillips, B. (2002) (Ed.), *Proceedings of the Sixth International Congress on Teaching Statistics*. Voorburg, the Netherlands: International Statistical Institute. CD ROM.

Podehl, W. M. (2002). Statistical literacy and the media. In B. Phillips, (Ed). *Proceedings of the Sixth International Congress on Teaching Statistics*. Voorburg, the Netherlands: International Statistical Institute. CD ROM.

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

Thistlewaite, L. L. (1990). Critical reading for at-risk students. *Journal of Reading,* 33(8), 586-593.

Utts, J. M. (1996). *Seeing through statistics*. Belmont, CA: Wadsworth.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association,* 88, 1-8.

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

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

IDDO GAL
Department of Human Services
University of Haifa
Eshkol Tower, Room 718
Haifa 31905, Israel

ASSESSING STATISTICAL REASONING

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

SUMMARY

This paper begins with a discussion of the nature of statistical reasoning, and then describes the development and validation of the Statistical Reasoning Assessment (SRA), an instrument consisting of 20 multiple-choice items involving probability and statistics concepts. Each item offers several choices of responses, both correct and incorrect, which include statements of reasoning explaining the rationale for a particular choice. Students are instructed to select the response that best matches their own thinking about each problem. The SRA provides 16 scores which indicate the level of students' correct reasoning in eight different areas and the extent of their incorrect reasoning in eight related areas. Results are presented of a cross-cultural study using the SRA to compare the reasoning of males and females in two countries.

Keywords: Statistics education research; Assessment; Reasoning; Misconceptions

1. INTRODUCTION

Two reform movements have been affecting the teaching and learning of statistics at all educational levels. One reform has focused on content and pedagogy, shifting the focus from computation and procedures to an emphasis on statistical reasoning and thinking (Moore, 1997). A second reform is in the area of student assessment, focusing on better alignment of instruction with important learning goals, and using assessment as a tool to improve student learning (Garfield, 1993, Garfield & Gal, 1999a, Chance & Garfield, 2002). Some of the statements identified with the assessment reform include:

- Traditional forms of assessment (e.g., multiple-choice exams) are too narrow to provide sufficient information about student learning.
- Providing single numbers or letters to students to represent their learning is inadequate and does not promote successful learning.
- Alternative types of assessment are needed, used in combination, and aligned with instructional and curricular goals.

While elementary and secondary schools have embraced new assessment methods such as portfolios and performance tasks, college courses still rely primarily on more traditional tests and exams (Garfield, Hogg, Schau & Whittinghill, 2002). Traditional assessments of statistical knowledge typically look like textbook problems that either rely heavily on numerical calculations or on the ability to recall isolated pieces of information. Although this type of assessment may succeed in providing instructors with a method for determining letter grades, these types of assessment rarely reveal information about how students understand

Statistics Education Research Journal 2(1), 22-38, http://fehps.une.edu.au/serj

and reason with statistical ideas or apply their knowledge to solve statistical problems. Garfield and Chance (2001) describe many alternatives to traditional assessment methods. One encouraging trend is the increased number of instructors who are using performance assessments in the form of student projects. However, many instructors of large classes have been hesitant about introducing this type of alternative assessment.

New forms of assessment are needed not only to provide information to students and instructors, but also to use in research on teaching and learning statistics, to evaluate the effectiveness of different curricula or pedagogical approaches, and to explore the development of statistical reasoning. Although student outcomes may best be assessed in personal interviews or via in-depth student work such as projects, there is a practical need to have an easily scorable instrument that captures students' thinking, reasoning, and application of knowledge, rather than a test where students "tell" the teacher what they have remembered or show that they can perform calculations or carry out procedures correctly.

2. THE NATURE OF STATISTICAL REASONING

Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information (Garfield & Chance, 2000). This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Much of statistical reasoning combines ideas about data and chance, which leads to making inferences and interpreting statistical results. Underlying this reasoning is a conceptual understanding of important ideas, such as distribution, centre, spread, association, uncertainty, randomness, and sampling.

Many people think of mathematics and statistics as the same thing, and therefore, confuse statistical reasoning with mathematical reasoning (Garfield & Gal, 1999b). Today's leading statistical educators view these disciplines and types of reasoning as quite distinct. Gal and Garfield (1997) distinguish between the two disciplines in the following ways:

- In statistics, data are viewed as numbers with a context. The context motivates procedures and is the source of meaning and basis for interpretation of results of such activities.
- The indeterminacy or "messiness" of data distinguishes statistical investigations from the more precise, finite nature characterizing mathematical explorations.
- Mathematical concepts and procedures are used as part of the solution of statistical problems. However, the need for accurate application of computations is rapidly being replaced by the need for selective, thoughtful, and accurate use of technological tools and increasingly more sophisticated software programs.
- Many statistical problems do not have a single mathematical solution, but instead, start with a question and result in an opinion supported by certain findings and assumptions. These answers need to be evaluated in terms of quality of reasoning, adequacy of methods employed, and nature of data and evidence used.

In recent years there has been an appropriate shift from traditional views of teaching statistics as a mathematical topic (with an emphasis on computations, formulas, and procedures) to the current view that distinguishes between mathematics and statistics as separate disciplines. As Moore (1992) argues, statistics is a mathematical science but is not a branch of mathematics, and has clearly emerged as a discipline in its own right, with characteristic modes of thinking that are more fundamental than either specific methods or mathematical theory.

As statistics is being distinguished from mathematics, so is statistical reasoning being distinguished from mathematical reasoning. DelMas (in press) provides an analysis and comparison of these two types of reasoning. He argues that while mathematical and statistical reasoning can be distinguished with respect to the content of reasoning (abstract versus contextual), certain statistical content has an abstract nature that proves difficult for students. He reviews the relevant research on reasoning and uses these findings to identify areas of statistical reasoning that students find most challenging. At the current time, studies on particular aspects of statistical reasoning are still in the early stages of identifying the development of different types of reasoning and how reasoning may be affected by particular teaching activities and technological tools (for a collection of these studies, see BenZvi &Garfield, in press and the STRL-2 summaries published in SERJ 1(1)). Despite a substantial body of knowledge on how to effectively promote statistical reasoning in students, a primary goal of statistics education is to enable students to produce reasoned descriptions, judgments, inferences, and opinions about data. Current mathematics curricula for students in elementary and secondary schools are being written or revised to help students comprehend and deal with uncertainty, variability, and statistical information in the world around them. This emphasis in statistical education on developing statistical reasoning illustrates the need for good methods to assess students' statistical reasoning.

2.1. ASSESSING STATISTICAL REASONING

Most assessment instruments used in research studies of statistical reasoning and understanding consist of items given to students or adults individually as part of clinical interviews or in small groups which are closely observed. Most paper-and-pencil assessment instruments focus on computational skills or problem solving, rather than on reasoning and understanding.

Traditional test questions involving statistical content often lack appropriate context and tend to focus on accuracy of statistical computations, correct application of formulas, or correctness of graphs and charts. Questions and task formats that culminate in simple "right or wrong" answers do not adequately reflect the nature of students' thinking and problem solving, and therefore provide only limited information about students' statistical reasoning processes and their ability to construct or interpret statistical arguments (Gal & Garfield, 1997).

Although statistical reasoning may best be assessed through one-to-one communication with students (e.g., interviews or observations) or by examining a sample of detailed, in-depth student work (e.g., a statistical project), carefully designed paper-and-pencil instruments can be used to gather some limited indicators of students reasoning. One such instrument is *The Statistical Reasoning Assessment* (SRA).

The SRA was developed and validated as part of the NSF-funded ChancePlus Project (Konold, 1990; Garfield, 1991), to use in evaluating the effectiveness of a new statistics curriculum for high school students in achieving its learning goals. At that time, no other instrument existed that would assess high school students' ability to understand statistical concepts and apply statistical reasoning.

The SRA is a multiple-choice test consisting of 20 items. Each item describes a statistics or probability problem and offers several choices of responses, both correct and incorrect. Most responses include a statement of reasoning, explaining the rationale for a particular choice. Students are instructed to select the response that best matches their own thinking about each problem. The SRA has been used not only with the ChancePlus project but also with other high school and college students in a variety of statistics courses, to evaluate the effectiveness of curricular materials and approaches as well as to describe the level of students' statistical reasoning. Items from this instrument have been adapted and used in research projects in other English-speaking countries such as Australia and the United Kingdom.

2.2. STATISTICAL REASONING GOALS FOR STUDENTS

The first step in developing or considering an assessment of statistical reasoning is to clarify the types of reasoning skills students should develop. The following types of reasoning were used to develop and select items to use in the SRA. These topics seemed appropriate for the purpose of the instrument, which was to be used with secondary school students who had been learning the basic techniques of data analysis as part of the ChancePlus Project described above.

Reasoning about data: Recognizing or categorizing data as quantitative or qualitative, discrete or continuous; and knowing how the type of data leads to a particular type of table, graph, or statistical measure.

Reasoning about representations of data: Understanding the way in which a plot is meant to represent a sample, understanding how to read and interpret a graph and knowing how to modify a graph to better represent a data set; being able to see beyond random artifacts in a distribution to recognize general characteristics such as shape, centre and spread.

These two types of reasoning (about data and representation) were not linked to specific items in the SRA but are related to many of the items assessing reasoning. The following types of reasoning are linked to particular items in the SRA as shown in Table 1.

Reasoning about statistical measures: Understanding what measures of centre, spread, and position tell about a data set; knowing which are best to use under different conditions, and how they do or do not represent a data set; knowing that using summaries for predictions will be more accurate for large samples than for small samples; knowing that a good summary of data includes a measure of centre as well as a measure of spread and that summaries of centre and spread can be useful for comparing data sets.

Reasoning about uncertainty: Understanding and using ideas of randomness, chance, likelihood to make judgments about uncertain events; knowing that not all outcomes are equally likely; knowing how to determine the likelihood of different events using an appropriate method (such as a probability tree diagram or a simulation using coins or a computer program).

Reasoning about samples: Knowing how samples are related to a population and what may be inferred from a sample; knowing that a larger, well chosen sample will more accurately represent a population and that there are ways of choosing a sample that make it unrepresentative of the population; being cautious when making inferences made on small or biased samples.

Reasoning about association: Knowing how to judge and interpret a relationship between two variables; knowing how to examine and interpret a two-way table or scatter plot when considering a bivariate relationship; knowing that a strong correlation between two variables does not mean that one causes the other.

3. INCORRECT STATISTICAL REASONING

In addition to determining what types of reasoning skills students should develop, it was also important to identify the types of incorrect reasoning students should not use when analysing statistical information. Kahneman, Slovic, and Tversky (1982) are well known for their substantial body of research that reveals some prevalent ways of thinking about statistics that are inconsistent with a technical understanding. Their research suggests that even people who can correctly compute probabilities tend to apply faulty reasoning when asked to make an inference or judgment about an uncertain event, relying on incorrect intuitions (Garfield & Ahlgren, 1988, Shaughnessy, 1992). Other researchers have discovered additional misconceptions or errors of reasoning when examining students in classroom settings (e.g., Konold, 1989; Lecoutre, 1992). Several of the identified misconceptions or errors in reasoning, used to develop the SRA, are described below:

Misconceptions involving averages: Averages are the most common number, to find an average one must always add up all the numbers and divide by the number of data values (regardless of outliers), a mean is the same thing as a median, and one should always compare groups by focusing exclusively on the difference in their averages.

The Outcome orientation: An intuitive model of probability that leads students to make yes or no decisions about single events rather than looking at the series of events (Konold, 1989). For example: A weather forecaster predicts the chance of rain to be 70% for 10 days. On 7 of those 10 days it actually rained. How good were his forecasts? Many students will say that the forecaster didn't do such a good job, because it should have rained on all days on which he gave a 70% chance of rain. They appear to focus on outcomes of single events rather than being able to look at series of events-70% chance of rain means that it should rain. Similarly, a forecast of 30% rain would mean it won't rain.

Good samples have to represent a high percentage of the population: It does not matter how large a sample is or how well it was chosen, it must represent a large percentage of a population to be a good sample.

The Law of small numbers: Small samples should resemble the populations from which they are sampled, so small samples are used as a basis for inference and generalizations (Kahneman, Slovic, & Tversky, 1982).

The Representativeness misconception: People estimate the likelihood of a sample based on how closely it resembles the population. Therefore, a sample of coin tosses that has an even mix of heads and tails is judged more likely than a sample with more heads and fewer tails (Kahneman, Slovic, & Tversky, 1982).

The Equiprobability bias: Events tend to be viewed as equally likely. Therefore, the chances of getting different outcomes (e.g., three fives or one five on three rolls of a dice) are incorrectly viewed as equally likely events (Lecoutre, 1992).

4. VALIDITY AND RELIABILITY ANALYSES

Once items had been written, borrowed, or adapted, to represent areas of correct and incorrect reasoning, all items went through a long revision process. The first step of this process was to distribute items to "experts" for content validation, to determine if each item was measuring the specified concept or reasoning skills, and to elicit suggestions for revisions or addition of new items. A second step was to administer items to groups of students and to investigate their responses to open-ended questions. These responses were used to phrase justifications of selected responses to use in a subsequent multiple-choice format in the instrument. After several pilot tests of the SRA followed by administration of the instrument in different settings, and after many subsequent revisions, the current version was created. A copy of the instrument is attached to this paper.

Table 1. Correct Reasoning Skills and Misconceptions Measured by the SRA and the Corresponding Items and Alternatives for Measuring Each Conception and Misconception

Correct Reasoning Skills	Items and Alternatives
1. Correctly interprets probabilities	2d, 3d
2. Understands how to select an appropriate average	1d, 4ab, 17c
3. Correctly computes probability	
a. Understands probabilities as ratios	8c
b. Uses combinatorial reasoning	13a, 18b, 19a, 20b
4. Understands independence	9e, 10df, 11e
5. Understands sampling variability	14b, 15d
6. Distinguishes between correlation and causation	16c
7. Correctly interprets two-way tables	5,1d
8. Understands importance of large samples	6b, 12b
Misconceptions	
1. Misconceptions involving averages	
a. Averages are the most common number	1a, 17e
b. Fails to take outliers into consideration when computing the mean	1c
c. Compares groups based on their averages	15bf
d. Confuses mean with median	17a
2. Outcome orientation misconception	2e, 3ab, 11abd, 12c, 13b
3. Good samples have to represent a high percentage of the population	7bc, 16ad
4. Law of small numbers	12a, 14c
5. Representativeness misconception	9abd, 10e, 11c
6. Correlation implies causation	16be
7. Equiprobability bias	13c, 18a, 19d, 20d
8. Groups can only be compared if they are the same size	6a

An attempt was made to determine criterion-related validity, by administering the SRA to students at the end of an introductory statistics course and correlating their scores with different course outcomes (e.g., final score, project score, quiz total, etc.). The resulting correlations were all extremely low, suggesting that statistical reasoning and misconceptions are unrelated to students' performance in a first statistics course.

In order to determine the reliability of the SRA, different reliability coefficients were examined. An analysis of internal consistency reliability coefficients indicated that the intercorrelations between items were quite low and that items did not appear to be measuring one trait or ability. A test-retest reliability coefficient appeared to be a more appropriate method to use, but first a new scoring method was needed.

Although individual items could be scored as correct or incorrect and total correct scores could be obtained, this single number summary seemed uninformative and did not adequately reflect students' reasoning abilities. Therefore, a method was created where each response to an item was viewed as identifying a correct or incorrect type of reasoning. Eight categories or scales of correct reasoning were created and eight categories of incorrect reasoning were also developed (see Table 1). One item (number 7) was not included in a correct reasoning scale because it seemed to be primarily effective in assessing misconceptions. Scores for each scale range from 2 to 8, depending on how many responses contribute to that scale. In addition to the 16 scale scores, total scores for correct and incorrect reasoning may be calculated by adding the 8 scale scores. A new set of data was gathered from a group of 32 students enrolled in an assessment course for preservice teachers. These students took the SRA and one week later took the same test again. A test-retest reliability analysis yielded a reliability of .70 for the correct total score and .75 for the incorrect reasoning scores (Liu, 1998).

5. CROSS CULTURAL STUDY

Once an appropriate scoring method was developed for the SRA, the instrument was used in cross-cultural study. Liu (1998) used the SRA to determine if gender differences exist in large samples of college students in the USA and in Taiwan. In this study, the SRA was administered to 267 subjects at the University of Iowa, and a translated, Chinese, version of the SRA was administered to 144 subjects at Cheng-Chi University and 101 subjects at Feng-Chia University in Taiwan. All students were tested at the end of an introductory course in business statistics. The students were of comparable ages and had no prior instruction in statistics. However, in the two Taiwan samples there were higher percentages of females (60% and 74%) than in the Iowa sample (43%). The first analyses compared scale scores for students in each country, as shown in Table 2.

Table 2. Comparison of scaled scores on SRA for Taiwan and Iowa Students (scale = 0 to 2 points)

	Taiwan (n=245)	Iowa (n=267)
Correct Reasoning Scales		
1. Correctly interprets probabilities	1.36	1.35
2. Understands how to select an appropriate average	1.19	1.22
3. Correctly computes probability	.90	.91
4. Understands independence	1.47	1.25
5. Understands sampling variability	.46	.44
6. Distinguishes between correlation and causation	1.30	1.31
7. Correctly interprets two-way tables	1.57	1.30
8. Understands importance of large samples	1.52	1.35
Misconceptions Scales		
1. Misconceptions involving averages	.43	.59
2. Outcome orientation (misconception)	.43	.45
3. Good samples have to represent a high percentage of the population	.18	.18
4. Law of small numbers	.68	.58
5. Representativeness misconception	.21	.34
6. Correlation implies causation	.20	.20
7. Equiprobability bias	1.12	1.12
8. Groups can only be compared if they are the same size	.78	1.20

Because each scale could have a different number of points, all scales were divided by the number of items to yield scores on a scale of 0 to 2. These scaled scores suggested that there are strong similarities in reasoning for the two samples of students in spite of cultural differences. These scores also show the types of reasoning that are most difficult for students (e.g., sampling variability, probability) and the misconceptions that are most prevalent (e.g., equiprobability bias).

The second set of analyses investigated gender differences, using total correct reasoning scores and total misconception scores. Table 3 presents two-way analysis of variance results for the total correct reasoning scores by country and gender, which indicates that the country effect is highly significant. As shown in Figure 1, students in Taiwan have higher correct reasoning scores than the students in the United States. While the male sample in Taiwan has higher scores than the females, in the United States males and females have more similar scores. Both the gender effect and the interaction effect between country and gender are not significant.

Table 3. Analysis of Variance Results for Total Correct Reasoning Scores by Country and Sex

Source of Variation	DF	Mean Square	F	F-prob
Country	1	307.268	13.945	<.001 **
Sex	1	69.051	3.134	.077
Country x Sex	1	73.028	3.314	.069
Error	508	22.035		

** significant at the alpha = .01 level

Table 4. Cell Means and Standard Deviations of Total Correct Reasoning Scores for Males and Females in Each Country

	Taiwan	United States
Male	22.90 (4.83)	20.55 (4.59)
Female	21.38 (4.80)	20.57 (4.58)

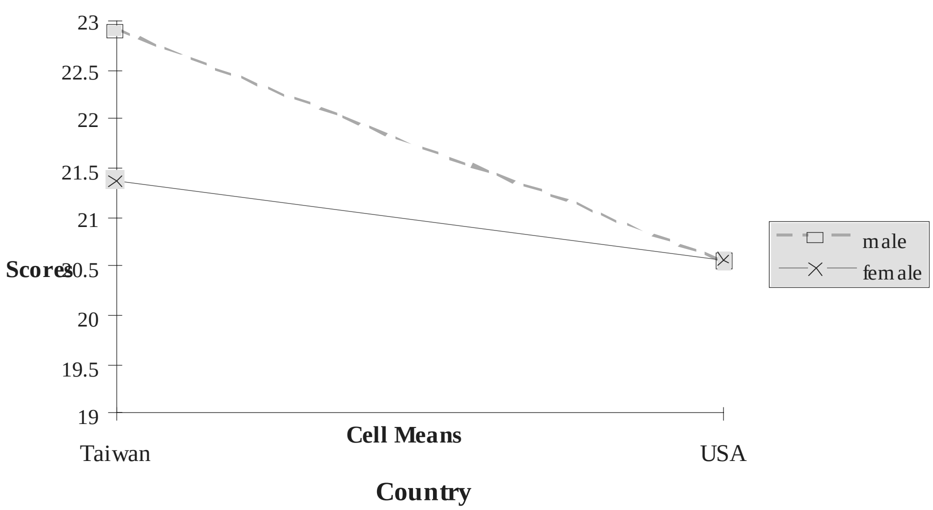

Figure 1: Cell Means of Total Correct Reasoning Scores for Each Sex by Two Countries

The ANOVA results for the total misconception scores by country and sex is presented in Table 5. Both country and sex effects are significant. Students in Taiwan have significantly lower misconception scores than students in the United States. As also shown in Figure 2, the female samples have significantly higher misconception scores than their male counterparts.

Table 5. Analysis of Variance Results for Misconception Total Scores by Country and Sex

Source	of Variation	DF	Mean Square	F	F-prob
Country	1	145.424	9.126	.003 **	
Sex	1	129.701	8.139	.005 **	
Country x Sex		1	31.258	1.962	.162
Error	508	15.935			

** significant at the alpha = .01 level

Table 6. Cell Means of Total Misconception Scores for Males and Females in Each Country

	Taiwan	United States
Male	11.28 (4.42)	12.87 (4.05)
Female	12.81 (3.53)	13.39 (4.11)

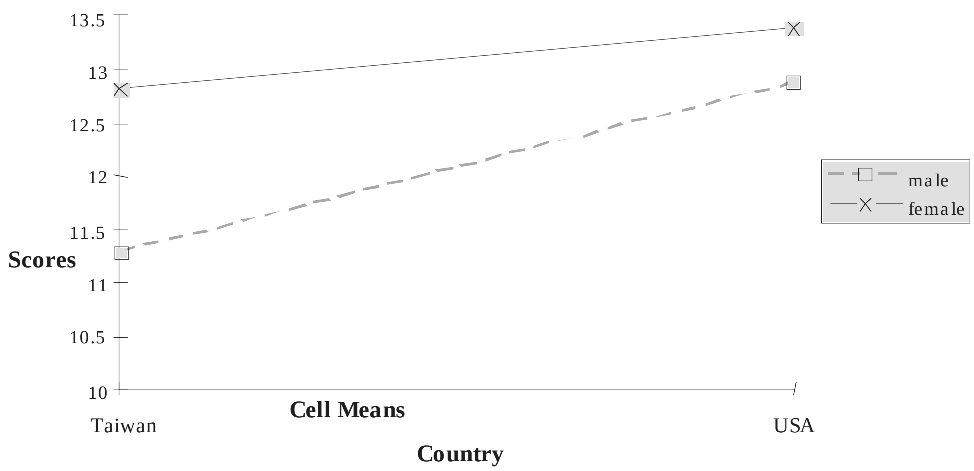

Figure 2: Cell Means of Total Misconception Scores for Each Sex by Two Countries

Liu concluded that based on her samples, males have higher total correct reasoning scores and lower total misconception scores than their female counterparts. Results were more striking in the Taiwan sample than the US sample. It is interesting to see that despite the seemingly similar scale scores for the students in the two countries, that there are actually striking differences when comparing the male and female groups. However, it is important to note that these non-random samples are not equivalent, and the results should be interpreted with caution. Nevertheless, it will be interesting to see if replications of this study in other countries will yield similar results. The SRA is currently being administered in other countries and similar comparisons will be useful for comparison purposes.

6. SUMMARY

Although there is a growing emphasis on developing students' statistical reasoning, assessing statistical reasoning remains a challenging task, and one that needs more attention in the research literature. Although the SRA is an easy to administer, paper-and-pencil instrument that provides some useful information regarding the reasoning of students, it is nonetheless an imperfect research and evaluation tool. The 16 scales represent only a small subset of reasoning skills and strategies, and attempts to establish the reliability and validity have raised new issues and yielded incomplete results. Konold (2003) is currently working with other researchers to establish an improved set of items to assess students' statistical reasoning that should be available in the near future, and include some of the original SRA items in a revised format. In addition, a new Web-based assessment resource for teachers of statistics is collecting and developing a large item bank for assessing many aspects of statistical reasoning and thinking (Garfield, delMas and Chance, 2003). Despite these two efforts, there is still ample room for more studies that develop new assessments of statistical reasoning, as well as studies that investigate or build on current instruments and items. A set of valid and reliable instruments will be of great use both to teachers and researchers who want to evaluate students' statistical reasoning.

REFERENCES

BenZvi, D. & Garfield, J. (In press). *The Challenge of Developing Statistical Reasoning, Literacy, and Thinking*. Dordrecht: Kluwer.

delMas, R. (In press). A comparison of mathematical and statistical reasoning. In D. BenZvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Reasoning, Literacy, and Thinking*. Dordrecht: Kluwer.

Chance, B. L. & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, 1(2), 38-44.

Gal, I. (1995). Statistical tools and statistical literacy: The case of the average. *Teaching Statistics*, 17 (3), 97-99.

Gal, I. & Garfield, J. (Eds.) (1997). *The Assessment Challenge in Statistics Education*. Amsterdam: IOS Press.

Garfield, J. (1991). Evaluating students' understanding of statistics: Development of the statistical reasoning assessment. In *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Volume 2* (pp. 1-7). Blacksburg, VA.

Garfield, J. (1994). Beyond testing and grading: Using assessment to improve instruction. *Journal of Statistical Education*, 1(2).

Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44-63.

Garfield, J. (2002) The challenge of developing statistical reasoning. *Journal of Statistics Education,* 10(3).

Garfield, J. & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning, 2,* 99-125.

Garfield, J., delMas, R. & Chance, B. (2003). The Web-based ARTIST Project. Paper presented at the *Annual Meeting of the American Educational Research Association*, Chicago.

Garfield, J. & Gal, I. (1999a). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67, 1-12.

Garfield, J. & Gal, I. (1999b). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Developing Mathematical Reasoning in Grades K-12: National Council Teachers of Mathematics 1999 Yearbook* (pp. 207-219). Reston, VA: N.C.T.M.

Garfield, J., Hogg, B., Schau, C. & Whittinghill, D. (2002). First courses in statistical science: the status of educational reform efforts, *Journal of Statistics Education,* 10(2).

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

Konold, C. (1989) Informal conceptions of probability. *Cognition and Instruction*, 6, 59- 98.

Konold, C. (1990). *ChancePlus: A computer-based curriculum for probability and statistics*. Final Report to the National Science Foundation. Scientific Reasoning Research Institute, University of Massachusetts, Amherst.

Konold, C. (2003) The reciprocal relation between learning goals and assessment. Paper presented at the *Annual Meeting of the American Educational Research Association*, Chicago.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics*, 23, 557-568.

Liu, H. J. (1998). *A cross-cultural study of sex differences in statistical reasoning for college students in Taiwan and the United States*. Doctoral dissertation, University of Minnesota, Minneapolis.

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

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan.

Joan B. GARFIELD
Department of Educational Psychology
University of Minnesota
Minneapolis, Minnesota 55455
USA

APPENDIX: STATISTICAL REASONING ASSESSMENT (SRA)

Purpose	The purpose of this survey is to indicate how you use statistical information in everyday life.
Take your time	The questions require you to read and think carefully about various situations. If you are unsure of what you are being asked to do, please raise your hand for assistance.

The following pages consist of multiple-choice questions about probability and statistics. Read the question carefully before selecting an answer.

1. 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.

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

___ a. Use the most common number, which is 6.2.

___ b. Use the 6.15 since it is the most accurate weighing.

___ c. Add up the 9 numbers and divide by 9.

___ d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. The following message is printed on a bottle of prescription medication:

WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.

Which of the following is the best interpretation of this warning?

___ a. Don't use the medication on your skin, there's a good chance of developing a rash.

___ b. For application to the skin, apply only 15% of the recommended dose.

___ c. If a rash develops, it will probably involve only 15% of the skin.

___ d. About 15 of 100 people who use this medication develop a rash.

___ e. There is hardly a chance of getting a rash using this medication.

3. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered <u>very</u> accurate if it rained on:

___ a. 95% - 100% of those days.

___ b. 85% - 94% of those days.

___ c. 75% - 84% of those days.

___ d. 65% - 74% of those days.

___ e. 55% - 64% of those days.

4. A teacher wants to change the seating arrangement in her class in the hope that it will increase the number of comments her students make. She first decides to see how many comments students make with the current seating arrangement. A record of the number of comments made by her 8 students during one class period is shown below.

Student Initials								
	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Number of comments	0	5	2	22	3	2	1	2

She wants to summarize this data by computing the typical number of comments made that day. Of the following methods, which would you recommend she use?

___ a. Use the most common number, which is 2.

___ b. Add up the 8 numbers and divide by 8.

___ c. Throw out the 22, add up the other 7 numbers and divide by 7.

___ d. Throw out the 0, add up the other 7 numbers and divide by 7.

5. A new medication is being tested to determine its effectiveness in the treatment of eczema, an inflammatory condition of the skin. Thirty patients with eczema were selected to participate in the study. The patients were randomly divided into two groups. Twenty patients in an experimental group received the medication, while ten patients in a control group received no medication. The results after two months are shown below.

	Experimental group (Medication)	Control group (No Medication)
Improved	8	2
No Improvement	12	8

Based on the data, I think the medication was:

___ 1. somewhat effective

___ 2. basically ineffective

<u>If you chose option 1,</u> select the one explanation below that best describes your reasoning.

___ a. 40% of the people (8/20) in the experimental group improved.

___ b. 8 people improved in the experimental group while only 2 improved in the control group.

___ c. In the experimental group, the number of people who improved is only 4 less than the number who didn't improve (12-8), while in the control group the difference is 6 (8-2).

___ d. 40% of the patients in the experimental group improved (8/20), while only 20% improved in the control group (2/10).

<u>If you chose option 2,</u> select the one explanation below that best describes your reasoning.

___ a. In the control group, 2 people improved even without the medication.

___ b. In the experimental group, more people didn't get better than did (12 vs 8).

___ c. The difference between the numbers who improved and didn't improve is about the same in each group (4 vs 6).

___ d. In the experimental group, only 40% of the patients improved (8/20).

6. Listed below are several possible reasons one might question the results of the experiment described above. Place a check by <u>every</u> reason you agree with.

___ a. It's not legitimate to compare the two groups because there are different numbers of patients in each group.

___ b. The sample of 30 is too small to permit drawing conclusions.

___ c. The patients should not have been randomly put into groups, because the most severe cases may have just by chance ended up in one of the groups.

___ d. I'm not given enough information about how doctors decided whether or not patients improved. Doctors may have been biased in their judgments.

___ e. I don't agree with any of these statements.

7. A marketing research company was asked to determine how much money teenagers (ages 13 - 19) spend on recorded music (cassette tapes, CDs and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $155 each year on recorded music.

Listed below are several statements concerning this survey. Place a check by <u>every</u> statement that you agree with.

___ a. The average is based on teenagers' <u>estimates</u> of what they spend and therefore could be quite different from what teenagers actually spend.

___ b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.

___ c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.

___ d. They should have asked teenagers coming out of music stores.

___ e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.

___ f. The average could be a poor estimate of the spending of all teenagers given that only teenagers in <u>malls</u> were sampled.

___ g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.

___ h. I don't agree with any of these statements.

8. Two containers, labeled <u>A</u> and <u>B</u>, are filled with red and blue marbles in the following quantities:

Container	Red	Blue
A	6	4
B	60	40

Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win $50. Which container gives you the best chance of drawing a blue marble?

___ a. Container A (with 6 red and 4 blue)

___ b. Container B (with 60 red and 40 blue)

___ c. Equal chances from each container

9. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

___ a. H H H T T

___ b. T H H T H

___ c. T H T T T

___ d. H T H T H

___ e. All four sequences are equally likely

10. Select one or more explanations for the answer you gave for the item above.

___ a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.

___ b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.

___ c. Any of the sequences could occur.

___ d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.

___ e. If you get a couple of heads in a row, the probability of a tails on the next flip increases.

___ f. Every sequence of five flips has exactly the same probability of occurring.

11. Listed below are the same sequences of Hs and Ts that were listed in Item 8. Which of the sequences is <u>least</u> likely to result from flipping a fair coin 5 times?

___ a. H H H T T

___ b. T H H T H

___ c. T H T T T

___ d. H T H T H

___ e. All four sequences are equally unlikely

12. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or a Oldsmobile. They first consulted an issue of <u>Consumer Reports,</u> which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.

The Caldwells then talked to three friends, two Oldsmobile owners, and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems, but nothing major. The Buick owner, however, exploded when asked how he liked his car:

First, the fuel injection went out — $250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.

The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?

___a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.

___ b. I would recommend that they buy the Buick in spite of their friend's bad experience. That is just one case, while the information reported in <u>Consumer Reports</u> is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.

___ c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

13. Five faces of a fair die are painted black, and one face is painted white. The die is rolled six times. Which of the following results is more likely?

___ a. Black side up on five of the rolls; white side up on the other roll

___ b. Black side up on all six rolls

___ c. a and b are equally likely

14. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

___ a. Hospital A (with 50 births a day)

___ b. Hospital B (with 10 births a day)

___ c. The two hospitals are equally likely to record such an event.

15. Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 p.m. on the evening before the test. The test scores for each group are shown in the graphs below. Each dot on the graph represents a particular student's score. For example, the two dots above the 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.

```
            .    .     .    .     .
       .    :     :     :    :    :     .    .     .         .
----------------------------------------------------------------------
  30         40      50        60      70      80     90      100
```

Test Scores: No- Sleep Group

```
                         .     .    .
    .    .    .     .    :    :     :    :    :     :   .
----------------------------------------------------------------------
  30         40      50        60      70      80     90      100
```

Test Scores: Sleep Group

Examine the two graphs carefully. Then choose from the 6 possible conclusions listed below the one you most agree with.

___ a. The no-sleep group did better because none of these students scored below 40 and the highest score was achieved by a student in this group.

___ b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.

___ c. There is no difference between the two groups because there is considerable overlap in the scores of the two groups.

___ d. There is no difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.

___ e. The sleep group did better because more students in this group scored 80 or above.

___ f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

16. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly. Listed below are several possible statements concerning the results of this research. Place a check by <u>every</u> statement that you agree with.

___ a. The sample of 500 is too small to permit drawing conclusions.

___ b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.

___ c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.

___ d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.

___ e. The research demonstrates that watching television causes poorer performance in school.

___ f. I don't agree with any of these statements.

17. The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following statements <u>must</u> be true if the average children per household is 2.2?

___ a. Half the households in the town have more than 2 children.

___ b. More households 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 household is 2.

___ f. None of the above.

18. When two dice are simultaneously thrown it is possible that one of the following two results occurs: *Result 1:* A 5 and a 6 are obtained. *Result 2:* A 5 is obtained twice.

Select the response that you agree with the most:

___ a. The chances of obtaining each of these results is equal

___ b. There is more chance of obtaining result 1.

___ c. There is more chance of obtaining result 2.

___ d. It is imposible to give an answer. (Please explain why)

19. When three dice are simultaneously thrown, which of the following results is MOST LIKELY to be obtained?

___ a. *Result 1*: "A 5, a 3 and a 6"

___ b. *Result 2*: "A 5 three times"

___ c. *Result 3*: "A 5 twice and a 3"

___ d. All three results are equally likely

20. When three dice are simultaneously thrown, which of these three results is LEAST LIKELY to be obtained?

___ a. *Result 1*: "A 5, a 3 and a 6"

___ b. *Result 2:* "A 5 three times"

___ c. *Result 3*: "A 5 twice and a 3"

___ d. All three results are equally unlikely

RELATIONSHIPS BETWEEN STUDENTS' EXPERIENCE OF LEARNING STATISTICS AND TEACHING STATISTICS

PETER PETOCZ
University of Technology, Sydney, Australia
Peter.Petocz@uts.edu.au

ANNA REID
Macquarie University, Sydney, Australia
Anna.Reid@mq.edu.au

SUMMARY

Students in the same statistics course learn different things, and view the role of the lecturer in different ways. We report on empirical research on students' conceptions of learning statistics, their expectations of teaching, and the relationship between them. The research is based on interviews, analysed using a qualitative methodology, with statistics students studying for a mathematics degree. Students expressed a range of conceptions of learning in statistics and a range of conceptions of their lecturers' teaching. These conceptions of learning and teaching were related, but not as closely or as exclusively as previous researchers have indicated. Looking at what students expect of teachers and their views of their own learning provides an opportunity for teachers to develop teaching practices that challenge students to move towards more integrated conceptions of statistics learning.

Keywords: Statistics education research; Learning; Teaching; Conceptual change

1. INTRODUCTION

Any teacher can tell you that it is obvious from the ways our students react in class, the sorts of questions they ask, the quality of their assessment tasks and the sorts of ways they integrate knowledge from one subject to another, that students learn in remarkably different ways. Research on learning often focuses on the ways that *teachers* understand teaching and learning (for example, Ho Yu et al., 2002; Weinberg and Abramovitz, 2000; McLean, 2000; Roiter & Petocz, 1996) based on the assumption that the teachers are best placed to make changes to the learning environment. Student evaluations of teaching, commonly undertaken by universities, quickly show that students have different expectations of what teaching should be about (Biggs, 2001). In investigating the problems of statistics education, some writers currently refer to a pedagogical 'reform' (Moore, 1997; Garfield et al., 2002) and discuss changes in content and methods of teaching, including explicit reference to the roles of assessment (Garfield & Gal, 1999) and attitudes (Gal et al., 1997). Many of these articles investigate and make comment about the teachers' perspective, implying again that changes and developments in teaching practice will result in some sort of change in learning. While these are important components, our approach has focused on investigating *students' understanding* of their own learning. Investigating the different ways that students understand learning in statistics will inform the current debate about the nature of reform in statistics

Statistics Education Research Journal 2(1), 39-53, http://fehps.une.edu.au/serj

education. We have previously reported on a study of the variety of different ways that statistics students understand learning statistics (Petocz & Reid, 2001) and teaching statistics (Petocz & Reid, 2002), and how students' epistemological beliefs about learning are related to a series of learning strategies and intended outcomes. Often, research of this kind focuses on teachers' beliefs about student learning, or students' understanding of their learning. Here we report on another component of our investigations, extending the scope of the previous discussions to explore in greater depth the relations between students' conceptions of learning statistics and their conceptions of teaching.

Kember (2001) has reported a relation between students' beliefs about learning and teaching with 'novice' and 'expert' part-time distance education students in Hong Kong. He concludes: "Overall this study has concluded that this set of beliefs about knowledge, learning and teaching is a fundamental factor in determining how well students cope with higher education and what they get out of it" (p.220). A similar relation was described by Beishuizen *et al.* (2001), who found that school-age children defined the characteristics of good teachers as personality and ability. They suggest that: "It is important to find out what students think about good teachers. Misunderstandings about mutual views of teachers and students may harm the efficacy and efficiency of teaching and learning" (p.186). Although our context is different, this study adds to these findings. We have also found a relation between students' conceptions of learning and teaching, and these conceptions of teaching (but not learning) have a strong affective component. Intriguingly, our study suggests that our statistics students consider teachers' enthusiasm for the subject and for their students to be a fundamental characteristic of good teaching. We describe the ways in which this, and students' other conceptions of teaching, are related to their understanding of learning.

Our research is based on interviews with 20 students from a first-year statistics class and a third-year class in regression analysis. All these students were studying on-campus, undertaking a degree in mathematics, with possible specialisations in statistics, finance or operations research. Each group was informed about the nature and aims of the project, and students were randomly selected from amongst those that had agreed to possible participation. The study was approved by the Human Research Ethics Committee of the University of Technology, Sydney, and the illustrative quotes that we use from the interviews are labelled with pseudonyms to avoid identification of individual students. We acknowledge the essential part played in the project by the students themselves.

These interviews showed that students expressed a range of conceptions of learning in statistics, some of them limiting, others expanding their view of statistics. It also showed that students expressed a range of ways that they experienced teaching, and that their ways of experiencing learning statistics and their expectations of teaching were related. In this paper, we summarise students' concepts of learning statistics and their conceptions of teaching statistics. We then discuss the relations between these two sets of conceptions, make comparisons with previous findings, and draw some implications for the continuing development of quality student learning environments (Reid & Petocz, 2001).

2. METHOD

Our choice of methodology – phenomenography – looks at how people experience, understand and ascribe meaning to a specific situation or phenomenon (Marton & Booth, 1997). Phenomenography can provide a rich description of an object of study through an emphasis on describing the variation in the meaning that is found in the participants' experience of the phenomenon. The outcome of a phenomenographic study is a set of logically related categories. These categories and the relations between them provide the *outcome space* for the research. The categories are usually reported in order of their inclusivity and sophistication, and they are defined by their qualitative difference from the other categories. However, it is the structure of the variation across the *group* that emerges through iterative readings of descriptions of the experience.

Data are typically collected through a series of in-depth, open-ended interviews that focus on allowing each person to fully describe their experience (Bowden, 1996; Ashworth & Lucas, 2000; Dortins, 2002). Questions are designed to encourage the participants to think about why they experience the phenomenon in certain ways and how they themselves constitute meaning of the phenomenon. In this case, students responded to questions on learning statistics: "What do you aim to achieve when you learn about statistics?", "What would you say learning in statistics was about?", "What do you do when you learn statistics?", "How do you know when you have learned something in statistics?". They were also asked questions about their perceptions of teaching: "What are your teacher's responsibilities?" and "How does your lecturer's teaching affect your learning?".

The questions were designed to focus students' awareness on different aspects of their experiences of learning and expectations of teaching, and were followed by responsive probing questions. The categories of description were developed on the basis of the range of responses. Interview transcripts were read by both authors, categories were suggested, refined and checked by repeated reading, and the final categories were confirmed by identification of appropriate quotes in the transcripts.

It is important to be aware of two important points when using this method for investigating people's experience of a certain phenomenon. Firstly, the outcome space describes a distillation of the *group's* experience and does not seek to describe one single person's experience. Secondly, this description of experience is situated; the participants are encouraged to focus their responses in relation to the specific investigative context. Hence, the categories describing variation in conceptions of the phenomenon represent a 'snapshot in time' in a specific context and do not represent a static view of any single individual.

Finally, for the purposes of this paper, we returned to our original transcripts once the initial categories had been obtained, and reanalysed each of them with the phenomenographic outcomes as the 'unit of analysis'. This process, which is an extension from the primary methodology, enabled us to consider the relations between a student's most integrated and sophisticated conception of learning statistics and their broadest conception of teaching statistics. It is *very important*, however, to recognise that these transcripts – and our decisions about them – only represent students' responses *at the time of interview*, and may not now represent the current state (and certainly not the permanent state) of any individual student's understanding of learning or teaching statistics. In fact, we would expect that, through the students' learning and our best teaching methods, they would develop their approach to learning, and maybe change their conceptions of teaching from time to time.

3. CATEGORIES DESCRIBING STUDENTS' CONCEPTIONS OF LEARNING STATISTICS

We have previously described students' conceptions of learning statistics in Petocz and Reid (2001). It is important to note that these categories are inclusive and hierarchical; they move from the most limited (Conception A) to the broadest (Conception F). Students who typically describe the more inclusive conceptions can use characteristics of the less inclusive conceptions if their perception of the situation demands; the reverse, however, is not generally true (Reid, 1997). Each conception is a complex relation between students' ideas about the subject, and their learning intentions and approaches. (A more detailed description of students' conceptions of *Statistics* appears in Reid and Petocz, 2002.) With each conception we have provided succinct quotes from the students' transcripts to illustrate the main characteristics of the category; these quotes are given verbatim and where interview questions are included, they are given in square brackets. Table 1 (below) shows these categories as one dimension of a table showing joint conceptions of learning and teaching.

3.1. CONCEPTION A– DOING: LEARNING IN STATISTICS IS DOING REQUIRED ACTIVITIES IN ORDER TO PASS OR DO WELL IN ASSESSMENTS OR EXAMS

Here, students focus on activities they have to do as part of their subject, which they think is sufficient to pass. They approach their study by attending lectures, reading, doing labs, repeating questions or examples until there are no mistakes, or doing previous exam papers. They aim simply to do well in assessment tasks and the exam.

Anne: *Well I think that the way in which they structure the course usually what happens is they have the revision before the exam, which is a good thing because most of the time it reinforces ideas and the main point that you are supposed to learn and also it forces you to study before the exam, and I think that is a very good thing for my case. … The truth is, I just learn what they teach me and I am not really sure about how broad statistics is yet.*

Hung: [How do you know you have learnt something in statistics?] *By doing different questions relating to the same problems over and over again without having to look back or getting any mistakes in them. … If they are going to cover a lot of the course in the final exam then it is probably better if they give out points relating to that at the beginning of the course.*

3.2. CONCEPTION B– COLLECTING: LEARNING IN STATISTICS IS COLLECTING METHODS AND INFORMATION FOR LATER USE

Here, students focus on gathering information, absorbing methods, increasing knowledge, and stockpiling examples or ideas. Students with this conception understand statistics to be about a group of techniques that need to be acquired in order to be used 'later'.

Emma: *It is furthering your knowledge, increasing what you know or coming across something you haven't thought of before. And remembering it for the future, putting it away somewhere. … I do all the examples, I think I learn by example, I read through the theory but I find that doing as many different types of problems as possible especially in statistics, like every different problem seems to pose a different way of doing it. I just try to do as many different problems as I can and I think I usually feel most confident when I have been able to do all the problems that I have been able to find if we were given past papers, or if we were given examples all through the book. If I can do all of them, then I am happy.*

Natasha: *What's learning?* (long gap) *Learning how to… hmmm…Going through the process of doing that, gathering… going through the… using stats, using your lecturer, that's learning. It's collecting information that in the end after you've absorbed everything you are able to use whatever you need to… It is very abstract.*

3.3. CONCEPTION C– APPLYING: LEARNING IN STATISTICS IS ABOUT APPLYING STATISTICAL METHODS IN ORDER TO UNDERSTAND STATISTICS

Here, students believe that actually carrying out the statistical activities provided will enable them to understand the subject of Statistics. They focus on doing practical things like examples, checking results and getting problems correct. In this regard their *approach* looks similar to Conceptions A and B: it is different, however, as their intention for their learning is to understand the subject Statistics.

Danny: *I try to focus on giving something a practical example in my own mind so that I can understand it. I can't, I would love to be able to rote learn things but I can't do it, I have to actually understand the concept to, get some sort of example or some kind of visualisation of it.*

Anne: *It was regression where it showed me the whole picture and that is when I understood that is what I was learning, and it is like oh that is how I use it. That's what I think every subject should have, to show how it is used in practice rather than just the theory.*

3.4. CONCEPTION D– LINKING: LEARNING IN STATISTICS IS LINKING STATISTICAL THEORY AND PRACTICE IN ORDER TO UNDERSTAND STATISTICS

This conception focuses on linking theory with practice. Students intend to find out how the practical exercises can inform their understanding of statistical theory, and vice versa. Students describe an intention to use statistics in 'real life' situations and they enjoy trying out their ideas on 'real' data.

Joe: *It is very important to be able to perform well in the laboratories side of things because if you only have an elementary understanding of the theory, of how it will work, it is actually the doing of the problems in regression analysis that is really important than understanding the theory behind it. The theory is important, it lays the foundation for the actual real world problems. You are not going to make money in the real world as a stats consultant or as an analyst knowing the theory behind the work. Whereas if you can solve a business's problems, that is more important I feel.*

Lucia: *Do the lab sheets, read text book and lecture notes. I think that helps me understand more by reading the text book and lecture notes, by doing the lab. It helps me understand more deeply about the information that I have read from the text and the lecture notes, in other words, reading gives more knowledge about statistics, doing labs gives more practice.*

3.5. CONCEPTION E– EXPANDING: LEARNING IN STATISTICS IS USING STATISTICAL CONCEPTS IN ORDER TO UNDERSTAND AREAS BEYOND STATISTICS

Here, students intend to connect statistical concepts with other areas. They aim to understand what they are doing, the meaning of data summaries, the broad subject area, and the real world meaning of what they are doing with numbers. They can see how statistics can be used outside the subject area or even outside the university context.

Paul: *I want to achieve understanding of what statistics actually means and how you interpret statistics in the real world and what is the value of statistics. ... All I want to achieve is how to get in-depth understanding of what statistics is. ... What I try to do is I try to apply the academic theory or concept into work and see if it is helpful or if it makes sense at work too.*

Jessica: *I guess, basically, by getting an answer right in the textbook or something, but also, like, when you do the work and also when you discuss with the lecturer, like, if you've got any problems you can go to the lecturer and you can ask them, and if you're on the right track then you know you're learning, I guess, and also just when you can find…, like, you can always get an answer from the textbook, but it's only when you start to understand it that you* *appreciate something, I guess, like…, I don't know… it's just like understanding the whole concept surrounding it, and things like that, and you're able to connect things with other things, like connect… just different aspects of what you're learning to something else and eventually figuring it out by yourself, and whatever…, something like that.*

Chris: *With statistics, ... it also helps me understand more in some of the financial models that are being used nowadays, whereas they don't teach it in business subjects. They might give you a brief explanation, 'blah, blah, blah', what it is, but they don't go in depth, and when they don't go in depth you don't get such a great understanding of what it is all about, and because I have done stats subjects it helps me understand a lot more. Even though the lecturer might not teach it, when I do my own reading, it actually helps me understand the context of that business subject, so it ties in with business and finance related subjects, so that's good for me.*

3.6. CONCEPTION F– CHANGING: LEARNING IN STATISTICS IS ABOUT USING STATISTICAL CONCEPTS IN ORDER TO CHANGE YOUR VIEWS

This is the most expansive and inclusive conception. Students focus on the changing quality of their own understanding of the broad idea of statistics and of the world. They see statistics as an intellectual tool that can be used to inform their understanding of many other areas, or to solve problems in other areas. They believe that their study of Statistics pushes them to change the way they view the world.

Lily: *I guess, you look at things differently when you have learned something. Like you know, this is totally non statistically based but if you learn about photography or light or stuff like that and how light focuses, I guess you'll always look at light differently. So whenever you see data and whenever you see graphs and things like that then you can look at them a little more critically. Then you look at tricks people use to change data and manipulate data.*

4. CATEGORIES DESCRIBING STUDENTS' CONCEPTIONS OF TEACHING STATISTICS

Our analysis of the transcripts identifies five qualitatively different ways in which students understand teaching in statistics. We have previously described these conceptions in Petocz and Reid (2002). As for the conceptions of learning statistics, they are presented here in an inclusive hierarchy from the most limited (Conception 1) to the most expansive (Conception 5). Table 1 (below) shows these categories as one dimension of a table showing joint conceptions of learning and teaching.

In the first conception, the focus is on the organisation and conditions for successful study.

4.1. CONCEPTION 1– PROVIDING MATERIALS, MOTIVATION, STRUCTURE

Here, students expect lecturers to provide them with good quality materials (e.g., course guides or lecture notes), motivation (e.g., interact, be enthusiastic, not boring); or structure (e.g., lectures for theory and labs for practice).

Jessica: *If you have any problems, even at the last minute, he'll always help you out, like, he'll tell you what to learn or what to do, so it's also like a personal kind of interaction with the students, not only just like a lecturer standing up there and teaching, or whatever, and like just… Obviously, he helps with the lecture notes and stuff like that, which is good, and* *then just when we get into the labs, he's always around and he helps around there. Like, he comes up and helps; it's personal, so it's no…, you're not so detached from your work and you feel like getting into it. I think it's good when you've got a lecturer that gives you motivation, and gives the interesting side to the subject…, basically, what I've said, it's just a personal thing.*

Lily: *I guess they should turn up and have an air of approachability. If they are not approachable if you have a problem then you won't approach them. I think that's pretty important. We have had a fair few lecturers, and some of them are kind of egotistical and you just can't approach them and if you do ask them a question you feel that they are looking down at you in a very stupid way, which is pretty bad. And they are very intelligent people so maybe it all comes naturally to them. I think social skills should be part of it.*

Melissa: *For regression analysis last semester we actually had lab times, so we'd have our lecture and then we would actually go into a lab and apply that straight away, so we would know what is actually going on. We would see the link between theory and practice. And also in Maths because it is such a small group we do have the one to one part as well in the labs and lectures. They even know your name, which is nice; you are not just a body. They actually do seem like they are willing to help and they actually care about your education. And they present things well. They have little booklets and things like that; you don't just learn from a textbook.*

In the next two categories, the focus is on the actual content of the course, and successful student learning within that course.

4.2. CONCEPTION 2– EXPLAINING MATERIAL AND HELPING WITH STUDENT WORK

In this conception, students expect that their lecturers will explain material coherently, providing clear guides for student work, assessment tasks, and ways of working. Lecturers should be able to deal with student problems, provide them with solutions, and review material at appropriate stages.

Anne: *He actually made a booklet so that we could read before we came to the lecture and so during the lecture he didn't actually go through each section of the book, rather he talked about tutorials in a lot of depth and he explained practical examples, so it was like if you went to the lecturer and read the book it is not the same content. So you feel you are not bored for one thing and you are getting two different sides of the same topic and it reinforces everything and you have the flexibility to read beforehand so you understand what he is talking about and I think that is one of the best ways.*

Melissa: *To actually teach things that you can apply in the real world and things like computer programs. Because you are not going to do it all by hand these days, so they have got to not only teach you the theoretical side but also the practical side, so that you can go out and actually get somewhere in your job. You know a lot of theory but you just can't apply it, or because you maybe have so many variables you can't do it by hand; it is impossible. They really have to show you how to use programs and what sort of things to look for in a program and commands.*

4.3. CONCEPTION 3– LINKING STATISTICAL CONCEPTS AND GUIDING LEARNING

In this conception students expect that lecturers will link statistical concepts by clarifying, explaining, elaborating on ideas, especially in unusual or different situations, and making connections between areas of the course.

Helen: *No matter how well read or intelligent that the lecturer may presume or the audience of students is really, the harsh reality of it is that the students, most of them are there to learn really a concept for the first time and really because they had a bit of background in the subject they don't really have any great idea about what the subject may be about. There may be some grey areas in that subject but simply the lecturer would need to go back on, elaborate on… and being able to make a smoother transition from one concept in a subject to another concept in the subject is really the key point to learning for students. Being able to relate one concept or idea closely to another is what really weaves all the different subjects together. ... To me, the important thing is that the lecturer can bind them altogether and show how one relates to the other, why this is important, why that goes hand in hand with the other and so forth.*

Chris: *I believe that their responsibility is to communicate the subject matter to us. They have to be able to communicate well, and basically teach us what the subject is about. They need to teach us the theory bit first, and then teach us how it's applied, and walk us through different examples and situations that we might get, and also you know how in some situations you can apply a certain theory, but then there's always a 'but' to it. I think they need to show us a few of the 'but' examples. So, basically, I believe that their main role is to communicate the point to us and help us to understand the subject matter, that's what I think the lecturers are there for.*

The next two conceptions have a focus on the students themselves that goes beyond their learning in a specific course or subject.

4.4. CONCEPTION 4– ANTICIPATING STUDENT LEARNING NEEDS

In this conception students expect that the lecturers will focus on the learning characteristics of each student in order to provide materials and methods that will best suit their learning needs. Students expect their lecturers to be teaching professionals, know the best methods to teach certain concepts and know what to do when students don't understand certain ideas.

Chris: *Some lecturers are incredibly passionate, like some of our stats lecturers in the maths faculty, they're really passionate about what they do. And because of that, they really help us to get the point across. They get the point across really well because first of all they know what they're talking about, and second of all they're very experienced, so they probably know what goes through a student's mind. They probably know how the students react to different subject matter. So powerpoint presentations do help, also the lecturer's tone, and I guess to me that's just about it; how the lecturers feel about their subject matter does help as well.*

Emma: *I think a teacher's responsibility is to have some idea of your capability and, even though I know it's sometimes impossible in a class of 200, in a small class, just what your students' strengths are, to know how best they respond to… if you're teaching a class in regression, knowing the best way to teach that subject. So, know the subject well enough to know the best way to teach it and the best way it should be learned, so whether it is by example to… to have the material well thought out, to just to be good at what they do, but to* *know it well enough to know the best way to get it across to a group of people. ... Going through the examples and not getting stuck on the finer points, leaving that up to the student to do their own reading, but ramming home the important points in the topic. What makes the topic different.*

4.5. CONCEPTION 5– BEING A CATALYST FOR 'OPEN MINDEDNESS'

In this conception students have an integrated view of a lecturer's responsibilities. They expect that lecturers will be a catalyst for their learning by showing them the importance of Statistics for general living, helping students change their view of the world, and opening the students' minds to new possibilities. This conception of teaching is different from the previous conceptions as it focuses on helping students develop high level understanding of Statistics.

Natasha: *Well, OK, different ways of looking at… well you are given data, different ways of looking at it and also helping you understand concepts and just opening your mind to… sometimes I have a one track mind so I wouldn't see a different scenario with some of the labs, different viewpoint, expanding my knowledge.*

5. LINKS BETWEEN STUDENTS' IDEAS ABOUT TEACHING AND LEARNING

Each of these conceptions of teaching or learning has been defined in terms of their qualitative *difference* from each other. However it is important to understand that there are also commonalities. Students' conceptions of both teaching and of learning statistics are inclusive and hierarchical. Students showing the broadest conceptions of teaching suggest that it is about focusing on their learning needs or helping them change their world view, but they also appreciate clear course notes, links between areas of the course and enthusiasm.

An example of this hierarchical nature is shown in the quotes from Melissa in Conceptions 1 and 2. Although she is aware of and appreciative of the materials and structure that her teachers provide, she also expects them to explain and help her with the computer programs that she uses; however, she did not make any comments about teaching that could be put into Conceptions 3, 4 or 5. In the context of her whole transcript, she was classed in Conception 2. Students showing the broadest conceptions of learning describe using statistics to change their views, but they are also able to discuss learning to apply statistical ideas and even learning basic statistical techniques. As another example, Anne describes understanding statistics by looking at practical examples in regression (quoted in Conception C) but also talks about learning by doing the required revision (Conception A); her views of learning statistics, taking her whole transcript into account, were classed under Conception C. Students showing the most limited conceptions, who suggest that good teaching is simply about the provision of materials and motivation and learning is focused on individual statistical techniques, will be happy with clear course notes describing these techniques, worked examples and lecturer enthusiasm. However, they seem to find it much harder to broaden their view to the more inclusive conceptions, to view teaching as a catalyst for 'open-mindedness' or learning as a way of changing their view of the world. Reid (1997) has demonstrated this inclusivity of conceptions in detail in another subject area (music), and it is also discussed by Prosser and Trigwell (1999, pp.108 -136).

If we look at the broad range of variation described in both sets of categories we can see that there are three main groupings in each. The two narrowest categories in learning and the narrowest in teaching form a group (AB, 1) that focuses on techniques of learning and teaching and emphasises these techniques as isolated activities; the mid-level categories represent a group (CDE, 23) that focuses on Statistics as a defined knowledge object; and the broadest categories in learning and in teaching form a group (F, 45) that focuses on the students and beyond the subject area of Statistics. These broad groupings show that the students in this study see that there are important areas of overlap between learning and teaching: (a) a focus on techniques, (b) a focus on the subject, and (c) a focus on the student. Kember (2001) suggested that students see learning and teaching as a coherent set of activities such that if a student's study focused on acquisition of techniques then their view of effective teaching would be supplying these specific techniques.

In some aspects, the conclusions from our research are broadly similar to Kember's ideas, but in other aspects they are quite different. Our descriptive categories indicate that there is a relation between conceptions of learning and teaching. Many students, such as Tran, Pat and Helen, show broadly consistent views on learning and teaching, as shown by the following excerpts.

Tran: *I think stats is not related to any other maths subject, or any other finance subject, just find out the relation between the independent variable and the dependent variable. ... I think I still need to learn the more basic things in stats.* [What are some of the more basic things?] *Hmm, it's very hard to think, because I'm not quite get to the point. ...* [What do you think your lecturers' responsibilities are?] *They give us a lot of tutorial work and mark them. If you ask questions the lecturer will tell you how to solve the problem.* [Are there any other things a good lecturer should do?] *Give us more work before the exam like sample papers. Some problems often occur in the exam: they should give us this to practice.*

These quotes, taken in the context of the whole transcript, suggest that Tran can be grouped as A1.

Pat: [What have you found is the most important area of statistics to learn?] *Understanding where it is coming from and how to see all the problems and how to apply it to things is important. Knowing how to do examples ...* [How do you go about learning statistics?] *I learn by doing it, doing all the examples. If you do enough, like, there is no point in me reading something because I won't understand it, to apply it ...* [What do you think your lecturers' responsibilities are?] *I don't know. Some of the lecturers, they basically just read out of a book, and it can't be just that, I think it is more than that, because a lot of them what they are saying you can read it out and it is no more clearer than the text book is, they have a slightly different thing to what the book says and add to it, and answer questions. I think that is important, someone that you can like ask questions and they make everything clearer, maybe it's a bit different to the way the textbook does it because it is just another version and could make it clearer.*

Pat's quotes suggest a grouping in C2.

Helen: *I think, to be able to sum it up in one sentence, it would be that it has changed to some effect the way I am viewing the world around me. I am finding that a lot of what I do learn in lectures with whatever subject does change the way I see things and the really effective lecturers will make sure that once you have left that room where the lecture has taken place to the real world you start to see things a lot differently, you start to understand how these figures apply, because we are talking about maths here, how different equations will apply, how this changing world has come to be what it is and why it has come to be what it is.*

This single quote (which we used in our previous writing about learning statistics, Petocz & Reid, 2001, and about teaching statistics, Petocz & Reid, 2002) shows that Helen holds the most inclusive conceptions of both learning and teaching, F5.

However, some individual students seem to show less consistency in their ideas. Maybe the most extreme example is Julie.

Julie: [How do you think learning about statistics will help you in your future career?] *Err, I know I can use it for… like, particularly jobs and things like that, and sometimes I get my neighbours and so forth asking me about it, and stuff, and just general use.* [What do you mean by 'general use'?] *Sometimes, I don't know, to work out my own percentages and so forth, and statistics has helped me with that, like the probability of this happening, and so forth.* [So your own world view?] *Yeah! It changes, like, sometimes, like, like I thought always studying was about just studying to get a job, and so forth, but it changes the way you view things, and so forth; I think of things differently!* ... [What do you think your teacher's responsibilities are?] *Err, to be there when they're needed, to make sure you understand, like if they get asked. My teacher's taught me how; he's very nice, he's always concerned, and he always makes time, and that's very important. Our tutes* [tutorials] *and so forth are very interesting, yeah, I guess that's what it turned out, kind of making maths fun and statistics fun.*

These quotes, and Julie's whole transcript, show that she holds the most inclusive conception of learning in statistics, but the least inclusive conception of teaching, F1.

Another example of this type is Jessica (E1); her earlier quotes show a sophisticated understanding of learning as understanding a whole concept and making connections with other areas, but she views her lecturer's role as providing motivation, interest and personal interaction. Reid (1995, 1999) previously identified this type of "inconsistency" and postulated that students with a sophisticated view of learning can view teaching as only a single component of their learning environment. This leads to the hypothesis that students will tend to have conceptions of teaching that are at the level of their conceptions of learning, or possibly at a lower level. An examination and categorisation of all 20 transcripts, shown in Table 1, lends some general support to this hypothesis, although the sample size is too small to draw definite conclusions (nor was this the original aim of the study). Almost all the classifications are on or above the 'diagonal' (interpreted in this case as the cells A1, A2, B2, C2, D2, E3, E4, E5 and F5). The obvious exception is Emma, classified as B4: her quotes (given earlier) show that she sees learning in statistics as collecting techniques and materials for future use, but she views her teachers' responsibilities as anticipating their students' learning needs.

Table 1. Students' Conceptions of Learning and Teaching in Statistics

Learning / Teaching	A. Doing	B. Collecting	C. Applying	D. Linking	E. Expanding	F. Changing
1.Providing essentials	3		1		2	2
2. Explaining ideas	3	1	1	2		
3. Linking concepts					1	
4.Anticipating learning needs		1			1	
5. Catalyst for open-mindedness					1	1

Interestingly, a significant area of difference between the conceptions of learning and of teaching is shown in the affective component. Almost all students in our sample indicated that an important component of good teaching was enthusiasm, interest and motivation. Beishuizen et al. (2001) have also found that students see this as an important feature of teaching, and it seems to be one component in the 'attitudes and beliefs' discussed by Gal et al. (1997). Student feedback on teaching (such as the Course Experience Questionnaire in Australia) consistently indicates that students rate lecturers' enthusiasm as very important to good teaching; however, scores on that scale are consistently independent of other aspects, such as teachers' knowledge of the subject (Ainley, 2001). This affective component seemed to be completely missing in students' discussions of their own learning. It seems that students value enthusiasm and motivation, but they believe that it is an aspect of their studies that comes from outside themselves rather than from within.

6. IMPLICATIONS FOR RESEARCH AND APPLICATION IN STATISTICS PEDAGOGY

Although our conclusions are based on interviews with only 20 students, they are supported by our broadly parallel findings in several other areas including design education (Davies & Reid, 2000) and accounting education (Jebeili & Reid, 2002), and also by the findings of other research (for example, Crawford et al., 1994, in mathematics education). Moreover, they are congruent with more general ideas about conceptions of learning such as those shown in the classic report of Marton et al. (1993) and with various aspects of the 'reform' movement in statistics education (Garfield et al., 2002). Our (and others') further research can investigate whether our broad categories remain valid in other statistics learning situations, and to what extent pedagogy utilising our findings can be effective in enhancing student learning during the course of a unit of study, or even of a whole degree.

These findings are important for the development of learning environments that can engage students' interest, broaden their understanding of statistics and enrich their own lives. Previously we have suggested that the development of learning environments must be 'total' (Reid and Petocz, 2001). Sustained change and development can only take place if there is alignment between intentions and actions of students, tutors and lecturers taking account of all teaching and learning activities, assessment tasks, and understandings about content.

The current study reveals a new dimension worth consideration; the different ways that students understand teaching and learning. The categories described show that this variation plays an important role in the way students approach their learning and their expectations for the lecturer's role. If we take these categories into account, then the development of the total learning environment will acknowledge the variation and seek to provide solutions that will enable students to change their ways of thinking about learning and teaching statistics towards the more inclusive levels. There are several things that we as statistics educators can do in response to these findings.

Firstly, we can help students become aware of the range of variation in conceptions of learning and conceptions of teaching. This may be enough in itself to encourage some students towards broader views. Any activity that allows students to explore the nature of their own learning and compare it to others in the same situation can form the basis of discussion. For instance, the 'usual' discussion surrounding assessment tasks can be directed towards an exploration of the different ways that students may tackle such tasks. Research on using phenomenographic outcomes, such as those presented in this study, has shown that simple awareness of difference can be a catalyst for change (Reid, 2002).

Secondly, we can provide activities and assessment that encourage students towards the broadest levels of learning. It is easy to construct classroom activities and assessment tasks that cater for the lower levels of learning statistics and that sit well within the realm of the lowest level of teaching statistics. For instance, a series of numbers and a request to find the mean, median and standard deviation will appear to satisfy everyone. The students have learned definitions and the technical processes to extract such information from data, the lecturer will find such a question quick and easy to mark, and the department will be likely to have a high pass rate. (Although this example may seem trivial, we observed exactly this in a recent examination question in a servicing statistics course.) However, the same question set in a specific situation where students are asked to explain the *meaning* of these observations and summary statistics for the *people* involved (such as a client or a colleague) immediately shifts students' focus. This sort of question also implies a more reflective style of teaching rather than the provision of simple definitions and worked solutions in class, and technically-focused assessment questions that are so often the result of time pressures, constraints in content, and ease of marking. Our findings imply that to change the total learning environment teachers need also to focus on the meaningfulness of statistics, and to constantly provide a variety of tasks that relate to real, important, interesting, messy, and complex data. If there is a consistency (or alignment, Biggs, 1999) between learning and teaching intentions and activities, students are more able to change their own thinking about learning, teaching, and statistics.

To encourage the highest levels of learning, a teacher can influence students' conceptions of teaching by moving the focus of their teaching from the provision of certain essentials, to the subject itself, and most importantly to the students' own learning. While acknowledging the importance of components of learning such as assessment (Garfield & Gal, 1999) or the technology of learning (Moore, 1997), this implies less focus on the curriculum itself, and certainly less focus on the traditional concern of material to be 'covered' or 'examined'. These research outcomes are critically important for the development of effective learning materials. For example, the laboratory exercises in Petocz (1998), which invite students to engage with the meaning of the data that they are analysing using a computer package, and to communicate this meaning in a variety of specific situations, or the book 'Reading Statistics' (described by Wood and Petocz, 2002), which encourages students to 'read' statistical papers, articles, and research in a variety of areas of application, with the aim of looking beyond the data to the 'real life' meanings. Now, the focus moves towards supporting students in their own learning, holistically, and beyond the arbitrary boundaries of the subject. This, in turn, can encourage students towards more inclusive views of their own learning.

REFERENCES

Ainley, J. (Ed.) (2001). *Course Experience Questionnaire*. Australian Council for Educational Research (ACER), Graduate Careers Council of Australia.

Ashworth, P. & Lucas, U. (2000). Achieving empathy and engagement: a practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education,* 25(3), 295–308.

Beishuizen, J., Hof, E., van Putten, C., Bouwmeester, S. & Asscher, J. (2001). Students' and teachers' cognitions about good teachers. *British Journal of Educational Psychology*, 71, 185–201.

Biggs, J. (1999). *Teaching for Quality Learning at University.* Society for Research in Higher Education, Open University Press, Great Britain.

Biggs, J. (2001). The reflective institution: assuring and enhancing the quality of teaching and learning. *Higher Education*, 41, 221–238.

Bowden, J. (1996). Phenomenographic research – some methodological issues. In G. Dall'Alba & B. Hasselgren (Eds.), *Reflections on Phenomenography* (pp. 49-66). Goteborg (Sweden): Studies in Educational Sciences 109.

Davies, A. & Reid, A. (2000). Uncovering problematics in design education: learning and the design entity. In C. Swann & E. Young (Eds.), *Re-Inventing Design Education in the*

University: Proceedings of the International Conference (pp. 178-184). Perth: School of Design, Curtin University.

Dortins, E. (2002). Reflections on phenomenographic process: interview, transcription and analysis. In A. Goody, J. Herrington & M. Northcote (Eds.), *Quality Conversations: Research and Development in Higher Education* (vol. 25, pp. 207–213). Perth: HERDSA.

Gal, I., Ginsburg, L. & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 37-54). Amsterdam: IOS Press.

Garfield, J. & Gal, I. (1999). Assessment and statistics education: current challenges and directions. *International Statistical Review*, 67(1).

Garfield, J., Hogg, R., Schau, C. & Whittinghill, D. (2002). First courses in statistical science: the status of educational reform efforts. *Journal of Statistics Education*, 10(2).

Ho Yu, C., Andrews, S., Winograd, D., Jannasch-Pennell, A. & DiGangi, S. (2002). Teaching factor analysis in terms of variable space and subject space using multimedia visualization. *Journal of Statistics Education*, 10(1).

Jebeili, S. & Reid, A. (2002). Constructive alignment in accounting education: Evaluating the use of case studies and problem-based approaches in large classes. *Proceedings of About Evaluations and Assessment Conference 2002.* Brisbane: Queensland University of Technology (CD ROM).

Kember, D. (2001). Beliefs about knowledge and the process of teaching and learning as a factor in adjusting to study in higher education. *Studies in Higher Education,* 26(2), 205–221.

Marton, F., Beaty, E. & Dall'Alba, G. (1993). Conceptions of learning. *International Journal of Educational Research*, 19, 277–300.

Marton, F. & Booth, S. (1997). *Learning and Awareness.* New Jersey: Lawrence Erlbaum.

McLean, A. (2000). The predictive approach to teaching statistics. *Journal of Statistics Education*, 8(2).

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

Petocz, P. (1998). *Statistical laboratory exercises using Minitab: A guide to understanding data*. Brisbane: Jacaranda Wiley.

Petocz, P. & Reid, A. (2001). Students' experience of learning in statistics. *Quaestiones Mathematicae, Supplement 1*, 37–45.

Petocz, P. & Reid, A. (2002). How students experience learning statistics and teaching. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, ICOTS-6*, Capetown: International Association for Statistical Education (CD ROM).

Prosser, M. & Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. United Kingdom: Society for Research into Higher Education.

Reid, A. (1995). The influence of teaching on learning in music: a case study. In L. Zelmer, L. (Ed.), *Higher Education: Blending Tradition and Technology*. Research and Development in Higher Education (vol 18, pp. 622–627). Rockhampton, Australia: HERDSA.

Reid, A. (1997). The hierarchical nature of meaning in music and the understanding of teaching and learning. *Advancing International Perspectives*, 20, 626–31.

Reid, A. (1999). *Conceptions of Teaching and Learning Instrumental and Vocal Music*. PhD Thesis. Sydney University of Technology.

Reid, A. (2002). Is there an 'ideal' approach to academic development? In A. Goody & D. Ingram (Eds.), *Spheres of influence: Ventures and visions in educational development*. Crawley, WA: University of Western Australia. Online at www.csd.uwa.edu.au/ICED2002/publication/.

Reid, A. & Petocz, P. (2001). Using professional development to improve the quality of assessment tasks and student learning environments. In C. Rust (Ed.), *Improving Student Learning Strategically*. Oxford: Brookes.

Reid, A. & Petocz, P. (2002). Students' conceptions of statistics: a phenomenographic study. *Journal of Statistics Education*, 10(2).

Roiter, K. & Petocz, P. (1996). Introductory statistics courses. A new way of thinking. *Journal of Statistics Education*, 4(2).

Weinberg, S. & Abramowitz, S. (2000). Making general principles come alive in the classroom. *Journal of Statistics Education*, 8(2).

Wood, L. N. & Petocz, P. (2002). Reading statistics. In A. Fernandez (Ed.), *UniServe Science Proceedings of Scholarly Inquiry in Flexible Science Teaching and Learning Symposium*. University of Sydney. Online at http://science.uniserve.edu.au/pubs/procs/.

PETER PETOCZ
Department of Mathematical Sciences
University of Technology, Sydney
PO Box 123, Broadway, NSW 2007, Australia

EVALUACIÓN DE LA COMPRENSIÓN DE LA CORRELACIÓN Y REGRESIÓN A PARTIR DE LA RESOLUCIÓN DE PROBLEMAS

ANTONIO ESTEPA
Universidad de Jaén, Spain
aestepa@ujaen.es

FRANCISCO T. SÁNCHEZ-COBO
Universidad de Jaén, Spain
fsanchez@ujaen.es

RESUMEN

En este trabajo presentamos un estudio exploratorio de caracterización del conocimiento que los estudiantes universitarios tienen sobre la correlación y regresión. Analizamos las soluciones a dos problemas de una muestra intencional de 193 estudiantes, que habían recibido un curso de estadística descriptiva en el ámbito universitario. Estudiamos los procedimientos de los estudiantes y discutimos las dificultades y errores que muestran los alumnos sobre el centro de gravedad del diagrama de dispersión, las rectas de regresión, el coeficiente de correlación, el tipo de relación entre las variables y la predicción.

Palabras Clave: Investigación en educación estadística; Aprendizaje; Evaluación; Correlación; Regresión

SUMMARY

In this paper we present an exploratory study intended to characterise University students' understanding of correlation and regression. We analyse the solutions to two problems from an intentional sample of 193 students who had previously received a course of descriptive statistics at the University. We study the student's procedures and discuss their difficulties and errors concerning the centre of gravity in the scatter plot, regression lines, correlation coefficient, type of relation between the variables and prediction.

Keywords: Statistics education research; Learning; Assessment; Correlation; Regression

1. INTRODUCCIÓN

Uno de los contenidos estadísticos de mayor importancia, tanto por su significación práctica como por su carácter instrumental para otros conceptos estadísticos, es la asociación estadística, entendiendo por tal la extensión del concepto de correlación a variables cualesquiera incluso no numéricas (Hildebrand, Lang y Rosenthal, 1977). La asociación

Statistics Education Research Journal 2(1), 54-68, http://fehps.une.edu.au/serj

estadística aparece en los currículos de enseñanza secundaria y en los cursos universitarios, bien en sus tres campos de problemas (tablas de contingencia, correlación y comparación de muestras), bien, en alguno de ellos. La correlación y regresión es un tema de estudio en los cursos de iniciación a la estadística de los primeros niveles de muchas carreras universitarias.

En el presente trabajo presentamos un estudio exploratorio en que se profundiza en el estudio de la comprensión de conceptos relacionados con la correlación y la regresión en situaciones de resolución de problemas, en los primeros cursos universitarios. De esta manera caracterizaremos la comprensión de los estudiantes de la muestra sobre estos conceptos, después del proceso de estudio seguido en un curso académico. Los resultados obtenidos son de utilidad para futuras planificaciones de la enseñanza de un tópico tan importante en estadística como la correlación y regresión.

2. INVESTIGACIONES PREVIAS

A pesar de la relevancia estadística y curricular de la asociación estadística, la investigación llevada a cabo sobre este tópico es escasa en general y, procede de la Psicología y de la Didáctica de la Matemática, siendo más numerosas las investigaciones psicológicas que las didácticas. A continuación, vamos a presentar un resumen de las aportaciones que la investigación ha proporcionado sobre la correlación y regresión.

Las investigaciones psicológicas sobre correlación de variables continuas están orientadas al estudio de la capacidad de estimación de la correlación existente por parte de los sujetos. Dicha estimación se lleva a cabo por las personas a partir de su impresión subjetiva (no se hace uso de cálculo formal) de la dirección e intensidad de la relación existente entre las variables, en una tarea donde los datos se presentan en un diagrama de dispersión o tabla de valores. Aunque estos trabajos no se relacionan de manera directa con el nuestro, sí nos alertan de que las percepciones subjetivas de la correlación pueden influir en las respuestas dadas por los alumnos en la resolución de problemas. Las principales conclusiones de estos trabajos son las siguientes:

1. La estimación de la correlación es más exacta cuando se dan los datos en forma gráfica, que cuando se dan en forma tabular (Lane, Anderson y Kellam, 1985).
2. La estimación de la correlación mejora cuando hay más datos y la correlación es alta (Erlick y Mills, 1967).
3. Si existen teorías previas sobre la intensidad y signo de la correlación, se sobrestima la correlación real existente, siempre que ésta coincida con las ideas previas, mientras que, en caso contrario, es necesaria la presencia de una correlación fuerte para que los sujetos la detecten (Jennings, Amabile y Ross, 1982).

Las investigaciones de carácter didáctico se han llevado a cabo con el objeto de obtener resultados que mejoren la enseñanza aprendizaje de este importante tópico. Entre ellas destacaremos las que tienen una relación más directa con el presente trabajo.

Estepa y Batanero (1996) analizan los juicios intuitivos sobre la correlación que 213 estudiantes preuniversitarios sin instrucción previa dan a partir del diagrama de dispersión de variables con distinto tipo de correlación (directa, inversa, intensa, moderada). Clasifican las estrategias utilizadas para realizar los juicios y la influencia de las variables de tarea en las dichas estrategias. Identifican tres concepciones incorrectas de la correlación:

1. *Concepción determinista de la asociación:* Cuando el estudiante sólo admite un sólo valor para la variable dependiente para cada valor de la variable independiente, para

considerar que las variables están relacionadas. Estos estudiantes consideran independientes las variables si no se cumple esta condición.

2. *Concepción local de la asociación:* Cuando el estudiante realiza el juicio de asociación basándose únicamente en una parte de los datos presentados. Si la parte de datos considerada presenta un tipo de correlación, por ejemplo, directa, considera que éste es el tipo de correlación para todo el conjunto de datos.
3. *Concepción causal de la asociación*: Cuando el estudiante se basa únicamente en la existencia o no de relación causal para afirmar la existencia de correlación, sin tener en cuenta la correlación empírica observada en los datos.

En el estudio de Batanero, Estepa, Godino y Green (1996) sobre interpretación de tablas de contingencia, se encontró además, otra concepción incorrecta de la asociación estadística para las variables presentadas en tablas de contingencia; La *concepción unidireccional de la asociación* se presenta cuando el estudiante percibe la existencia de asociación, solamente cuando la relación entre las variables es directa, considerando la relación inversa como independencia.

Morris (1997) estudia las concepciones de los estudiantes sobre la correlación y sugiere las tareas más adecuadas para valorar su comprensión. Encuentra la concepción causal de la asociación y dificultades y confusiones que los sujetos exhiben con el signo de la misma. También preguntó a los estudiantes si encontraban la estadística difícil y útil para su formación, encontrando que era difícil para el 60% de la muestra estudiada y que ningún estudiante piensa que la estadística es irrelevante como materia de estudio. Nuestro trabajo completa el de Morris (1997), ya que añadimos el estudio de la regresión y estudiamos el interés que tiene el tópico correlación y regresión para los estudiantes, dentro de la asignatura de estadística.

Truran (1997) compara los datos de evaluación en dos cursos impartidos en universidades de Australia y Malasia sobre correlación y regresión. Estudia la interpretación que los estudiantes hacen del coeficientes de correlación y determinación, la pendiente y ordenada en el origen de la ecuación de regresión, la predicción y sus restricciones. Sus conclusiones reflejan pocas diferencias entre los estudiantes de ambos países. Casi todos los alumnos identifican correctamente la correlación moderada y negativa, aunque también encuentra la concepción determinista de la asociación. Asimismo, indica que algunos estudiantes argumentan, de modo razonable, las reservas que se deben tener con las extrapolaciones como son la intensidad de la correlación y el tamaño de la muestra. Este trabajo será una referencia para el nuestro.

Estepa y Sánchez-Cobo (1998) estudian la presentación de la correlación y regresión en los libros de texto de bachillerato, examinando, además, los ejercicios propuestos. Entre los hechos encontrados debemos subrayar el enfoque teoría-práctica en el desarrollo de los temas y la aplastante mayoría de ejemplos y ejercicios con conjuntos de datos con correlaciones positivas, frente a las negativas e independencia. Esto puede llevar a los estudiantes a considerar que sólo existe la correlación positiva (concepción unidireccional) o que para que exista correlación, ésta debe ser muy fuerte. Además, en los libros de texto de Educación Secundaria se presta poca atención al centro de gravedad de los diagramas de dispersión$(\bar{x}, \bar{y})$, es decir, el punto definido por las medias de las dos variables, a la distinción entre la variable explicativa y la explicada y a los problemas de predicción, utilizando las rectas de regresión. Este estudio nos proporcionó la base para diseñar la presente investigación.

3. METODOLOGÍA DE LA INVESTIGACIÓN

3.1. OBJETIVO DEL ESTUDIO

Estos resultados nos motivaron a profundizar en el conocimiento de la comprensión de los conceptos de correlación y regresión por parte de los estudiantes. El presente artículo reanaliza parte de los datos obtenidos por Sánchez-Cobo (1999), sobre la comprensión de la correlación y regresión en estudiantes universitarios. En el citado trabajo se estudió la estimación de la correlación que hacen los alumnos a partir de diversas representaciones (descripción verbal, diagrama de dispersión, tabla y coeficiente de correlación), observando que la exactitud de la estimación depende del tipo de representación (tabla, gráfico o expresión verbal), intensidad y signo de la correlación y tipo de covariación (Sánchez-Cobo, Estepa y Batanero, 2000). Este tipo de tarea es también explorada por Moritz (2002) con alumnos de educación primaria, llegando a la conclusión de que es importante tener en cuenta los procesos de traslación entre estas diferentes representaciones.

También completamos la investigación de Estepa y Batanero (1996), cuyos estudiantes no habían estudiado el tema de correlación y regresión y a los que se les pedía que estimasen la relación existente entre dos variables en forma intuitiva, guiándose tan sólo del diagrama de dispersión. Nuestra muestra de estudiantes ha seguido un curso de iniciación a la estadística, estudiando el tema de correlación y regresión y se les ha pedido calcular el coeficiente de correlación para interpretarlo y decidir que tipo de relación presenta el conjunto de datos del problema propuesto.

Los conceptos estudiados a continuación son: la relación existente entre el centro de gravedad de la nube de puntos $(\bar{x}, \bar{y})$ y la recta de regresión, el cálculo e interpretación del coeficiente de correlación, el tipo de relación entre las variables, determinación de las rectas de regresión, interpolación y discriminación entre las dos rectas de regresión. El estudio está basado en el análisis de los procedimientos, aciertos, errores y dificultades que los estudiantes manifiestan en sus respuestas a las cuestiones de los problemas planteados. Identificamos los procedimientos de resolución utilizados por los estudiantes, estudiando su adecuación y corrección en su desarrollo. Con esto completamos, en parte, las investigaciones anteriores aportando conocimientos válidos para la enseñanza del tema.

3.2. PROBLEMAS PLANTEADOS

En lo que sigue se estudian y discuten los resultados de dos problemas, que formaron parte de un cuestionario más amplio completado por los alumnos de la muestra, parte de los cuales fueron publicados en Estepa y Sánchez-Cobo (2001). Estos problemas que analizaremos a continuación se presentan en forma de prueba de ensayo, con respuestas abiertas y, en consecuencia, produce datos de tipo cualitativo.

Problema 1. Una recta de regresión tiene una pendiente de 16 y corta al eje de ordenadas en el punto y = 4. Si la media de la variable independiente es 8, ¿Cuál es la media de la variable dependiente?

Problema 2. La siguiente tabla muestra el número de bacterias por unidad de volumen que está presente en un cultivo después de un cierto número de horas.

Número de horas	1	2	3	4	5
Número de bacterias por unidad de volumen	18	21	33	54	61

a) Calcule el coeficiente de correlación lineal.

b) Indique qué tipo de relación (directa, inversa o independencia) existe entre ambas variables.

c) Determine la recta de regresión de y, número de bacterias por unidad de volumen, sobre x, número de horas.

d) ¿Qué número de bacterias cabe esperar que habrá, transcurridas 2.5 horas?

e) ¿Qué tiempo deberá pasar para que el número de bacterias del cultivo sea 27?

El primer problema fue tomado de Cruise, Dudley y Thayer (1984, pp. 288). Para su resolución los estudiantes deben utilizar los siguientes conceptos matemáticos: recta, pendiente, ordenada en el origen. Respecto a los conceptos estadísticos, son necesarios recordar el hecho de que la recta de regresión pasa por el centro de gravedad del diagrama de dispersión, fundamental para la aproximación intuitiva de la recta de regresión a partir del diagrama de dispersión. De esta manera se relaciona el contexto geométrico y el estadístico de esta situación (Truran, 1997). Por último, el estudiante debe usar la predicción, esencial en el estudio de la regresión.

El segundo problema es una adaptación del que aparece en Vizmanos y Anzola (1988, pp. 372). Además de los cálculos del coeficiente de correlación y de las rectas de regresión, se pretende caracterizar el uso que los alumnos hacen de ellos. En el apartado b) se busca que el estudiante interprete el coeficiente de correlación, ya que algunos autores (Morris, 1997; Truran, 1997) sugieren dificultades en esta tarea. Los apartados d) y e) se dirigen a explorar si el alumno ha interiorizado debidamente la predicción y si llega a comprender su naturaleza estocástica (Truran, 1997), haciendo un uso conveniente de ella, es decir, empleando la recta de regresión adecuada y haciendo referencia al valor promedio.

Antes de decidir el cuestionario definitivo, una primera versión del mismo fue ensayada en una muestra piloto, compuesta por un reducido número de alumnos, distintos de los que participaron en la muestra final, con el fin de estudiar su legibilidad y facilidad de comprensión. Teniendo en cuenta los resultados obtenidos decidimos el cuestionario definitivo.

3.3. MUESTRA

En los centros educativos, los alumnos están agrupados en tipos de estudios y niveles de enseñanza, y en consecuencia, cuando se plantea una investigación educativa, es difícil conseguir una muestra aleatoria. Por tanto, muchas veces, las muestras están constituidas por grupos naturales de estudiantes que cursan una asignatura de una carrera determinada, y son intencionales, como en nuestro caso, constituyendo la investigación un cuasi-experimento, en el sentido de Cook y Campbell (1979). Al ser nuestra muestra de carácter intencional, somos conscientes de las limitaciones en la generalización de nuestros resultados.

La muestra final estuvo formada por 193 estudiantes de la Universidad de Jaén, 57 hombres (30%) y 136 mujeres (70%), con una edad media de 20 años. De ellos, 104 (53.8%) estudiaban la Diplomatura de Empresariales y 89 (42.2%) la Diplomatura de Enfermería; 109 (56.4%) estudiantes tenían una formación preuniversitaria orientada a ciencias y 84 (43.6%) a letras. Por otra parte, 117 (60.6%) estudiantes no habían cursado Estadística en su formación preuniversitaria. Además, 155 (80.3%) de los estudiantes indicaron que los temas de correlación y regresión tienen suficiente, bastante o mucho interés dentro del programa de la asignatura de estadística que estudiaban. Finalmente, 162 (83.9%) de los estudiantes sugirieron que la asignatura de Estadística tiene suficiente, bastante o mucho interés para su formación dentro del plan de estudios. Estos resultados son ligeramente inferiores a los de Morris (1997), ya que todos sus estudiantes encontraron relevante la asignatura de estadística.

El programa de la asignatura de Estadística que estudiaban los alumnos de nuestra muestra comprendía contenidos fundamentales de estadística descriptiva, tales como: tablas de frecuencias y gráficos, medidas de centralización y dispersión, simetría y apuntamiento, variables estadísticas bidimensionales, tablas de contingencia, covarianza y correlación, regresión lineal y polinómica, muestreo, distribución en el muestreo, intervalos de confianza y test de hipótesis. Los problemas que presentamos fueron elegidos teniendo en cuenta el análisis de la enseñanza realizada en la programación del profesor del tema de correlación y regresión y los apuntes de dos estudiantes.

A los 193 estudiantes de la muestra se les pidió que cumplimentaran un cuestionario completo, dándoles, antes de comenzar unas breves instrucciones verbales, que enfatizaban y aclaraban las expuestas por escrito en el cuestionario. Como se puede observar en la lectura de los problemas propuestos, las respuestas esperadas de los estudiantes son abiertas.

Para analizar las respuestas de los estudiantes en cada uno de los apartados de los problemas, se codificaron las respuestas en cada cuestión y se agruparon según su similitud, obteniendo las distintas categorías de respuestas con las que se llevó a cabo un análisis cualitativo (Miles y Huberman, 1984; Huberman y Miles, 1994). Finalizamos el análisis con la clasificación de las categorías de respuestas, según el procedimiento utilizado y la corrección de las mismas, obteniendo su frecuencia y porcentaje.

4. RESULTADOS Y DISCUSIÓN

4.1. CÁLCULO DE LA MEDIA DE LA VARIABLE EXPLICADA

En el problema 1 el estudiante ha de utilizar sus conocimientos analíticos sobre la regresión lineal, incluyendo el centro de gravedad $(\bar{x}, \bar{y})$, su relación con la recta de regresión, la pendiente y la ordenada en el origen (Truran, 1997). Asimismo, debe tenerse en cuenta que, mientras que la correlación es una relación simétrica, pues su valor no depende de cuál sea la variable elegida como explicativa en la variable estadística bidimensional, en el caso de la regresión, hay que tener en cuenta el tipo de covariación (Barbancho, 1973).

Tabla 1. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 1 según procedimiento de resolución

Procedimiento	Tipo de solución			Número de alumnos que usa el procedimiento
	Correcta	Incorrecta	No responde	
$y = mx + n$	31 (38.3)	50 (61.7)		81
$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$	8 (34.8)	15 (65.2)		23
Uso de las dos expresiones anteriores	2 (100)			2
	2 (22.2)	7 (77.8)		9
$x - \bar{x} = \frac{S_{xy}}{S_y^2}(y - \bar{y})$		4 (100)		4
Uso de las dos rectas de regresión	2 (66.7)	1 (33.3)		3
Uso de un parámetro estadístico		7 (100)		7
No responde			64(100)	64
Total	45 (23.3)	84 (43.5)	64 (33.16)	193

Si la dependencia es causal unilateral, las variables explicativa y explicada quedan determinadas de forma unívoca. Si el tipo de covariación es uno de los cuatro restantes - interdependencia, dependencia indirecta, concordancia o covariación casual - el resolutor debe decidir qué recta de regresión hay que emplear, *y* sobre *x* o *x* sobre *y*, y, desde ese momento, quedan determinadas ambas variables, explicativa y explicada. Puesto que en nuestro caso no se propone un contexto, el estudiante puede elegir entre las dos rectas.

Este problema ha sido difícil para esta muestra de estudiantes, ya que han dado respuesta solamente 129 estudiantes (66.8% de la muestra), siendo correctas sólo 45 (34.9% de la muestra). En la Tabla 1, hemos clasificado las respuestas, por procedimiento seguido y tipo de solución. De los 129 estudiantes que han respondido, 122 (94.57% de respuestas) han utilizado una o ambas rectas de regresión para resolver el problema. La principal dificultad encontrada en la resolución de este problema es que algunos estudiantes no tienen en cuenta que el centro de gravedad $(\bar{x}, \bar{y})$ está contenido en la recta de regresión. En la enseñanza este tema sólo se trató de manera implícita, aunque se hizo referencia a tal punto, como punto de corte de las dos rectas de regresión. A continuación detallamos las soluciones.

Uso de la Recta de Regresión de Y sobre X

El 80.62% de los alumnos que han respondido han utilizado en sus cálculos la expresión de la recta de regresión de *y sobre x*. Distinguimos tres casos:

1. *Uso de la expresión explícita de la recta de regresión* $y = mx + n$, siendo *m* la pendiente y *n* la ordenada en el origen de la recta de regresión de *y* sobre *x* (81 sujetos). Las condiciones expresadas en el enunciado no determinan de forma unívoca las variables dependiente e independiente. En consecuencia, una de las dificultades más frecuentes ha sido confundir la variable explicativa con la explicada (27 alumnos de los 50 que han llegado a una solución incorrecta a partir de la ecuación $y = mx + n$). Nueve estudiantes han hecho una interpretación inadecuada de los parámetros *m* y *n* en dicha ecuación, coincidiendo con una dificultad ya señalada en Truran (1997), conmutando sus valores, o bien, asignándoles valores no apropiados. Un ejemplo de respuesta es: *"y = n +mx, y = 4, a = 16, x = ?, 4 = 16 - 8x, 12 = 8x, x = 12 / 8 = 1.5"*. Las demás dificultades no son significativas debido a su baja frecuencia.
2. *Uso de la expresión punto-pendiente de la recta de regresión* (23 alumnos han empleado la expresión $y - \bar{y} = (S_{xy} / S_x^2)(x - \bar{x})$). Una dificultad es que, al tener la necesidad de conocer un punto de la recta para determinar esta ecuación, no se han dado cuenta que éste podría ser el correspondiente a la ordenada en el origen. Otra dificultad ha sido el uso de una expresión inadecuada de la recta de regresión *y sobre x* (7 estudiantes). También vuelve a aparecer la confusión entre las variables dependiente e independiente (8 estudiantes). Con este procedimiento, sin embargo, ha sido más fácil para los estudiantes interpretar cuál es la pendiente de la recta (S_{xy} / S_x^2), que dar un valor adecuado a la pendiente *m,* cuando emplean la ecuación explícita de la recta ($y = mx+n$).
3. *Otros procedimientos.* Dos alumnos se sirven de las dos expresiones anteriores de la recta de regresión de *y* sobre *x*: la ecuación explícita y la ecuación punto-pendiente. Estos dos alumnos no recordaban que el centro de gravedad pertenece a la recta de regresión, pero observan que la media de la variable dependiente se encuentra en la expresión de la ecuación punto-pendiente, comparan ambas expresiones y determinan la media de la variable explicada, llegando a una solución correcta, como se muestra en la siguiente respuesta:

$$\text{“}y\ sobre\ x\text{:}\ \ y = mx + n,\ \ y = 16x + 4,\ \ y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x}),\ \ m = \frac{S_{xy}}{S_x^2} = 16$$

$$n = \bar{y} - \frac{S_{xy}}{S_x^2}\bar{x},\ \ n = \bar{y} - 16.8,\ \ \bar{y} = 132\text{”}$$

Uso de la Recta de Regresión de X sobre Y

Trece alumnos, (10.1 % de los 129 que han respondido este problema), han utilizado la recta de regresión de *X* sobre *Y*. Dado que el problema se plantea en forma abierta, no se puede considerar que confundan la variable dependiente con la independiente. De ellos, 9 emplean la expresión de la ecuación explícita de la recta de regresión, $x = m'y + n'$, de los cuales sólo 2 responden de forma correcta. Los 4 restantes usan la ecuación punto-pendiente de la recta de regresión, $x - \bar{x} = (S_{xy}/S_y^2)(y - \bar{y})$, repitiéndose las dificultades encontradas anteriormente: confusión entre la variable explicativa y la explicada, intercambio de papeles entre la ordenada en el origen y pendiente de la recta de regresión o intercambio de parámetros de la ecuación explícita de la recta de regresión. Un ejemplo de respuesta de esta última dificultad es el siguiente:

$$\text{"}n' = \bar{x} - m'\bar{y},\ \ 16 = 8 - 4\bar{y},\ \ \bar{y} = \frac{-16 + 8}{4} = 2\text{"}$$

Uso de las dos Rectas de Regresión Y sobre X y de X sobre Y

Tres alumnos se sirven de ambas rectas de regresión, *y* sobre *x* y *x* sobre *y*. Podemos considerar que ésta es la mejor respuesta, pues, como indicábamos anteriormente, el problema se planteó de forma abierta y, por tanto, es plausible considerar a *x* como variable explicativa o como variable explicada.

Síntesis de dificultades

En resumen, podemos concluir que la dificultad más importante con la que se enfrentan los estudiantes de la muestra al resolver el problema 1 es la de discriminar entre la variable explicativa y la variable explicada, lo que ocurre a 38 alumnos (50% de los estudiantes que responden de forma incorrecta). Aunque las expresiones de la ecuación de la recta de regresión, punto-pendiente y explícita, son, obviamente, equivalentes, estos alumnos emplean como herramienta más operativa la punto-pendiente, pues en ella identifican mejor la pendiente que en la forma explícita; sin embargo, utilizan con más frecuencia la forma explícita, probablemente debido al enunciado del problema.

Finalmente, hemos de advertir que 7 alumnos de 129 intentan resolver este problema apoyándose en conceptos estadísticos relacionados con los datos suministrados; para ello usan parámetros estadísticos como, por ejemplo, la media, la covarianza, etc., siendo aplicados todos estos procedimientos de forma inadecuada.

4.2. CÁLCULO DEL COEFICIENTE DE CORRELACIÓN

Las respuestas de 168 estudiantes (87.05% de la muestra) al apartado a) del problema 2 se resumen en la tabla 2. La enseñanza recibida proporcionaba dos fórmulas para calcular el coeficiente de correlación: $r = S_{xy}/(S_xS_y)$ y el coeficiente de determinación $r^2 = m \cdot m'$, siendo m y m' los coeficientes de regresión lineal de las rectas *y sobre x* y *x sobre y*. Los estudiantes de esta muestra utilizan más frecuentemente la primera, ya que es más operativa.

Tabla 2. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 2.a según procedimiento de resolución

Procedimiento	Tipo de solución			Número de alumnos que usa el procedimiento
	Correcta	Incorrecta	No responde	
Fórmula $r = S_{xy} /(S_x S_y)$	106 (65.8)	55 (34.2)		161
Fórmula incorrecta de r		5 (100.0)		5
Coeficiente de determinación		2 (100.0)		2
No responde			25 (100)	25
Total	106 (54.9)	62 (32.1)	25 (13.00)	193

Aproximadamente un tercio de los 161 alumnos que han utilizado una fórmula correcta para calcular el coeficiente de correlación, han dado una respuesta inadecuada, resultados superiores a los de Morris (1997), quien sólo obtuvo un 20% de respuestas correctas, a pesar de dar la fórmula para calcular el coeficiente de correlación. Los principales errores se producen en los cálculos (22 de las 55 respuestas erróneas, lo que supone un 40 %, mientras que Morris obtuvo el 62.5% en su investigación); también ha influido un uso inadecuado de la calculadora o el uso de expresiones inadecuadas para el cálculo de la desviación típica y covarianza. Algunos alumnos dibujan un diagrama de dispersión para deducir el resultado pedido. Entre las argumentaciones ofrecidas es interesante reseñar las siguientes:

1. Tener conciencia de que el resultado obtenido no es correcto, ya que no está comprendido en el intervalo *[-1, 1],* pero no ser capaz de calcular el correcto. Una respuesta de este tipo es la siguiente: *"r = 2.24. Debe salir entre -1 ≤ r ≤ 1. Este dato debe estar mal".*
2. Reconocer que el resultado obtenido es incorrecto a la vista del diagrama de dispersión. Una respuesta de este tipo es la siguiente: *"Debe estar mal porque se ve claramente que existe correlación"* (cuando observa el diagrama de dispersión que el propio estudiante ha construido).

Las dos respuestas anteriores sugieren que el alumno conoce el algoritmo adecuado para responder a la pregunta formulada, pero no es capaz de realizarlo correctamente. A la vista de estos resultados, podemos concluir que la mayoría de los alumnos utiliza una expresión adecuada para calcular el coeficiente de correlación, que ciertos errores son debidos al cálculo, coincidiendo con los resultados de Morris (1997) y son pocos los errores debidos a conocimiento incorrecto sobre la correlación.

4.3. DETERMINACIÓN DE LA RECTA DE REGRESIÓN DE *Y* SOBRE *X*

Las respuestas de los 162 alumnos que han contestado a esta cuestión (83.9% de la muestra) se presentan en la tabla 3, junto con sus procedimientos. Los estudiantes han calculado sólo una recta de regresión (82.1 % del total de respuestas), o las dos (17.9 % del total de respuestas). Es de resaltar que la mitad de los estudiantes que responden (81 de 162) lo hacen de forma inadecuada, siendo ésta una actividad tan primordial para este tema.

De los 133 alumnos que calculan una sola recta de regresión, 97.0% utilizan la expresión de la ecuación punto-pendiente de la recta de regresión de *y* sobre *x,* mientras que el resto (3.0%) usan la recta de regresión de *x* sobre *y.* El 17.9 % hallan tanto la recta de regresión de *y* sobre *x* como la de *x* sobre *y*. Esto puede ser debido a que no saben claramente qué recta hay que determinar en este apartado o a que, como posteriormente - Problema 2 apartado e) -

necesitarán la recta de regresión de *x* sobre *y*, deciden obtenerla en este apartado, pero no hacen ninguna indicación al respecto. En cuanto a los errores que cometen son, esencialmente, similares a los expuestos cuando hemos tratado el cálculo del coeficiente de correlación.

Tabla 3. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 2c según procedimiento de resolución

Procedimiento	Tipo de solución			Número de alumnos que usa el procedimiento
	Correcta	Incorrecta	No responde	
Calcula la recta de regresión de y sobre x	81 (62.8)	48 (37.2)		129
Calcula la recta de regresión de x sobre y		4 (100.0)		4
Calcula ambas rectas de regresión		29 (100.0)		29
No responde			31(100)	31
Total	81 (42.0)	81 (42.0)	31 (16.0)	193

4.4. INTERPRETACIÓN Y PREDICCIÓN

Interpretación de la Relación entre las Variables

Los procedimientos que han empleado los estudiantes de la muestra para llevar a cabo el juicio de asociación pedido en el problema 2, apartado b) se han agrupado en 6 categorías, cuyas frecuencias y respuestas obtenidas se ofrecen en la Tabla 4.

Tabla 4. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 2b según procedimiento de resolución

Procedimiento	Respuesta			No interpreta	Número de alumnos que usa el procedimiento
	Dependencia directa (correcto)	Dependencia inversa	Independencia		
Coeficiente de correlación	46 (90.2)	1 (2.0)	3 (5.8)	1 (2.0)	51
Variación conjunta	17 (94.4)		1 (5.6)		18
Covarianza	17 (94.4)	1 (5.6)			18
Coeficiente determinación	3 (50.0)	1 (16.7)	2 (33.3)		6
Otras	3 (42.9)		3 (42.9)	1 (14.2)	7
Sin estrategia	57 (61.3)	2 (2.2)	5 (5.4)	29 (31.1)	93
Total	143 (74.1)	5 (2.6)	14 (7.2)	31 (16.1)	193

- *Uso del coeficiente de correlación*. Si el cálculo e interpretación son correctos, la respuesta es correcta, como la del siguiente estudiante: *La relación es directa ya que el coeficiente de correlación es positivo*. Aunque 168 alumnos han calculado el coeficiente de correlación en el apartado a), menos de un tercio de ellos, lo utilizan explícitamente para interpretar la relación entre las variables. Un hecho estadístico destacado es que 'la magnitud (fuerza) del coeficiente de correlación es por completo independiente de su dirección (positiva o negativa)" (Phillips, 1992, pp. 113). Sin embargo, unos pocos estudiantes confunden la intensidad y el sentido de la correlación, como se ve en la respuesta: *Existe una relación directa ya que el coeficiente de correlación es alto*. Probablemente estas respuestas se deban a que en la enseñanza se pone mucho énfasis en la correlación como medida de la intensidad de la relación, del mismo modo que se hacía en los libros de texto de bachillerato (Estepa y Sánchez-Cobo, 1998), lo que puede incidir en que los alumnos confundan magnitud con dirección de la correlación.

- *Variación conjunta de las variables.* Cuando el estudiante argumenta la existencia de relación al observar la variación paralela de las dos variables, como arguye este alumno: *Relación directa puesto que a medida que va pasando el tiempo, el número de bacterias aumenta*. Esta argumentación puede constituirse en un obstáculo, si el estudiante tiende a pensar que existe proporcionalidad entre las variables.
- *Uso de la covarianza*. Los estudiantes que usan la covarianza emplean el signo de ésta para decidir el sentido de la relación, como argumenta este alumno: *Es una relación directa ya que la covarianza es positiva.* Algunos estudiantes creen erróneamente que el coeficiente de correlación se utiliza para decidir si hay relación lineal y la covarianza para especificar el tipo de correlación existente, como manifiesta este estudiante: *A través del coeficiente de correlación se puede afirmar que existe una dependencia lineal alta. Para determinar si es directa o inversa se utiliza la covarianza* $S_{xy} = 23.8$. *Al ser positiva será directa.*
- *Uso del coeficiente de determinación*. Algunos estudiantes usan la bondad del ajuste como criterio para juzgar la relación, por ejemplo: *La relación es directa e intensa ya que* r^2 *se aproxima a 1*. Aquí aparece nuevamente la confusión entre intensidad y dirección de la relación, un estudiante responde: *Relación inversa ya que el valor de* $r^2 = 0.4568$ *no está próximo a 1.*
- *Otras estrategias*. En esta categoría se incluyen diversos procedimientos y argumentaciones, como utilizar un intervalo de variación del coeficiente de correlación, usar el diagrama de dispersión o comparar la asociación y la proporcionalidad.
- *Sin estrategia*. En esta categoría el estudiante no justifica o razona la respuesta, aunque como hemos comentado al principio de esta sección, es probable que hayan utilizado el coeficiente de correlación, como parece deducirse de la siguiente respuesta: *La relación que existe entre las dos variables es directa con intensidad fuerte.*

Estepa y Batanero (1996) pidieron a estudiantes que no había recibido instrucción sobre la correlación que estimasen el tipo de relación existente entre dos variables, dado el diagrama de dispersión, obteniendo un 47.5% de respuestas correctas. En nuestro caso la proporción de respuestas correctas es muy superior (74.1%), y la mayoría de estudiantes interpretó correctamente la correlación, a pesar que los datos se dan en forma de tabla, aunque los estudiantes habían recibido instrucción sobre la correlación. La justificación más utilizada se ha basado en el coeficiente de correlación, seguido de la covarianza y la variación conjunta. Es de destacar la confusión entre el sentido e intensidad de la correlación que manifiestan algunos alumnos.

Predicción de los Valores de la Variable Explicada a partir de la Variable Explicativa

El apartado 4 d) del problema 2 se refiere al uso de la recta de regresión para predecir, cuando se debe hacer una interpolación.

Los procedimientos de resolución utilizados en las respuestas a la pregunta sobre interpolación se exponen en la tabla 5. Han respondido 162 alumnos (83.93% del total). Aunque el valor obtenido de los cálculos es 31.625 bacterias, hemos considerado como respuesta correcta 31 ó 32 bacterias, ya que la respuesta a la pregunta formulada debe tener sentido en el contexto del problema, considerando erróneas las respuestas que ofrecen un valor decimal, lo que explica las pocas respuestas correctas, ya que los alumnos han prestado menos atención al contexto y al resultado que ofrecen, que al modelo estadístico y los cálculos.

Tabla 5. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 2.d según procedimiento de resolución

Procedimiento	Tipo de respuesta			Número de alumnos que usa el procedimiento
	Correcta	Incorrecta	No responde	
Recta regresión y sobre x	5 (3.5)	137 (96.5)		142
Recta regresión x sobre y		8 (100)		8
Uso de proporcionalidad		10 (100)		10
No especifica estrategia		2 (100)		2
No responde			31 (100)	31
Total	5 (2.6)	157 (81.3)	31 (16.1)	193

Se puede observar otra vez la confusión entre las variables dependiente e independiente, en los alumnos que utilizan la recta de regresión de *x sobre y*. En otros casos se expresa una creencia en la existencia de proporcionalidad entre las variables, que se manifiesta por el uso de una regla de tres, como en la siguiente respuesta: *Si a 2 le corresponde 21 a 2.5 le corresponderá x.* Otro caso interesante es el alumno que observa que 2.5 es el punto medio del intervalo [2,3] y razona que la respuesta estará en el punto medio del intervalo correspondiente [21,33].

Predicción a partir de la Recta de Regresión de X sobre Y

Tal y como están planteadas las preguntas del problema 2, el objetivo de la cuestión e) era indagar si el alumno es consciente de la existencia de las dos rectas de regresión y si hace un uso adecuado de ellas, pues en Batanero, Godino y Estepa (1991, 2001) algunos estudiantes no discriminaban estas dos rectas En la tabla 6 mostramos los resultados.

Tabla 6. Frecuencia y (porcentaje) de soluciones correctas e incorrectas al problema 2.e según procedimiento de resolución

Procedimiento	Tipo de respuesta			Número de alumnos que usa el procedimiento
	Correcta	Incorrecta	No responde	
Recta regresión x sobre y	49 (50.5)	46 (47.4)		95
Recta regresión y sobre x		45 (100.0)		45
Uso de proporcionalidad		9 (100.0)		9
Otras estrategias		2 (100.0)		2
No especifica estrategia		2 (100)		2
No responde			40 (100)	40
Total	49 (25.4)	104 (53.9)	40 (20.7)	193

153 estudiantes (79.3% de la muestra), responden a la pregunta, la mayoría de los cuales (62.1%) utilizan la recta *x sobre y,* y en consecuencia han discriminado las dos rectas de regresión y la aplican de manera pertinente, salvo en errores de cálculo o a usar una expresión incorrecta de la recta de regresión.

Los que usan la recta de regresión *y sobre x* no son conscientes de la existencia de dos rectas diferentes ni de la utilidad de la recta de regresión de *x sobre y*. Los que emplean un procedimiento proporcional, tienen la creencia de que la recta de regresión es una aplicación lineal, hecho que, en general no es correcto, y suelen establecer una regla de tres. En consecuencia, se aprecia un razonamiento proporcional con ciertas carencias. Dentro de éstos, hay una minoría que comparan el resultado obtenido en el apartado d) y, sin hacer ningún cálculo, responden que el tiempo que deberá transcurrir es de 2.5 horas, el mismo que el ofrecido en dicha cuestión. Esto es indicativo de una concepción determinista de la dependencia aleatoria (Estepa y Batanero, 1996) además de no discriminar las dos rectas de regresión.

5. SÍNTESIS DE RESULTADOS E IMPLICACIONES

En este estudio exploratorio hemos analizado los procedimientos de los estudiantes de la muestra cuando se enfrentan a la resolución de los problemas simples de regresión y correlación, destacando sus aciertos, errores y dificultades, cuyo análisis nos ha proporcionado una caracterización del conocimiento de los estudiantes sobre conceptos relacionados con la correlación y regresión analizados, al finalizar un curso de Estadística Descriptiva en los primeros cursos universitarios.

Casi las dos tercera partes de los estudiantes calcula correctamente el coeficiente de correlación, mientras que sólo la mitad llegan a una expresión correcta de la recta de regresión. A pesar de los errores de cálculo o uso de fórmulas incorrectas, muchos estudiantes reconocen que sus resultados no son correctos a la vista del diagrama de dispersión o de la magnitud del valor obtenido para el coeficiente de correlación. Tres de cada cuatro alumnos dan un juicio de asociación correcto, usando diversas estrategias correctas para tomar la decisión (coeficiente de correlación, covarianza, variación conjunta).

Los estudiantes tienen dificultades al tratar de relacionar conceptos matemáticos y estadísticos en la resolución del problema 1, por ejemplo, al interpretar los parámetros m y n en la ecuación de la recta, la ordenada en el origen o su centro de gravedad; estas dificultades disminuyen la capacidad de utilizar la recta de regresión con eficacia.

Un porcentaje elevado de estudiantes conoce el procedimiento matemático para efectuar las predicciones, pero ha habido errores de cálculo y sobre todo no se ha tenido en cuenta el contexto. Por otra parte, la confusión entre las variables dependiente e independiente da lugar, de manera sistemática y persistente a una serie de errores que dificulta la resolución adecuada de este tipo de problemas.

Las confusiones entre el sentido y la intensidad de la correlación, entre la variable dependiente e independiente (en varias cuestiones), o bien, la creencia de que la variación conjunta es proporcional son dificultades importantes a tener en cuenta en la planificación de la enseñanza.

A la vista de estos resultados, podemos concluir que este tema es complejo y una enseñanza deficiente puede dar lugar a concepciones erróneas y confusiones que obstaculicen una comprensión amplia del mismo, tan necesaria, tanto por su utilidad, como por ser requisito para seguir profundizando en el conocimiento estadístico. Estas dificultades nos alertan sobre la necesidad de reforzar la enseñanza de este tema, a través de situaciones didácticas apropiadas que permitan superar las dificultades observadas. Consideramos que sin una plena integración e interrelación de los diferentes conceptos y su interpretación dentro del contexto del problema, difícilmente los futuros graduados podrán llevar a cabo una práctica correcta de la estadística.

Dado que el estudio llevado a cabo es de tipo exploratorio, no podemos concluir con soluciones definitivas a los problemas detectados en el aprendizaje del tema, sino que estimamos que algunos de los resultados expuestos aquí precisan de nuevas investigaciones para quedar convenientemente clarificadas las causas de los errores, dificultades y concepciones erróneas que hemos detectado y, en consecuencia, poder proponer soluciones a los problemas encontrados en la adquisición de conocimiento de los alumnos. Esto es especialmente importante puesto que estudios previos sugieren que los adultos tienen escasa capacidad intuitiva para estimar la correlación, salvo cuando ésta es muy fuerte y confirma sus teorías previas.

Agradecimientos: Este trabajo forma parte del Proyecto PB97-0851 (Ministerio de Ciencia y Tecnología, Madrid).

REFERENCIAS

Barbancho, A. G. (1973). *Estadística elemental moderna*. Barcelona: Ariel (Cuarta edición. Reimpresión de 1975).

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

Batanero, C., Godino, J. D. y Estepa, A. (2001). La construcción del significado de la asociación mediante actividades de análisis de datos: Reflexiones sobre el papel del ordenador en la enseñanza de la Estadística. En J. R. Pascual (Eds.), *II Simposio de la Sociedad Española de Investigación en Educación Matemática (SEIEM),* (pp. 151-169). Pamplona: Universidad Pública de Navarra.

Cook, T. D. y Campbell, D. T. (1979). *Quasi-experimentation. Design & Analysis Issues for Field Settings*. Chicago: Rand M^{c}.Nally.

Cruise, J. R., Dudley, R. L. y Thayer, J. D. (1984*). A Resource Guide for Introductory Statistics*. Dubuque, IO: Kendall / Hunt.

Erlick, D. E. y Mills, R. G. (1967). Perceptual quantification of conditional dependency. *Journal of Experimental Psychology*, 73(1), 9-14.

Estepa, A. y Batanero, C. (1996). Judgments of correlation in scatter plots: student's intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 25-41.

Estepa, A. y Sánchez-Cobo, F. T. (1998). Correlation and regression in secondary school text books. En L. Pereira-Mendoza, L. Seu, T. Wee y W.K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (vol. 2, pp. 671-676). Voorburgo (Holanda): International Statistical Institute.

Estepa, A. y Sánchez-Cobo, F. T. (2001). Empirical research on the understanding of association and implications for the training of researchers. En C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 37-52). Granada: International Association for Statistical Education e International Statistical Institute.

Hildebrand, D. K., Lang, J. D. y Rosenthal, H. (1977) *Analysis of Nominal Data*. Londres: Sage University Paper.

Huberman, A. M. y Miles, M. B. (1994). Data management and analysis methods. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 445-462). Londres: Sage.

Jennings, D. L., Amabile, T. M. y Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases,* (p. 211-230). Nueva York: Cambridge University Press.

Lane D. M., Anderson, C. A. y Kellam, K. L.(1985). Judging the relatedness of variables: The psychophisics of covariation detection. *Journal of Experimental Psychology. Perception and Performance*, 11(5), 640-649.

Miles, M. B. y Huberman, A. M. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods.* Londres: Sage Publications.

Morris, E. J. (1997). *An investigation of students' conceptions and procedural skills in the statistical topic correlation*. Londres: The Open University, Centre for Information Technology in Education, Informe n. 230.

Moritz, J. (2002). Conflicting representations of statistical association. *Statistics Education Research Journal,* 1(1), 39-40.

Phillips, J. L. (1992). *How to Think about Statistics*. Nueva York: W.H. Freeman.

Sánchez-Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Departamento de Didáctica de la Matemática. Universidad de Granada.

Sánchez-Cobo, F. T., Estepa, A. y Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, 18(2), 297-310.

Truran, J. M. (1997). Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. En J. B. Garfield y J. M. Truran (Eds.), *Research Papers on Stochastics Educations from 1997* (pp. 205-212). Universidad de Minnesota.

Vizmanos, J. R. y Anzola, M. (1988). *Matemáticas II. Opción C: Ciencias Sociales. Opción D: Humanística / Lingüística.* Madrid: SM.

ANTONIO ESTEPA
Universidad de Jaén
Departamento de Didáctica de las Ciencias
Campus Las Lagunillas s/n
23071 Jaén

RECENT PUBLICATIONS

Aberson, C. L., Berger, D. E., Healy, M. R. & Romero, V. L. (2002). An interactive tutorial for teaching statistical power. *Journal of Statistics Education,* 10(3).

This paper describes an interactive Web-based tutorial that supplements instruction on statistical power. This freely available tutorial provides several interactive exercises that guide students as they draw multiple samples from various populations and compare results for populations with differing parameters (for example, small standard deviation versus large standard deviation). The tutorial assignment includes diagnostic multiple-choice questions with feedback addressing misconceptions, and follow-up questions suitable for grading. The sampling exercises utilize an interactive Java applet that graphically demonstrates relationships between statistical power and effect size, null and alternative populations and sampling distributions, and Type I and II error rates. The applet allows students to manipulate the mean and standard deviation of populations, sample sizes, and Type I error rate. Students (n = 84) enrolled in introductory and intermediate statistics courses overwhelmingly rated the tutorial as clear, useful, easy to use, and they reported increased comfort with the topic of statistical power after using the tutorial. Students who used the tutorial outperformed those who did not use the tutorial on a final exam question measuring knowledge of the factors influencing statistical power.

Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In Phillips, B. (Ed.), *Proceedings of the Sixth International Conference of Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute. CD-ROM.

This paper contrasts two types of educational tools: a route-type series of so-called statistical minitools (Cobb et al., 1997) and a landscape-type construction tool, named Tinkerplots (Konold & Miller, 2001). The design of the minitools is based on a hypothetical learning trajectory (Simon, 1995). Tinkerplots is being designed in collaboration with five mathematics curricula and is open to different approaches. Citing experiences from classroom-based research with students aged ten to thirteen, I show how characteristics of the two types of tools influence the instructional decisions that software designers, curriculum authors, and teachers have to make.

Bakker, A. (2003). The Early History of Average Values and Implications for Education. *Journal of Statistics Education*, 11(1),

http://www.amstat.org/publications/jse/v11n1/bakker.html

The early history of average values is used as a source of inspiration for instructional materials for seventh-grade classrooms. This historical study helped to define different layers, aspects, and applications of average values and to look through the eyes of the students, who do not have the same concepts as teachers and instructional designers have. As a result of this study, a few possible implications for education have been tried out, such as estimation as a starting point for a statistics course, allowing the midrange as an initial strategy, a visual way of estimating the mean using bar representations in a simple computer tool, and the reinvention of midrange, median, mode, and mean. There turn out to be striking parallels but also important differences between the historical and students' individual development of statistical understanding.

Broers, N.J. (2002). Selection and use of prepositional knowledge in statistical problem solving. *Learning and Instruction,* 12(3), 323-344.

Central in this study is the question why subjects who possess the necessary factual or propositional knowledge needed to solve a particular statistical problem, often fail to find the solution to that problem. 10 undergraduate psychology students were trained so as to possess all the relevant knowledge needed to solve five multiple choice problems on descriptive regression analysis. They were asked to think aloud while attempting to solve the problems. Analysis of the think-aloud protocols showed that a failure to select the relevant information in the text, together with a failure to retrieve relevant propositional knowledge from memory and a difficulty with logical reasoning combined to produce incorrect responses. Factual knowledge was less likely to be successfully retrieved when it was acquired only recently or when it concerned relationships of a highly abstract nature. Furthermore, the existence of misconceptions proved to inhibit the use of correct factual knowledge.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education,* 10(3).

This paper focuses on a third arm of statistical development; statistical thinking. After surveying recent definitions of statistical thinking, implications for teaching beginning students (including non-majors) are discussed. Several suggestions are given for direct instruction aimed at developing “habits of mind” for statistical thinking in students. The paper concludes with suggestions for assessing students’ ability to think statistically. While these suggestions are primarily aimed at non-majors, many statistics majors would also benefit from further development of these ideas in their undergraduate education.

delMas, R. C. (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education,* 10(3).

Similarities and differences in the articles by Rumsey, Garfield and Chance are summarized. An alternative perspective on the distinction between statistical literacy, reasoning, and thinking is presented. Based on this perspective, an example is provided to illustrate how literacy, reasoning and thinking can be promoted within a single topic of instruction. Additional examples of assessment items are offered. I conclude with implications for statistics education research that stem from the incorporation of recommendations made by Rumsey, Garfield and Chance into classroom practice.

Friedman, H. H., Friedman, L. W. & Amoo, T. (2002). Using humor in the introductory statistics course. *Journal of Statistics Education,* 10(3).

This paper discusses reasons for using humor in the statistics classroom. Humor strengthens the relationship between student and teacher, reduces stress, makes a course more interesting, and, if relevant to the subject, may even enhance recall of the material. The authors provide examples of humorous material for teaching students such topics as descriptive statistics, probability and independence, sampling, confidence intervals, hypothesis testing, and regression and forecasting. Also, some references, summarized strategies, and suggestions for becoming more humorous in the classroom are provided.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education,* 10(3).

This paper defines statistical reasoning and reviews research on this topic. Types of correct and incorrect reasoning are summarized, and statistical reasoning about sampling distributions is examined in more detail. A model of statistical reasoning is presented, and suggestions are offered for assessing statistical reasoning. The paper concludes with implications for teaching students in ways that will facilitate the development of their statistical reasoning.

Ludlow, L. H. (2002). Rethinking practice: Using faculty evaluations to teach statistics. *Journal of Statistics Education,* 10(3).

This article explains why and how a course in general linear models was restructured. This restructuring resulted from a need to more fully understand traditional teaching evaluations, coupled with a desire to introduce more meaningful data into the course. This led to the incorporation of a longitudinal dataset of teaching evaluations into the lecture material and assignments. The result was a deeper appreciation of how students perceive my teaching, specifically, and a greater understanding of how statistics courses, in general, can be taught more effectively.

Paparistodemou, E., Noss, R. & Pratt, D. (2002). Exploring in sample space: Developing young children's knowledge of randomness. In B. Phillips (Ed.), *Proceedings of The Sixth International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute. CD-ROM.

This paper focuses on 6-8 year-old children's thinking about randomness. It reports the findings of a study in which the children engaged with a game-like environment in which they could construct for themselves spatial representations of sample space. The system was designed so that the rules governing the relationship between the selection of elements of the sample space and the outcomes of the game were available for inspection and reconstruction by the children. In response to a range of tasks, the children manipulated the sample space in ways that generated corresponding outcomes to the game. We present a case study of children's activities, which illustrates how the novel medium mediates the children's expression and understanding of chance events.

Pratt, D. (2002). Probability and randomness. In Haggerty, L. (Ed.), *Aspects of Teaching Secondary Mathematics: Perspectives on Practice*. Milton Keynes, UK: Open University.

This is a chapter aimed at providing teachers of mathematics with a research-based rationale for an approach to teaching probability. Evidence is offered that probabilistic ideas are not well understood amongst the population at large. The chapter then focuses on some research-based ideas and principles to guide the teaching of the fundamental ideas in the England and Wales National Curriculum.

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

In this paper we explore the growth of mathematical knowledge and in particular, seek to clarify the relationship between abstraction and context. Our method is to gain a deeper appreciation of the process by which mathematical abstraction is achieved and the nature of abstraction itself, by connecting our analysis at the level of observation with a corresponding theoretical analysis at an appropriate grain size. In this paper we build on previous work to take a further step towards constructing a viable model of the micro-evolution of mathematical knowledge in context. The theoretical model elaborated here is grounded in data drawn from a study of 10-11 year olds' construction of meanings for randomness in the context of a carefully designed computational microworld, whose central feature was the visibility of its mechanisms - how the random behavior of objects actually "worked". In this paper, we illustrate the theory by reference to a single case study chosen to illuminate the relationship between the situation (including, crucially, its tools and tasks) and the emergence of new knowledge. Our explanation will employ the notion of situated abstraction as an explanatory device that attempts to synthesize existing micro- and macro-level descriptions of knowledge construction. One implication will be that the apparent dichotomy between mathematical knowledge as de-contextualized or highly situated can be usefully resolved as affording different perspectives on a broadening of contextual neighborhood over which a network of knowledge elements applies.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education,* 10(3).

In this paper, I will define statistical literacy (what it is and what it is not) and discuss how we can promote it in our introductory statistics courses, both in terms of teaching philosophy and curricular issues. I will discuss the important elements that comprise statistical literacy, and provide examples of how I promote each element in my courses. I will stress the importance of and ways to move beyond the "what" of statistics to the "how" and "why" of statistics in order to accomplish the goals of promoting good citizenship and preparing skilled research scientists.

Zimmermann, G. M., & Jones, G. A. (2002). Probability simulation: What meaning does it have for high school students? *Canadian Journal of Science, Mathematics, and Technology Education*, 2, 221-236.

This study investigated high school students' reasoning and beliefs when confronted with contextual tasks involving the assessment and construction of two-dimensional probability simulations. Nine students enrolled in an advanced algebra course, with little formal instruction in probability, engaged in clinical interviews focusing on the simulation tasks. All students showed evidence of being able to identify an appropriate probability generator to model a contextual problem. However, their thinking in probability simulation was constrained by their inability to deal with two-dimensional trials. In assessing the validity of a given simulation, only one student was able to identify a flaw that resulted from the use of one-dimensional trials rather than two-dimensional trials. Additionally, when asked to construct a simulation, only two students were able to define an appropriate two-dimensional trial and develop a valid solution. The study also revealed evidence of students' beliefs about probability simulation - some of which could be helpful in informing instruction, others problematic.

RECENT DISSERTATIONS

Baloglu, M. (2001). *An application of structural equation modeling techniques in the prediction of statistics anxiety among college students.* PhD Texas A&M University - Commerce. Supervisor: Paul F. Zelhart.

No general theory has been formulated to show interrelations among a collection of variables that are related to statistics anxiety. The present study made an attempt to develop a comprehensive model that would predict statistics anxiety from several dispositional, situational, and environmental antecedents derived from the current literature. Two hundred forty-six college students who were enrolled in introductory statistics courses completed a survey packet that included a set of questions and five standardized assessment instruments that measured statistics anxiety, mathematics anxiety, attitudes toward statistics, test anxiety, and general anxiety. Extensive preliminary data screening assured the appropriateness of the data for parametric statistics. Independent - test results showed significant differences between low-and-high anxious students in terms of attitudes toward statistics, test anxiety, mathematics anxiety, general anxiety, previous mathematics experience, satisfaction, and pace. A direct discriminant function analysis was used to discriminate between low-and-high statistics-anxious students. A significant discriminant function, based on the attitudes toward statistics and test anxiety, classified the groups accurately approximately 80% of the time. Five measurement models and one structural model were specified, identified, estimated, and tested. Results showed that the modified structural model did not fit the data well. However, the dispositional and situational antecedents models fit the data well. The original environmental antecedents model was modified to fit the data. In the final model, the dispositional and situational antecedents models contributed significantly to statistics anxiety. The environmental antecedents model was not a significant contributor; however, it was significantly related to the other variables in the model. The dispositional antecedents model alone accounted for 58% and the situational antecedents model alone accounted for 23% of the variance in statistics anxiety scores. The present study showed that statistics anxiety is a complicated construct that is difficult to measure and investigate. Findings of the present study also suggest that personality-related factors may be one of the most important effects of statistics anxiety. More studies are needed to clarify the construct of statistics anxiety and its relationships with other variables.

Espasandin Lopes, C. A. (2003). *O conhecimento profissional dos professores e suas relações com estatística e probabilidade na educação infantile* (Teacher's profesional knowledge in relation to statistics and probability in primary school education). PhD University of Campinas (Brazil). Supervisor: Anna Regina Lanner de Moura.

This is a collaborative research linking researcher and teachers working together at schools. Our assumption is that teachers' professional knowledge, resulting from the integration of theory and practice, is personal and can be mainly observed from teacher's work while developing the curriculum. Moreover, teachers' professional development emerges when they intentionally engage in an educational project, and reflect on their practice, as individuals and as a group. We also took into account the relevance of researchers' participation, professional knowledge, questioning capacity and empathy when encouraging teachers to learn more about themselves and their practice.

Freire's concept of reflective teacher was taken into account to investigate the contributions that Statistics and Probability concepts can bring to a group of Kindergarten teachers' professional education and pedagogical practice in a private school in Campinas. A planned intervention led to collaborative production, which allowed the enlargement of professional knowledge concerning Mathematics and Statistics, the curriculum, and the teaching learning process for these teachers. Information was collected along three years of teaching, using questionnaires, interviews, papers, the researcher's notes, collective discussion of texts, videotaped lessons and analysis of activities planned and performed by the teachers. Case study of teachers and coordinators participating in the experience when trying to identify important points in their mathematics, statistics and didactics knowledge and their professional development were used.

Main conclusions were: The curriculum knowledge was linked to teachers' conception about the meaning of statistics and probability in Kindergarten education. Teachers were aware of the curricula goals, and were capable to design projects related to their context. Their didactical knowledge was visible in solving the problems and in the diversity of strategies and solutions. Their professional development increased through their ethical and solidary work when jointly producing concept developments and didactics knowledge for Mathematics and Statistics. Consequently, we suggest an educational process that values the teachers' knowledge, challenges their reasoning, encourage teachers' research on their own practice and allow them to contribute to collective production of knowledge.

Miller, G. F. (2001). *The relationship between college student learning styles and assessment methods in elementary statistics.* EdD Columbia University Teachers College. Supervisor: Phillip J. Smith.

This study investigated the relationship between students' learning styles and their choice of grade weightings for and performance on three classroom assessment instruments. The student learning styles were measured using McCarthy's Learning Type Measure. Students chose weightings for a take home examination, a research project, and an in-class examination. The sample for the study consisted of 44 students in two sections of an elementary statistics course at an urban community college. Using analysis of variance of the weighting and performance data when organized by student learning type, the investigator found no significant relationship between learning types and the grade weightings and no significant relationship between learning types and performance on the assessment instruments. There was a significant positive correlation between the score for Type 2 learners and the performance on the in-class examination and also a significant positive correlation between the score for Type 4 learners and the weight assigned for the project. Both of these correlations validate the characteristics of these learning types as they are conceptualized by McCarthy, Kolb, and others. Type 4 learners are generally risk takers and learn by perceiving through concrete experience and processing through active experimentation. Type 2 learners generally perceive information through reflective observation data, the investigator conducted clinical interviews to ascertain differences between students' rationales for their grade weightings and how student learning styles may have affected students' performances on the various assessment instruments.

Sharkey, C. K. (2001). *Secondary school students conceptions, factors behind achievement, and problem solving strategies with stochastic problems.* PhD The American University. Supervisor Lyn Stallings.

The purpose of this study was to investigate the conceptions secondary school students have when dealing with stochastic questions and the heuristics these students use to solve stochastic questions. The second purpose of this study was to determine if there were any effects of gender, grade level, mathematical placement, reading ability and prior stochastic experience on the students' stochastic achievement. The students' stochastic achievement was based on the percentage correct on a multiple-choice stochastic test and the students' conceptions and the heuristics they used were based on the answers students gave on a stochastic reasoning test. The analysis sample for the study consisted of 392 secondary school mathematics students in the Toms River, New Jersey, school district who took the multiple-choice stochastic test. Eighteen of the 392 students volunteered to take the reasoning test, where six students were from each group of students who scored in the top third, middle third and bottom third of the multiple-choice test. Statistical methods were used to test if there were any effects of the variables mentioned earlier on students' stochastic achievement, and whether there was a difference in the proportion of correctly answered questions on the multiple-choice test between probability and statistics questions. The results indicated that, at the 0.05 significance level, reading ability, grade level (Grade 9), the interaction between gender and mathematical placement (track 3), and the interaction between reading ability and stochastic experience had a significant effect on students' stochastic achievement. In addition, there was a significant difference in the proportion of correct answers between probability and statistics questions. Another question that was investigated in this study was if secondary school students use heuristics to solve stochastic questions. This question was qualitatively researched. From the results of the reasoning test, it was concluded that secondary school students use the following heuristics to solve stochastic problems: Belief Strategy, Equiprobable Bias, Bigger is Better, Prior Experience and Normative Reasoning. Belief strategy was used more often than the other heuristics. Also, it was determined that students do not always use the same heuristics to solve similar types of problems.

INFORMATION ON PAST CONFERENCES

1. INTER-AMERICAN CONFERENCE ON TEACHING OF STATISTICS

The *Inter-American Conference on Teaching of Statistics* (Jornadas Interamericanas de Enseñanza de la Estadística-JIEE), jointly organized by IASI and the Universidad Nacional de Tres de Febrero (UNTREF), of Argentina, was held in Caseros, Buenos Aires, Argentina on October 28-November 1, 2002. This Meeting was sponsored by the International Association for Statistical Education (IASE).

The Conference was linked to the Fifth Latin American Congress of Statistical Societies (CLATSE V), organized by the Argentine Society of Statistics (SAE), the Argentine Group of Biometry (GAB), the Chilean Society of Statistics (SOCHE), and the Uruguayan Society of Mathematics and Statistics (SUME). About 500 people, from Argentina, Brazil, Canada, Chile, Honduras, Italy, Spain, Panama, United States, and Uruguay, attended these events.

Plenary lectures:

Orual Andina: The Teaching of Statistics in Public Offices,

Carmen Batanero: The Challenges of the Statistical Culture

Estela Dagum: Some Thoughts on University Statistical Education for the XXI Century

Pedro Silva: The Teaching of Statistics for Modern Quantitative Research

Short courses

Pedro Silva: Development of Sampling Techniques

Martha Bilotti-Aliaga: Interactive Statistics

2. STOCHASTIC WORKING GROUP AT CERME III

DAVE PRATT
University of Warwick
dave.pratt@warwick.ac.uk

The third conference of the European Society for Mathematics Education took place at Bellaria, Italy, from 28th February to 3rd March. CERME has a distinctive style in that it encourages discussion rather than presentation of papers. In order to facilitate this approach, the conference was organised into working groups, which spent 12 hours together considering and discussing papers previously read by the participants.

This issue of SERJ reports on the work of the working group on Stochastic Thinking, which incorporated issues pertaining to the teaching and learning of both probability and statistics, as well as the interface between. This group was led by four organisers: Dave Pratt (UK, Chair), Carmen Batanero (Spain), Rolf Biehler (Germany) and Michel Henry (France). 17 papers were accepted and discussed within this working group. They represented work from authors spread across four continents and ten countries.

The working group began with an ice-breaker, planned and led by Dor Abrahamson (USA). We were required to collect data about each other, a process which succeeded in helping us to get to know each other but which also raised some interesting issues about emergent behaviour. Perhaps because of this ice-breaker, the group subsequently engaged in productive discussion, a central aim of the conference. There were two special sessions; one involving the hands-on use of computer software, namely Fathom (www.keypress.com/fathom/) and NetLogo (ccl.northwestern.edu/netlogo/), and another final session in which the working group divided into smaller teams with the task of pulling together the main threads of the previous discussion.

The other four sessions were divided into four themes: Teacher Education, Computer-based Tools, Statistical Thinking and Probabilistic Thinking. In each session, papers relevant to that theme were briefly re-introduced with the aim of reminding the group about their key ideas. The aim was that most of the session was devoted to asking specific clarification questions or raising important discussion points. In most sessions that aim was achieved though sometimes there was more presentation and less discussion than had been envisaged.

Some important conclusions were drawn. The Teacher Education theme focussed on the impoverished nature of training for teachers of statistics, who were often not especially knowledgeable in that area. There was also concern about the accuracy of many text books on statistics.

The work on Computer-based Tools highlighted the importance of technology for the teaching and learning of stochastics. New approaches for the spatial dynamic representation of stochastic ideas were discussed, and the potential for students to appreciate probability distribution as an emergent phenomenon was acknowledged as exciting and novel.

The Statistical Thinking theme reported on studies into the mean and graphing. There was also discussion about the important role that the construction of a task and the subsequent social interaction has on the quality of observed statistically-related discussion.

Finally the work on Probabilistic Thinking put forward new theoretical perspectives and evidence that recognised the context-sensitive nature of student's probabilistic thinking. This research also proffered the view that a modelling perspective on probability was especially accessible through advances in technology.

At the end of the conference, there was a feeling that much progress had been made, not just in terms of the discussion content, but also in terms of establishing a new network that will support research in this important area in the future. In the immediate future, the working group will continue to improve through email collaboration its 17 papers, which will eventually be published on the internet (fibonacci.dm.unipi.it/~didattica/CERME3/) and on CD. We look forward to the continued work of this group at the next CERME conference.

Papers presented

The following papers were presented at the stochastic group and will be published in the conference proceedings:

Abrahamson, D. & Wilensky, U. The quest of the bell curve: A constructionist designer's advocacy of learning through designing.

This paper introduces the rationale, explains the functioning, and describes the process of developing 'Equidistant Probability', a NetLogo microworld that models stochastic behavior. In particular, we detail the phases in attempting to choose suitable parameters and create such graph displays as will permit an observer to witness the incremental growth of a bell-shaped curve. Drawing on a constructionist perspective (e.g., Papert, 1980), the paper argues that the process of building the model, and in particular the accountability, motivation, and frustration experienced, were conducive to 'connected learning' (Wilensky, 1993), through which the design of this microworld is grounded. The microworld is part of a larger middle-school curriculum on "understanding complexity".

Alldredge, J. R. & Brown, G. Association of course performance with attitudes and beliefs: An analysis by gender and instructional software environment.

We desired to improve student learning in our introductory, algebra-based statistical methods course. Marketing claims as well as anecdotal evidence suggested that electronic forms of educational material improve student learning. We desired to assess how the use of instructional software influences, and is influenced by, student attitude and if these influences are related to gender.

Azcárate, P., Serradó, A. & Cardeñoso, J. M. Hazard's treatment in secondary school.

The research work that we present, is framed in a investigation agenda whose aim is to study the teaching and learning process of the probabilistic knowledge, analysed from the perspective of the teacher; his ideas, knowledge, capacities and instruments that he uses when planning and developing his teaching skills in the field of mathematical knowledge.

Batanero, C., Cobo, B. & Díaz, C. Assessing secondary school student's understanding of averages.

In this paper we describe results from a questionnaire given to a sample of secondary school students to assess the personal meaning they attribute to mean, median, and mode. The questionnaire is made up of 9 open-ended tasks (26 sub items) where students provide detailed reasoning to their responses. Comparative results from two samples of 14 year-olds (n=168) and 16 year-olds (n=144) and multivariate analysis for the combined sample will be analysed.

Callaert, H. In search of the specificity and the identifiability of stochastic thinking and reasoning.

The unique format of the CERME conference invites interdisciplinary collaboration in an intense and personalized way. Research on stochastic thinking and reasoning, with its implications on the teaching of statistics, benefits from professional input from psychology and education as well as from statistics. That's why statisticians, even with no formal training in educational psychology, but with a concern about the teaching of statistics, should be involved in the discussion. Reading research papers on stochastic thinking and reasoning, one wonders which processes can be identified as being specific and unique. Such identification could be helpful, possibly leading to guidelines for optimizing statistical teaching strategies. Through a couple of examples, this paper hopes to stimulate mutual discussion.

Cañizares, M. J., Batanero, C., Serrano, L. & Ortiz, J. J. Children's understanding of fair games.

This paper analyses the responses given by children from two samples (n=320; n=147) to two test items concerning the fairness for a game of chance. We study the influence of age and mathematical ability on the percentage of correct responses. Interviews with a small sample of pupils serve to describe children's conceptions of fair games.

Carvalho, C. Solving strategies in statistical tasks.

The role of Statistics is becoming increasingly important in today's society. Collaborative work has shown to be one of the most adapted forms of facilitating knowledge appropriation and the mobilisation of competencies. In a school context we are aware of it. It is a fact that individuals construct explanations and solving strategies by themselves and also when interacting with others. The question is to understand how it works and how we can benefit from it in our classes. This is a research challenge. Responding to the challenge, we analysed the protocols of 136 dyads of a quasi-experimental study. One of the main results show five resolution strategies used by the dyads. In this paper we will discuss two of them.

Karian, Z. Z. A new approach to probability and statistics instruction.

Symbolic manipulators provide a dynamic educational environment that can increase the productivity of novices as well as expert statisticians. They enable students to ask for a random sample of any size, compute various sample statistics, plot a histogram, see the relationship between the sample statistics and population parameters or the relationship between the sample histogram and the probability histogram—all without interrupting the flow of the reasoning process. Such flexibility develops deeper insights and the opportunity to explore different models significantly enhances intuition. This paper describes a statistics package based on the Maple symbolic computing system and gives several examples of its educational use.

Espasandin Lopes, C. A. Teachers' development and developing children's stochastic knowledge.

This paper is about a research, performed in Brazil, and based on the epistemological reflection of the teacher concerning the stochastic ideas in elementary education and their critical reflexive practice. Throughout preparation and development of activities in statistics and probability, the teachers built different ways in their pedagogical practices to widen their professional development, writing and publishing their experience reports. The teachers needed to encourage learners to socialize their solutions, learning to listen to criticism, appreciate their own jobs as well as those of other students. The work with probability and statistics can be of great contribution, having in mind its natural problems, helping the enrichment of the reflexive process. The work with Stochastic in the classroom must promote discussions and reflections to solve a problem-situation which was asked by the students or instigated by the teacher.

Meletiou-Mavrotheris, M. & Stylianou, D. A. On the formalist view of mathematics: Impact on statistics instruction and learning.

In the paper, we argue that the persistence of students' difficulties in reasoning about the stochastic despite significant reform efforts in statistics education might be the result of the continuing impact of the formalist mathematical tradition. We first provide an overview of the literature on the formalist view of mathematics and its impact on statistics instruction and learning. We then re-consider some well-known empirical findings on students' understanding of statistics, and form some hypotheses regarding the link between student difficulties and mathematical formalism. Finally, we briefly discuss possible research directions for a more formal study of the effects of the formalist tradition on statistics education.

Monteiro, C. & Ainley, J. Developing critical sense in graphing.

In current social contexts there are various situations in which people participate in graphing activities. The school has an important role in the teaching of graphing knowledge to citizens. Several researchers have stressed critical sense as an important aspect of the data handling process. This paper reports on a pilot study exploring some tasks in which primary school teachers might approach graphing, using critical sense as an important element. Analysis of the results suggests factors, which may be significant in the design of such tasks.

Nilsson, P. Experimentation as a tool for discovering mathematical concepts of probability.

This paper puts focus on students' ability to handle the component of probability, while acting under uncertainty in an experimental environment. In particular, we are interested in to what extent seventh grade students are able to develop secondary intuitions of probability during interaction with mathematical modelling in a co-operative setting. To catalyse this kind of learning situation, involving consciousness and reflective processes, a competitive game is introduced to the students. We argue that this kind of experimental mathematical activity supports the desired development of secondary intutitions.

Pange, J. & Talbot, M. Literature survey and children's perception on risk.

Risk perception differs between people. There are studies presenting the different risk perception according to their character. The aim of this paper is to make a short literature review on risk and also is trying to present papers which consider the hypothesis: 'is any correlation between risk perception by children and mathematical thinking?'.

Paparistodemou, E. & Noss, R. Fairness in a spatial computer environment.

This paper focuses on how children (6-8 year-old) construct a fair sample space in a spatial environment. It illustrates what kind of intuitions about chance and randomness children bring from their experiences and it refers to their expressions of fairness, while they are involved in a spatial computer-based game. The toolset offered children the opportunity to manipulate sample space and distribution in order to achieve fairness in their game. The paper describes children's constructions of symmetric and asymmetric fairness, the two categories that children employed for the construction of fairness in a computer spatial environment.

Pitarch Andrés, I. & Orús Báguena, P. Logic and treatment of data in secondary school.

In summary, in this communication, we try to show the possible coexistence of the treatment or analysis of data and of the logic, in the obligatory secondary education.

Pratt, D. The Emergence of probabilistic knowledge.

In this paper, I summarise a theoretical framework for the growth of probabilistic knowledge. Through reflection on two theories that seek to model sense-making activity at quite different grain sizes, a synthetic view is proposed that draws its power from three sources: (i) its connection with the two original theories, (ii) its ability to model the behaviour of children working with a particular computer microworld, and (iii) its consistency with work in the literature. In particular, the theory offers a coherent way of thinking about inconsistency in children's responses and proposes principles that could underpin effective teaching or curriculum development.

Way, J. The development of young children's notions of probability.

Task based interviews were conducted with 74 children aged four to twelve years from three schools. These children had not received any formal instruction in probability as it was not part of their school curriculum. The study confirmed the presence of three developmental stages, but also revealed two distinct transitional stages. This paper focuses on the characteristics of children's strategies for making probabilistic judgements in each stage, and on the implications for teaching.

Poster presentations

Villarroya Grau, G, & Orús Báguena, P. Didactic software for graphical data handling through classification.

We present an original software which focuses on graphical data handling in secondary education and is based on graphic semiology ideas. The software potential is shown through the various options for graphical classification of binary data (EXCEL files), which are transformed into a two-colour grid (codifying the presence and absence of each variable for each subject). It also enables manual classification from a direct permutation of rows and columns and the table automatic treatment produces classifications using Gras and Lerman's similarity index or classes. The software has been designed as a didactic tool to teach logical-statistical and classificatory reasoning.

FORTHCOMING CONFERENCES

Hawaii International Conference on Statistics and Related Fields, June 5-8, 2003

The main goal of the 2003 Hawaii International Conference on Statistics and Related Fields is to provide an opportunity for academicians and professionals from all over the world to come together to meet and learn from each other. The Conference will provide a meeting place for academicians and professionals from related fields and, for those with cross-disciplinary interests, a chance to interact with each other from inside and outside their own particular disciplines, as well as for presenting research. Web page: http://www.hicstatistics.org/

The 2003 Mathematics Education Research Group of Australasia [MERGA] Conference, Deakin University, Geelong, Australia, July 6- 10, 2003

*Further information:*http://www.deakin.edu.au/fac_edu/numeracy_and_merino/merga/ or from the Chair of the organising committee: Judy Mousley (judym@deakin.edu.au)

IX Seminar on Applied Statistics: "Statistics in Education and Education in Statistics", Interamerican Statistical Institute, Rio de Janeiro, Brazil, July 7-10, 2003.

This seminar should provide an excellent opportunity for exchange of ideas, dissemination of recent work and developments that took place in Brazil and the Americas over the last few years, together with discussion of perspectives for advancement of both areas in the future. The goals are to attract wide participation from researchers, university teachers, and professionals, students and high-school teachers. More information is available from Pedro Luis do Nascimento Silva (pedrosilva@ibge.gov.br) or the Web page: http://www.indec.mecon.gov.ar/newindec/proyectos/iasi_ingles/act_seminarios.htm

The 27th PME Conference, Honolulu, Hawaii, July 13-18, 2003

You are cordially invited to participate in the Twenty-Seventh Annual Meeting of the International Group for the Psychology of Mathematics Education (PME) and the Twenty-Fifth Annual Meeting of PME-North American Chapter to be jointly held in Honolulu, Hawaii on July 13-18, 2003.The conference will be held in the Hawaii Convention Center. Web site: http://www.hawaii.edu/pme27. Coordinators of the Stochastical Thinking Learning and Teaching Discussion Group are Michael Shaughnessy (mike@mth.pdx.edu) and Jane Watson (Jane.Watson@utas.edu.au).

Third International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3), University of Nebraska, USA, July 23-28, 2003

The third in a series of International Research Forums is to be held in the United States of America in July 2003. The Teachers College Institute and the Department of Educational Psychology at the University of Nebraska-Lincoln will host the Forum. This gathering offers an opportunity for a small, interdisciplinary group of researchers from around the world to meet for a few days to share their work, discuss important issues, and initiate collaborative projects. The topic of the third Forum will be *Reasoning about Variability*. This conference is supported by the International Association for Statistical Education (IASE) and the American Statistical Association (ASA) Section on Statistical Education. More information from William T. Mickelson (wmickelson2@unl.edu) or from the Web site: http://tc.unl.edu/srtl.

ICTMA 11, Marquette University, Milwaukee, Wisconsin, USA, July 27-31, 2003

The International Conferences on the Teaching of Mathematical Modelling and Applications (ICTMA Conferences) have been held biennially since 1983. These conferences provide an international forum for the presentation and exchange of research, information, methods, materials, and ideas related to the teaching, learning, and assessment of mathematical modelling, mathematical models, and applications of mathematics. The theme for ICTMA 11 is *Mathematical Modelling: A Way of Life.* More information from ICTMA2003@aol.com or the Web page: http://mscs.mu.edu/~sue/ICTMA/ictma_11.html

JSM (the Joint Statistical Meetings) San Francisco, California, USA, August 3-7, 2003

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. It is held jointly with the American Statistical Association, the International Biometric Society (ENAR and WNAR), the Institute of Mathematical Statistics, and the Statistical Society of Canada. Attended by over 4000 people, activities of the meeting include oral presentations, panel sessions, poster presentations, continuing education courses, exhibit hall (with state-of-the-art statistical products and opportunities), placement service, society and section business meetings, committee meetings, social activities, and networking opportunities. San Francisco is the host city for JSM 2003 and offers a wide range of possibilities for sharing time with friends and colleagues. For information, contact meetings@amstat.org or see the Web site: http://www.amstat.org/meetings/jsm/2003/

International Conference on Creativity in Mathematics Education and the Education of Gifted Students, Rousse, Bulgaria, August 3-9, 2003

This conference is to be held at the University of Rousse. The main aim of the conference is to formulate the problem and globally define the direction of the development of creative mathematics education of gifted students. The basic issues to be discussed are (i) how to stimulate mathematical creativity in students and their teachers; (ii) what areas, methods and problems in mathematics are appropriate for stimulating the creative activity of students and (iii) who the gifted students are and how can they be identified. For further information see the websites www.cmeegs3.rousse.bg or www.ami.ru.acad.bg/conference2003 or email conf_orgcom@ami.ru.aced.bg

IASE Satellite Conference on Statistics Education and the Internet Max-Planck Institute for Human Development, Berlin, Germany, August 11-12, 2003

This conference will be organised by the IASE in cooperation with the Section on Stochastics of the German Mathematics Education Association, the Max-Planck-Institute for Human Development, and Stochastics Section of German Mathematical Association. The aim is to discuss the implications of the Internet for teaching and learning statistics: web based teaching, learning, materials and resources. Electronic proceedings and a limited number of printed copies of the set of papers presented will be produced after the conference.

More information from: the Chair Larry Weldon (weldon@sfu.ca) or the Chair of the Local Organising Committee, Joachim Engel (JoaEngel@aol.com).

IASE IPM at ISI-54 Berlin, Germany, August 13-20, 2003

The following Invited Papers Meeting will be organized:

- IPM44 Teaching probability with a modelling approach. Organizer: Michel Henry (henry@math.univ-fcomte.fr)
- IPM45 Statistics training for consultants or collaborators. Organizer: Gabriella Belli (gbelli@vt.edu).

- IPM46 International co-operation in research on statistics education. Organizer: Lisbeth Cordani (lisbeth@maua.br).
- IPM47 Mathematics teachers teaching statistics. Organizer: Susan Starkings (starkisa@sbu.ac.uk).
- IPM48 Statistics education for media reports. Organizer: Maxine Pfannkuch (pfannkuc@math.auckland.ac.nz).
- IPM49 Teaching and learning approaches aimed at developing statistical reasoning, thinking or literacy. Organizers: Joan Garfield (jbg@tc.umn.edu) and Dani Ben-Zvi (dani.ben-zvi@weizmann.ac.il) .
- IPM50 Statistics Teaching in the Internet Age. Organizer: Wolfgang Haerdle (haerdle@wiwi.hu-berlin.de).
- IPM68 Education and assessment of literacy, numeracy and other life skills. Organizer: Denise Lievesly (d.lievesly@unesco.org).
- IPM69 Impact of developments in information systems on statistics education (joint with IASC). Organizer: Annie Morin (Annie.Morin@irisa.fr) and Albert Prat.
- IPM70 Teaching biostatistics (joint with the International Biometrics Society). Organizer: Elisabeth Svensson (elisabeth.svensson@esa.oru.se) and Els Goetghebeur.
- IPM71 Educational implications of statistical method and modelling developments in psychometry. (Joint with the European Mathematical Psychology Group). Organizer: Helena Bacelar (hbacelar@fc.ul.pt).

Chair of the IASE Programme Committee Gilberte Schuyten, Gilberte.Schuyten@rug.ac.be

European Conference on Educational Research, University of Hamburg, Germany, September 17-20, 2003

The European Educational Research Association (EERA) will be holding its annual conference in Hamburg (Germany) in cooperation with the University of Hamburg and the Deutsche Gesellschaft für Erziehungswissenschaft (DGfE). More information from the web page: http://www.eera.ac.uk/events.html

6th International Conference of The Mathematics Education into the 21st Century Project, Brno, Czech Republic, September 19-25, 2003

The Mathematics Education into the 21st Century Project has just completed its fifth successful international conference in Sicily, following conferences in Egypt, Jordan, Poland and Australia. The next conference will be in Brno, Czech Republic. The title of the conference is *The Decidable and the Undecidable in Mathematics Education*, a tribute to Kurt Godel who was born and educated in Brno. More information from Allan Rogerson (arogerson@vsg.edu.au) or the Web page: http://math.unipa.it/~grim/21project.htm

The 10th International Congress on Mathematics Education, Copenhagen, Denmark, July 4-11, 2004

venue will be the Technical University of Denmark, located in a northern suburb of Copenhagen. The IASE will be collaborating in the organisation of specific statistics education activities in the conference. Chair International Programme Committee: Mogens Niss (ICME10-IPC@ruc.dk). Chair Local Organising Committee Morten Blomhøj (ICME10-LOC@ruc.dk). Conference Web page: http://www.icme-10.dk/

IASE 2004 Research Round Table on Curricular Development in Statistics Education, Lund, Sweden, June 28 - July 3, 2004

The Round Table dates coordinate with those of the Tenth International Congress on Mathematical Education, which takes place in Copenhagen, Denmark 4-11 July 2004. 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, E-mail: (burrill@msu.edu).

IASE Activities at the 55th Session of the ISI, 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. As such he also is Chair of the IASE Programme Committee, which is in charge of preparing a list of Invited Paper Meetings to be organised by the IASE alone or in co-operation with other ISI Sections, Committees and sister societies. The committee will pay special attention to new topics that have been not discussed at the previous ISI Session. There is still time for you to propose a session theme for the IASE sessions for ISI55 in Sydney in 2005. Sessions that are of joint interest to IASE and another ISI section are also sought. Suggestions should normally include the name of the session organiser, a short description of the theme and an indicative list of possible speakers. Please email your proposals to Chris Wild at c.wild@auckland.ac.nz.

ICOTS-7, Working Cooperatively in Statistics Education, Brazil, 2006

We are also glad to announce that the IASE Executive accepted the proposal made by the Brazilian Statistical Association to hold ICOTS-7 in 2006 in Brazil. The proposal is also supported by the statistical associations in Argentina and Chile. Pedro Morettin <pam@ime.usp.br> is the Chair of the Local Organising Committee and Lisbeth Cordani <lisbeth@maua.br> is acting as a link between the IASE Executive and the local organisers. Scientific Committee IPC: Carmen Batanero (Chair), Susan Starkings (Chair Scientific Programme), John Harraway (Scientific Secretary), Allan Rossman and Beth Chance (Editors of Proceedings). More information from Carmen Batanero (batanero@ugr.es).