Statistics Education Research Journal

Volume 4 Number 2 Nov 2005

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

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

Statistics Education Research Journal

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

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

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

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

Submissions

Manuscripts must be submitted by email, as an attached Word document, to co-editor Tom Short <tshort@iup.edu>. These files should be produced using the Template available online. Full details regarding submission are given in the Guidelines for Authors on the Journal's Web page: http://www.stat.auckland.ac.nz/serj

Publication: IASE/ISI, Voorburg, The Netherlands
Technical Production: University of New England, Armidale, NSW, Australia

ISSN: 1570-1824

SERJ EDITORIAL BOARD

TABLE OF CONTENTS

EDITORIAL

In this second issue of 2005 we start with a reflection on some of the issues we face at SERJ, and follow up with updates about changes and plans.

As an international journal, SERJ seeks both to report on research in diverse contexts in which teaching and learning of statistics is being studied, as well as to contribute knowledge relevant to researchers and practitioners across many countries and contexts of educational work. This dual task presents many challenges to members of SERJ's editorial board and to reviewers. We have to be aware of diversity and accommodate variations in many aspects of scientific reporting, from seemingly technical but non-trivial things such as terminology or writing styles, to understanding the nuances and impact of differences in systems of education or teaching/learning contexts, all the way to conceptions of what constitutes "current knowledge", "research", and equally important, "good research".

However, our conception of "current knowledge" is limited, not only by the small (though growing!) number of researchers interested in statistics learning and teaching, but also by language barriers. While English is the dominant language for scientific reporting, we as a research and practice community are hampered because many prospective authors and researchers, as well as practitioners, cannot read or communicate well in English. Potentially important research being carried out in countries where English is not spoken or read well cannot gain broad exposure. It is therefore not surprising that a disproportionate number of the papers being submitted to SERJ come from English-speaking countries. We hope to see this situation change over time.

The above serves as a backdrop for an important aspect of the present SERJ issue. In addition to three papers in English, we also publish a paper in French (the first in French that SERJ has published) *and* a paper in Spanish, yet with an important twist. The international standard used by journals that accept papers in more than one language is to publish the paper in its original language, with a brief abstract of 100-150 words in English. We find that such an abstract, while giving English readers a general idea of the issues addressed, is too short to be of much value in terms of the paper's contribution to scholarly and applied knowledge for a wide audience. Hence, for this issue we asked authors whose papers were not written in English to add an extended summary in English. Of course, this new feature places an extra burden both on authors and on our editorial board and it also makes for a slightly more complicated structure of the elements which precede the paper itself (i.e., title, keywords, and abstracts in two languages, and an extended summary in English). Yet, we hope that this new feature can give all readers better and deeper access to new research findings. Please give us your feedback.

We now turn to a report of changes and future plans at SERJ. We are happy to announce that Tom Short (Indiana University of Pennsylvania, USA) will be the next SERJ co-editor and will start his four-year tenure in January 2006. Tom will replace Flavia Jolliffe, whose term is coming to an end in December, and joins Iddo Gal, who will continue as co-editor until the end of 2007. Tom was warmly nominated by the Search Committee appointed by the IASE president, Gilberte Schuyten, consisting of Chris Wild (chair), Iddo Gal (continuing editor), and Mike Shaughnessy (IASE member). The IASE Executive has unanimously approved this nomination. Tom brings with him, in addition to high motivation and interest, a strong background in the area of statistics

Statistics Education Research Journal, 4(2), 2-3, http://www.stat.auckland.ac.nz/serj

education and a substantial editorial experience. He has already worked successfully in the past with a diverse editorial board and with many referees and authors as editor of the Journal of Statistics Education, and has been involved both in applied work as a statistician as well as in various projects focused on statistics education and teacher training. In our next issue Tom will share his ideas and intentions regarding the Journal.

There are other changes in the SERJ editorial board. We welcome Carol Joyce Blumberg who joins us as an Associate Editor for a 3-year term. She has been involved with SERJ from the beginning in her role as IASE vice-president responsible for IASE publications. Chris Reading, who has been our trustworthy Assistant Editor in charge of editing and producing each issue ever since SERJ's first issue was published in 2002, will cease to take this role at the end of this year and instead continue as a regular Associate Editor. Please see more information and messages regarding Carol and Flavia later on this page, as well as a Call for Nominations for a new Assistant Editor.

Finally, we have started to process papers submitted for our special issue on research on "learning and reasoning about distributions", planned for November 2006. We are thinking about topics for other Special Issues later on, and encourage SERJ readers to propose relevant topics which can benefit from a focus by researchers. We also look forward to ICOTS7 in July 2006, where SERJ will arrange a workshop for prospective authors, similar to that which took place at ISI55 in Sydney in April 2005. There will also be a session on statistics education journals at ICOTS7 similar to one organised by Carol Joyce Blumberg at the Joint Statistical Meetings in August 2005. At this meeting Flavia presented a paper jointly written with Iddo which describes lessons learned at SERJ regarding typical problems with submitted papers and provides advice for prospective authors. This paper, along with other papers and presentations from JSM, is available at http://www.stat.auckland.ac.nz/~iase/publications.php?show=jsm and on the SERJ website.

IDDO GAL AND FLAVIA JOLLIFFE

Thanks from the SERJ editorial Board to Flavia Jolliffe: Our departing Co-Editor, Flavia Jolliffe, has been a founding editor of SERJ since 2001. Flavia has tirelessly devoted time and energy to SERJ as she handled incoming submissions and managed a continuous flow of correspondence with authors, reviewers, and associate editors. We thank Flavia for giving several years for work on behalf of SERJ and IASE.

Call for Nominations - New Assistant Editor: The Journal is looking for someone to join its Editorial Board as Assistant Editor starting in January 2006 for a 3-year term. The Assistant Editor is in charge of copy-editing and preparation of manuscripts accepted for publication, and for producing each issue in PDF format. Work is intermittent during the year and increases during the weeks leading to the publication of a new issue each May and November, when communication with authors and editors is also needed. Depending on qualifications, the Assistant Editor may also be involved in managing the SERJ website or take part in other activities of the SERJ editorial board. The ideal candidate will have excellent command of the English language, interest in editorial work, familiarity with basic desktop publishing or PDF-producing software, and some familiarity with research in statistics education.

Interested colleagues should send a letter of intent and a short curriculum vitae to Tom Short <tshort@iup.edu>, to whom any queries about the position may be addressed.

NEW CO-EDITOR

SERJ welcomes the following new Co-Editor who has joined the Editorial Board for a 4-year appointment 2006-2009.

Thomas H. Short is an Associate Professor in the Mathematics Department and Coordinator of the Applied Research Lab at Indiana University of Pennsylvania. From 1991 through to 2002 Tom was in the Department of Mathematical Sciences at Villanova University. Tom earned his B.S. in Mathematics from John Carroll University, and his M.S. and Ph.D. in Statistics from Carnegie Mellon University. He was Editor of the *Journal of Statistics Education* from 2001 through to 2003, and served on the *JSE* editorial board both before and after his term as editor. From 2000 through to 2005 Tom was a member of the American Statistical Association Advisory Committee on Teacher Enhancement, serving as Vice-chair (2003) and Chair (2004 and 2005). Tom is a Fellow of the American Statistical Association and received the 2005 Mu Sigma Rho National Statistics Honorary Society Statistical Education award.

NEW ASSOCIATE EDITOR

SERJ welcomes the following new Associate Editor who has joined the Editorial Board for a 3-year appointment 2005-2007.

Carol Joyce Blumberg is a Professor in the Department of Mathematics and Statistics at Winona State University. She was a Vice-President of IASE 2001-2005 and has been a Member of the ISI Publications Committee since 2004. Her current interests include survey research and statistics education, and she presently coordinates for IASE the International Statistical Literacy Project (course1.winona.edu/cblumberg/islplist.htm).

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

CHARACTERIZING YEAR 11 STUDENTS' EVALUATION OF A STATISTICAL PROCESS

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

ABSTRACT

Evaluating the statistical process is considered a higher order skill and has received little emphasis in instruction. This study analyses thirty 15-year-old students' responses to two statistics assessment tasks, which required evaluation of a statistical investigation. The SOLO taxonomy is used as a framework to develop a hierarchy of responses. Focusing on the quality of response allowed insight into and suggestions for how instruction might be improved. The implications for teaching, assessment, and the curriculum are discussed.

***Keywords*:** *Statistics education research; Evaluating statistical investigations; Assessment; SOLO taxonomy; Secondary students*

1. INTRODUCTION

In 2002 a new approach to national assessment in New Zealand was introduced at Year 11 (15-year-olds). Instead of one final external examination in mathematics, one third of the course is now internally assessed, with external moderation, and the rest is an external examination (New Zealand Qualifications Authority, 2001). Statistics is internally assessed and students are given data sets to investigate. The assessment is standards-based with three performance levels: achievement, merit, and excellence. Achievement requires students to interpret statistical information and answer straightforward questions. For merit, students must also draw inferences, justify their answer to their question, and comment on features in the data, whereas for excellence the requirement is to evaluate the statistical process. Students must provide evidence that they can meet these levels in two tasks. The tasks are designed so that it is clear what performance level each question is measuring. The level of statistical thinking required at Year 11 with this new internal assessment, compared to the previous external assessment that largely asked students to read and interpret graphs and calculate measures of central tendency, has produced real challenges for teachers and students. The focus of this paper is on characterizing student responses to the excellence part of the assessment, which requires students to evaluate the statistical process.

1.1. RELATED RESEARCH

Evaluation of the statistical process (problem, plan, analysis, conclusion) requires thinking tools such as a list of criteria or "worry questions" for each stage (see Section 4, Wild & Pfannkuch, 1999). These thinking tools need to be an integral part of students' analytic techniques that can be triggered to stimulate thought processes on what issues

Statistics Education Research Journal, 4(2), 5-26, http://www.stat.auckland.ac.nz/serj

need to be considered and taken into account when conducting a statistical investigation. Wild and Pfannkuch (1999) proposed that a checklist of basic questions could be drawn up for students which could be underlain with more and more sophisticated questions in an internet-type procedure. These underlying questions could be accessed as students' understanding progressed. This proposal emerged from the realization that the statistics discipline had developed tools for the analysis stage of the statistical process but had not paid attention to developing analytic tools for the other stages of the investigative cycle. One discipline that has developed such tools is quality management. The students in this study did not have access to any thinking tools for the problem, plan, and conclusion stages and this raises the question about what issues beginning students will consider when they evaluate a statistical investigation.

Gal and Garfield (1997, p. 4) stated that students should "be aware of possible biases or limitations or the generalizations that can be drawn from the data" but according to Gal (1997, p. 49) "little has been written about issues involved in assessing students' opinions about data." Research has been carried out on assessing students' opinions about media articles (e.g., Watson, Collis, & Moritz, 1994) but there is limited research on evaluating and analyzing the quality of students' opinions about a statistical investigation. People have written about the assessment procedures used when students conduct their own statistical investigations (e.g., Starkings, 1997; Holmes, 1997) but have not reported an analysis of students' responses. However, the critical evaluation of statistically-based reports in relation to statistical literacy has been a recent focus in research, and since there is considerable overlap in the skills required between evaluating someone else's report and one's own statistical investigation this literature will be drawn upon.

For the interpretation of media reports Watson (1997) identified a three-tier hierarchy of skills. These skills were: basic understanding of terminology; embedding of language and concepts in a wider context; and questioning claims. The first two skills are relevant for interpretation of the problem by the students but the third skill of challenging claims presented in the media is only partially relevant as the students would be challenging their own claims. From another perspective, Gal (2002, p. 3-4) believes that critical evaluation of statistically-based information is predicated on "a *knowledge* component (comprised of five cognitive elements: literacy skills, statistical knowledge, mathematical knowledge, context knowledge, and critical questions) and a *dispositional* component (comprised of two elements: critical stance, and beliefs and attitudes)."

Considering Gal's (2002) perspective, each of the five cognitive elements of the knowledge component is elaborated upon with respect to how each can be used as a criterion for the setting of assessment tasks. A Year 11 student's capability and the prescribed curriculum are also taken into account. For general literacy, the first cognitive element, students need to understand the text as well as distinguish the meaning of statistical terms (e.g., spread) from their everyday meaning (Watson, 1997). A written assessment task for students should ensure the text is written in a meaningful way for the particular age group and that the statistical terms that are used in the text are part of their statistical knowledge base, the second cognitive element. When evaluating a statistical investigation students may be required to draw upon a wider statistical knowledge base such as having knowledge about sampling variability. The statistical knowledge base element, which Year 11 students are currently exposed to, is problematic. Questions, for example, have been raised about the type of conceptual experiences Year 11 should have as they move towards formal inference (Pfannkuch, 2005). Mathematics knowledge, the third cognitive element, at its basic level refers to 'number sense', which refers to an ability to correctly interpret numbers such as fractions and percentages in a report (Gal, 2002), which would be assumed knowledge at Year 11. At another level 'number sense' means evaluating whether the data or numbers are plausible, which requires an ability to spot basic arithmetic errors, inconsistencies, and massaging of data, knowledge that probably cannot be assumed.

Context knowledge, the fourth cognitive element, is not only necessary for interpreting and gleaning information from statistical data but also is a prerequisite for critical reflection about statistical information (Wild & Pfannkuch, 1999; Pryor, 2001; Gal, 2002). Therefore the Year 11 assessment tasks should use contexts that are sufficiently well known to students that their 'real world' knowledge could be used not only to understand the problem but also to suggest possible improvements or alternative explanations in their evaluations. Pryor (2001), in her research on tertiary students' ability to critique media reports, identified critical thinking from a context knowledge base as a precursor to critical thinking from a statistical knowledge base. Gal's (2002) list of critical questions, with which students should be familiar, addresses the fifth cognitive element. The list contains mainly statistical knowledge worries but it would be justifiable to have more context knowledge worries. This raises the question as to what critical worry questions would be suitable for Year 11 students when evaluating a statistical process.

When considering Gal's (2002) dispositional component for statistical literacy, Wild and Pfannkuch (1999) claimed that a person's propensities to adopt a critical stance and to be curious and imaginative were dispositions that drive a statistical investigation. Hence, it would seem that for Year 11 students to evaluate an investigation the adoption of a critical stance would help them in their ability to critique an investigation. An implication is that the assessment tasks should be sufficiently motivating and interesting to the students to invoke a critical stance and, if possible, the tasks should also challenge their beliefs. According to Pfannkuch (1996), students' non-awareness of their own beliefs and attitudes or of community assumptions affect their ability to evaluate media reports. Such findings have implications for teaching the evaluation of a statistical process. A willingness to think beyond one's own beliefs at the metacognitive level should be part of students' learning experiences in the classroom. Indeed, Gal (1997) believed that the development of students' ability to generate sensible and justifiable opinions should be a focus of instruction. He suggested that teachers should first elicit the student's opinion and then follow up with a question asking the student to provide evidence for the opinion. A climate of "explaining one's reasoning" should be fostered in the classroom in order for students to learn how to evaluate a statistical process.

Evaluation implies that there exist criteria upon which judgements are made (Bloom, 1956). This raises the question as to what criteria should be used for evaluating the statistical process. Starkings (1997, p. 144) stated in her marking schedule the criterion for evaluation of the statistical process: "Clearly relates solution to the problem. Shows a good understanding and appreciation of the solution." The New Zealand Qualifications Authority (2001) marking schedule exemplars referred to sources of bias, improvements, limitations, and appropriateness of the statistical process and a few suggestions were given on how a student might answer a particular question. These examples were general in that they could be applied to any evaluation such as stating another graph that could be drawn, more accurate measurements that could be taken, or more data that should be collected. The ability to evaluate a statistical process is considered to be indicative of achieving "excellence" in the given task. When considering the exemplars given to teachers, however, the judgement of excellence does not seem to be based on a high level quality of response.

Pegg (2003, p. 252) stated that the SOLO model (Biggs & Collis, 1982) not only offered a method to categorize the quality of the responses but also allowed "teachers an insight into where instruction might most profitably be directed." It is this twofold applicability that is pertinent to this research. First, teaching evaluation of the statistical process is new to teachers and hence they are uncertain what cognitive level or patterns of thought are present in their students and what constitutes a quality response. Second, such an analysis will aid their understanding of how to foster and scaffold students' thinking in the evaluation of a statistical process. From a research perspective more knowledge will be built up in this area.

1.2. RESEARCH QUESTIONS

As part of a larger project on developing Year 11 students' statistical thinking, the following three research questions are addressed in relation to the evaluation of the statistical process and are based on responses to two assessment questions:

- What response category types describe Year 11 students' evaluation of a statistical process?
- What issues do students consider when evaluating a statistical process?
- What SOLO levels do students attain when evaluating a statistical process?

2. RESEARCH METHOD

The research described in this paper is concerned with an identified problematic area from the first year of a planned three-year project on developing students' statistical thinking. The Wild and Pfannkuch (1999) statistical-thinking framework underpins the research project. The framework is initially employed to communicate to teachers the nature of statistical thinking and habits of thinking that should be fostered in students. It is then not only concretized by the teachers in their instruction but also is employed as a thinking tool to critically reflect upon and to describe and analyze teaching and learning situations.

2.1. APPROACH TO RESEARCH

A developmental research method is used that is based on the ideas of Gravemeijer (1998), Wittmann (1998), and Skovsmose and Borba (2000) (see Pfannkuch & Horring, in press, for a fuller account). The research method is developmental in that an action-research cycle is set up whereby problematic areas are identified by teachers and researcher through observations and critical reflections on the implementation of a teaching unit and by the researcher through analysis of student assessment responses. The students also identify areas of concern about their learning through a questionnaire. The teachers and researcher then discuss how the current situation might be changed for the following year when the unit is taught again. The teachers then rewrite the teaching unit.

An initial approach to the mathematics teachers by the researcher during 2002 resulted in them selecting Year 11 for the project. The case-study teacher was self-selected. A workshop, which focused on communicating the nature of statistical thinking to the teachers, was conducted by the researcher. After the workshop the case-study teacher and another teacher were interviewed to identify problematic areas in their 2002 statistics-teaching unit (Pfannkuch & Wild, 2003). These two teachers and the researcher then discussed teaching ideas that could be implemented to enhance the development of students' statistical thinking. The teachers wrote a new four-week statistics unit for 2003. Although all Year 11 teachers implemented the new teaching unit research data were mainly collected from the case study classroom. These data were videotapes of 15 lessons, student bookwork, student responses to the assessment tasks, student questionnaires, and the teacher's weekly audio-taped reflections on the teaching of the unit. Two main areas of concern, identified by the researcher and case-study teacher after the first teaching implementation in 2003, were informal inference and evaluation of the statistical process. Thus the first analysis of these data focused on these identified problematic areas. The informal inference analysis and its implications are reported in Pfannkuch (2005).

2.2. PARTICIPANTS

The school involved in the project draws on students from low socio-economic backgrounds, is culturally diverse, and has teachers interested in improving their statistics teaching. This secondary girls' school like many other schools in Auckland city has a high percentage of new immigrants to New Zealand (about 60%), many of whom have English as their second or third language. In the case-study classroom there were thirty students who were regarded by their teacher as above average in mathematical ability. In this particular class 45% were Pakeha (New Zealand European), 40% were Maori (New Zealand indigenous) or Pasifika (Pacific Islands), and 15% were Asian or Indian. Two students chose not to be video-taped for the research project. The teacher is Pakeha, in her mid-thirties, has a first degree majoring in education, a Masters degree in mathematics education, and has taught secondary mathematics for twelve years.

The class is taught mathematics by the teacher for four hours per week. The teacher is in charge of Year 11 mathematics and therefore, in consultation with the other Year 11 teachers, writes an outline of the content to be covered together with suggested resources and ideas for teaching the unit. She also writes the internal assessment tasks, which are moderated at the national level. The researcher previously knew the teacher on a professional basis. The researcher was used as a source of teaching ideas before and during the teaching of the unit and was consulted about the assessment tasks.

2.3. THE ASSESSMENT TASKS

Students were given two assessment tasks, Task One (Appendix A) and Task Two (Appendix B) which were created by the case-study teacher. The assessment occurred in two stages. In the first stage the students were given only the story and data for Part A of Task One and asked to pose a question. The teacher then marked their ability to pose a question. For students who could not pose a question, the teacher gave a question to them. In the second stage the students were given one hour and forty minutes to complete both Part B of Task One and Parts A and B of Task Two.

For Task One the students were given a table of data showing the maximum temperatures of two cities Napier and Wellington, which were taken from some summer newspapers. A story involving a decision about where to go for a summer holiday was communicated to the students. Students were required to pose a question (e.g., Which city has the higher maximum temperatures in summer?), analyze the data, draw a conclusion, justify the conclusion with three supporting statements, and evaluate the statistical process with three statements (see Appendix A, Task One, Question 4). For Task One it should be noted that the data were presented to the students as two independent samples. A statistician might have recorded the maximum temperatures for each city each day and then conducted a paired comparison test.

Task Two had two parts. In Part A of the task weather data from the Pacific, Australia, and New Zealand regions were used to generate questions for the students to answer not given in Appendix B). In Part B students were required to evaluate the statistical process carried out by another person, named Jason (see Appendix B, Task Two, Question 2). For the evaluation of the statistical process it was decided to prompt the students to consider each stage of the process (problem, plan, analysis, conclusion). The prompt for considering the problem posed was omitted from the second task but on reflection should have been included. This research is focused on the students' evaluation of a statistical process and hence it is the student responses to Question 4 of Task One and Question 2 of Task Two that are analyzed.

2.4. APPROACH TO ANALYSIS OF ASSESSMENT RESPONSES

The analysis of the evaluation of the statistical process occurred in two stages. First, the student assessment responses to the two evaluation questions were analyzed. The analysis used a spreadsheet whereby a clustering procedure was used to sort the responses into categories (Miles & Huberman, 1994). The classification of the quality of the response for each level within a category used a hierarchical performance level approach based on the SOLO taxonomy (Biggs & Collis, 1982). The approach recognizes that within the concrete-symbolic mode, in which these students would most likely be functioning, there are at least one and possibly two distinguishable cycles of thinking operating through four levels (PUMR): pre-structural (P) – no use of relevant aspects; unistructural (U) – focuses on one piece of relevant data; multistructural (M) – two or more pieces of data used without integration; relational (R) – all data integrated into coherent whole (Pegg, 2003). These hierarchical levels were determined again by using a clustering approach within the spreadsheet. Based on the student responses qualitative descriptors for each level within a category were written and coded by the author, and then another person independently coded all responses. A consensus was reached between them on the final codes for each student response.

Second, the transcriptions of the video-tape data from the case-study classroom were qualitatively analyzed for instances of the evaluation process in operation in the classroom. This analysis was used to inform the discussion about the assessment responses.

3. RESULTS

The three research questions are addressed respectively in this section. First, descriptors of the category types for student responses to the evaluation of a statistical process, which were derived from the data, are discussed. Second, examples of the student responses are discussed in terms of the issues students considered when evaluating the statistical process, and third, a summary of the SOLO levels attained by the students is presented.

3.1. RESPONSE CATEGORY TYPES

Within the four stages of the statistical process most responses were to the analysis and plan stages, giving four distinct categories whereas there was little response to the problem and conclusion stages, which were combined into one category. The five main categories of response identified with respect to the students' evaluation of the statistical process were: *My/Someone Else's Analysis* and *Another Analysis* that could be conducted for the analysis stage; *More Data* and *Other Data* that could be collected for the plan stage; and *Other* which mainly related to the problem and conclusion stages (Fig. 1). These categories turned out to be similar to the prompts given to the students in the questions. Within these categories hierarchies of responses were identified and qualitatively described, reflecting the use, combining, and relating of elements suggested in the SOLO model. The descriptors for all the categories were similar in that they followed a sequence of *specify*, *justify*, and *relate*. The latter three categories, however, relied mainly on contextual knowledge of the situation whereas the former relied mainly on statistical knowledge. A possible transition into a higher-level mode requiring statistical knowledge, which was more abstract than contextual knowledge, was also identified for the latter three categories (Fig. 1).

	Evaluation of the Statistical Process			
	Analysis Stage Categories		**Plan Stage Categories**	**Problem/ Conclusion Stage Category**
SOLO Level Description	**1. My /Someone Else's Analysis**	**2. Another Analysis**	**3. More Data 4. Other Data**	**5. Other**
Prestructural (P) Inappropriate response	Gives an inappropriate reason why analysis is a good/bad choice.	Gives an inappropriate improvement or non-specific improvement.	Gives an inappropriate improvement or non-specific improvement.	
Unistructural (U) Single elements	Specifies one appropriate reason why analysis is a good/bad choice.	Specifies one appropriate statistical improvement.	Specifies one appropriate contextual improvement.	
Multistructural (M) Multiple elements	Justifies/critiques the choice of analysis in relation to the original question.	Justifies or gives an appropriate reason for the statistical improvement.	Justifies or gives an appropriate contextual reason for the improvement or an appropriate broad statistical justification.	
Relational (R) Relates to investigation	Justifies/critiques the choice of analysis in relation to the information that can be derived from that analysis or to the ability to reason from that analysis to answer the original question.	Relates the improvement to the original question under consideration.	Relates the improvement to the original question under consideration.	
Extended Abstract (U(2)) Brings in extra statistical elements			Specifies a statistical improvement.	

Figure 1. Categories and hierarchical descriptors for evaluation of the statistical process

The hierarchical descriptors for each category will now be explicated more fully in terms of the student responses. The data suggested it was necessary to have a separate descriptor for *My/Someone Else's Analysis* as full integration at the relational level seemed to occur when the students amplified how a particular analysis allowed them to reason about the question, such as this response for Task One:

S5: I believe that the box-and-whisker graph was the most appropriate graph to use because it is very easy to read and at a glance you can see that Napier is overall warmer than Wellington. It is an appropriate graph for a comparison question (R).

The *Another Analysis* hierarchical descriptors were similar to categories 3, 4, and 5. The difference was that the students were commenting on the analysis and hence needed to use their statistical knowledge to justify the suggested improvement. An example of each level of response for Task One is:

S21: Use histogram graph (P).

S17: A back to back stem-and-leaf may have been a better graph because it would have shown all the figures (U).

S18: Could draw a stem-and-leaf graph (back to back) and look to see if there are any peaks (M).

S29: A back to back stem-and–leaf would also have been a good graph for me to draw because it would have shown me the shape of the data and given me a good idea where most of the temperatures were for each city (i.e. 20 something degrees or something teen degrees etc.) (R).

The prestructural response was considered an inappropriate improvement, as it did not clearly specify how the data would be compared. The unistructural response gave an appropriate alternative graph for the comparison of data but the reason was inappropriate. The multistructural response recognized that a stem-and leaf graph allows peaks to be seen implying that these could not be seen in boxplot graphs. The relational response extended the idea further by relating this advantage to being able to find out more information about the temperatures of the cities.

For categories 3, 4, and 5 one set of descriptor levels was sufficient. The first identified cycle was based on and characterized mainly by contextual knowledge of the situation. Occasionally a student gave a broad statistical justification in the sense that the statement could apply generally to any investigation and hence it was classified as multistructural within the first cycle rather than a second cycle response. The following response was classified as multistructural in the first cycle since such a general statement was considered a broad statistical justification rather than a specific and full statistical justification and explanation:

S2: His analysis can be improved if he had more information of temperatures from other days and maybe also from another country so that he will be able to generalize his findings (M).

In comparison, the beginnings of specific statistical improvements were identified in two students' responses. These students seemed to be moving beyond the relational level as they began to think about how they might analyze those data, for example, suggesting the possibility of a graph:

S24: As my friend and I are wanting to know the warmest place to go perhaps it would be to our benefit to collect a range of data and graph the coldest temperatures of these two cities as well. That way we would find out how cold it may get, and this may well alter our perspective of where we wanted to travel on holiday. That is other data that could also be collected (U(2)).

Such a response was classified as U(2) because it went beyond the level expected in the concrete-symbolic mode by bringing in the beginnings of statistical knowledge rather than being solely based on contextual knowledge.

3.2. ISSUES CONSIDERED BY STUDENTS

Considering the students were evaluating the statistical process (problem, plan, analysis, conclusion) their responses to and the criteria they used for judging each stage of the process are highlighted. It should be noted that *students were only required to make three statements for each question*. Therefore the no response category means that a student did not respond in that particular category.

Problem Stage Improvements to the question posed were classified with the conclusions stage under *Other* as few students responded to this stage. Two students responded successfully by suggesting an improvement to the question they posed for Task One (see Table 1), for example:

S23: Next time I would improve the question I posed by looking into the maximum temperatures not trying to draw conclusions on finding a warmer climate with only maximum temperature statistics. I would change my question to, does Napier or Wellington have a higher maximum temperature? Since that is more to do with the data I was given (U(2)).

The two successful students realized that the measures used were possibly not relevant to the question they had posed.

Table 1. Task One: Details of student responses

	Analysis Stage		Plan Stage		Problem/ Conclusion Stages
SOLO Level	My Analysis	Another Analysis	More Data	Other Data	Other
No response	8	22	10	12	23
Prestructural	2	1	13	4	5
Unistructural	4	5	3	6	
Multistructural	13	1	3	5	
Relational	3	1	1	2	1
Unistructural(2)				1	1
Total number of students	30	30	30	30	30

Plan Stage Improvements to the plan centered on whether *More Data* or *Other Data* should be collected before making a decision or drawing a conclusion. In specifying *More Data* that should be collected, the student responses revealed a prevalent misconception. Twelve students for Task One and two students for Task Two (Table 2) mentioned that the sample size should be the same for each data set. A typical response for Task One was:

S6: In the data given there were 3 temperatures not given for Wellington. If they were given, the statistics could have increased or decreased and affected the results. There should have been an even amount of data for both sides – Wellington and Napier (P).

Table 2. Task Two: Details of student responses

	Analysis Stage		Plan Stage		Problem/ Conclusion Stages
SOLO Level	Someone else's Analysis	Another Analysis	More Data	Other Data	Other
No response	13	8	18	24	27
Prestructural	3	6	3	3	1
Unistructural	7	6	6	2	2
Multistructural	7	9	2	1	
Relational Unistructural(2)		1	1		
Total number of students	30	30	30	30	30

The more successful students focused on whether a reasonable sampling method had been used and suggested what data should be collected and why:

S2: I think that the analysis can be improved if she had another set of data to compare because temperatures can vary anytime of the year (M).

This response was considered a borderline multistructural response as the student did not clearly state that temperatures should be collected from other years and could vary from year to year. Acknowledging that temperatures could vary, however, is the beginnings of understanding sampling variation from a contextual perspective.

For the category *Other Data,* over half the students responded in Task One and one-fifth of the students responded in Task Two. Some students did not specify the actual data that should be collected and hence their responses were classified as prestructural such as the following statement for Task Two:

S4: Other data could have been collected to verify or support Jason's statement in a more trustworthy way (P).

If the student was able to specify the appropriate weather data to collect the response was considered to be unistructural:

S22: Jason could have improved this by: Using data from the whole world not just NZ and Pacific / Oz (U).

This response suggested an emergent realization that proving a theory in one region of the world was insufficient and that such an observation should be replicated elsewhere. Specifying appropriate weather data to collect and giving a reason that related to the warmth of the climate, such as the following response to Task One, was considered to be multistructural:

S26: Other data that could be collected to improve the analysis is rainfall over summer because to me places can be humid and raining, it would be important to know other aspects of the weather to compare regions (M).

Such responses from students indicated that they were beginning to realize that capturing the notion of 'warm' with a single measure was insufficient and that other measures for warmth such as humidity and minimum temperatures, should be considered in the comparison of regions. A relational response was considered to be a coherent whole when the specified data and reason for collecting them were justified in terms of the question posed:

S5: It may be that Napier gets colder during the nights than Wellington does. The minimum temperature should also have been gathered as the posed question asked "Which city is warmer over the summer period, Napier or Wellington?" and this data does not give the adequate information to correctly answer the question (R).

Eighteen student responses were classified under *Other Data* for Task One. Of those students, eight were classified at multistructural and above (Table 1).

Analysis Stage Giving reasons why *My Analysis* was a good choice in Task One (22 students) and suggesting *Another Analysis* for Task Two (22 students) prompted the most response (Tables 1 & 2). Most responses for these categories focused on suggesting either that the student's own graph was the best choice for Task One or a box-and-whisker graph was more appropriate for Task Two. A typical multistructural response for *My Analysis* in Task One was:

S7: I think I made the best choice in picking a box-and-whisker graph as it clearly shows the comparison between Napier and Wellington and their temperatures (M).

In Task Two in the *Another Analysis* category a prevalent multistructural response was:

S15: He should have drawn a box-and-whisker so you could actually compare the results (M).

The notion that box-and-whisker plots were useful for comparing grouped data was a typical response, with 16 and 10 students responding in Task One and Task Two respectively at the multistructural level and above.
For *Someone Else's Analysis* in Task Two, however, there were four comments on the categorization of the data such as:

S30: Could have kept the shower/rain in a different table and surveyed more days to see if rain affected the outcome (M).

These four students were beginning to realize that the categorization of data was relevant to a statistical analysis and that a different categorization might produce a different conclusion.

Conclusion Stage. Responses to the conclusion classified under *Other* were limited. One student attempted an alternative explanation for Jason's theory in Task Two, which was very convoluted, but showed she was willing to challenge the assumption that the clouds keep in the heat. Part of her response was:

> S5: My point is that it may not actually be the clouds keeping in the heat – it may just be that when it's colder it's less likely to heat up or get dramatically colder than it is, and that when it's hotter it gets much colder during the night (U).

Another student wondered whether her conclusion for Task One made sense with what she knew about the real world situation and reasoned from an individual event explanation:

> S4: Although these statistics have shown that Napier tends to have a higher maximum temperature than Wellington this statement in my opinion is not all that accurate. Because weather is unpredictable, Wellington might have a nice sunny day but then wind comes along and the temperature drops giving a low reading, e.g. 16.0 rather than 19.9 (P).

Students did not attempt to evaluate whether their conclusions were valid from the perspective of inference space judgement. No student responses were classified at the multistructural level or above for judging the conclusion.

3.3. SUMMARY OF THE SOLO LEVELS ATTAINED BY STUDENTS

The results tables (Tables 1 & 2) give an overview of the level of the responses that students demonstrated according to this method of analysis. It should be noted that students were only required to make six statements with respect to evaluating the statistical process, that is three statements for Task One and three statements for Task Two. In Task One the average was 2.5 statements per student whereas for Task Two it was 2 statements per student. The case-study teacher did observe that some students ran out of time to fully answer the Task Two questions. When considering the data overall there seemed to be some students operating fairly consistently at the same levels in the PUMR cycle. A summary table, calculated by taking the best four statements that a student made out of a possible six statements and assigning 0 to a prestructural or no response, 1 to a unistructural response etc., and then finding the mean score, produced a student profile of the class (Table 3).

Table 3. Summary of student responses

SOLO level and Mean Score	Number of students
Prestructural (0-)	7
Unistructural (0.75 -)	14
Multistructural (1.75 -)	8
Relational (2.75 -)	1
Total	30

Two-thirds of the class appeared to be operating at the unistructural level and below. It should be noted that, even though the case-study teacher's marking schedule for the evaluation questions was not based on these SOLO criteria, only one student was awarded an "excellence" grade, and this was the same student who was categorized as thinking at the relational level.

4. DISCUSSION

The student patterns of thought observed in the data could be said to be in response to the method of instruction (1), or to the students' general or contextual knowledge (2), disposition (3), statistical knowledge (4), cognitive development (5), or general literacy levels (6), which include both text comprehension and ability to communicate, or to the task which specifically mentioned aspects to comment on (7). It was considered that all seven factors could be operating on the level of student response, the contributory effect of each being unknown.

For the first factor, an analysis of the videotape data that recorded the teaching of the unit revealed that the instruction only once briefly focused on the evaluation of the whole statistical process. In one lesson the students compared the prices of second-hand exercycles and home gyms, which were data gathered from the newspaper advertisement columns. The teacher prompted an evaluation of the statistical process by asking: "How could I improve my investigation?" The students suggested these ideas: "Find out what kind of gym"; "Look at it everyday for a couple of weeks"; "Brand"; "Its quality"; and "How old it is". The teacher elaborated on their ideas but at no stage asked them to justify their opinions. In another two lessons the students evaluated graphs. The instruction was:

> Okay, I'm going to give you 5 minutes to look at the graph and I want you to write anything that pops into your mind, okay about the graph. Any conclusions that you can make, anything that you think is confusing on the graph, and also other questions that you might ask. So I want you to think quite generally. I want you to think about conclusions and other things that arise. In fact to start you off, you might want to go back to that phrase "I notice", and "I wonder".

Basically the students focused on whether graphs were misleading, on interpreting the information, and on thinking of contextual reasons for the distribution of the data.

Even though statistical investigations with an evaluation of the investigation have been an internally assessed component at Year 13 for about twenty years, this teaching approach would suggest that the evaluation part has not been a focus of the taught curriculum. Indeed the school textbooks have only cursorily covered this aspect. The main emphasis in textbooks and teaching has been misleading graphs. The current exemplars for Year 11 and the teacher workshops, provided by the New Zealand Qualifications Authority, gave minimal direction to teachers. Hence the outcomes of this study suggest that new understandings of how one teaches evaluation of the statistical process are needed if students are to improve their responses.

The responses to the evaluation task revealed a misconception that was not evident during teaching. Even though students had dealt with data sets of unequal sample size in class, they were never asked to evaluate those investigations and hence the prevalent misconception that data sets should have the same sample size before being compared was not uncovered. When reflecting on this misconception, and thinking of Curcio's (1987) hierarchical model for interpreting graphs, the author's observation, corroborated by the case-study teacher, was that the students had experience of reading the data, less experience at reading between the data, and little experience of reading beyond the data. If these students had some experience of inferring "missing data" from a data set they may have predicted that the missing summer temperatures were likely to be within the interquartile range or at least within the range. The problems of missing data are well known in statistics and students could be given opportunities to impute values for observations and to analyze data with and without the imputations. Specific attention could be drawn to students' beliefs and to whether their conclusions would change with unequal sample sizes. Although evaluation of the statistical process might be considered a higher level skill (Bloom, 1956) and may not currently be a strong feature of teaching, it would seem that allowing students to express their opinions on the overall investigation might allow some different insights into their thinking. These insights need to be reflected upon critically to determine new teaching approaches.

The other six factors that were identified as possibly affecting student responses raised three main issues. Firstly, cognitive development and the disposition or the willingness to adopt a critical stance might have had an effect on student responses but such effects could only be ascertained through a large longitudinal study. Secondly, the tasks were written with the students' contextual and statistical knowledge and text comprehension in mind but presumably these had an effect on student outcomes. Thirdly, general literacy was observed in the students' ability to communicate. The two coders learned a salutary lesson when they were reaching a consensus about the level assigned to two students. One of the coders was challenged on her ascribed levels for the two students and asked to justify the levels in terms of the given hierarchical descriptors. The coder then realized that the student who did not express herself fluently actually should have been awarded a higher level for her response, and that the student, fluent in English, was awarded one level too high for her response. Thus the level descriptors appeared to ensure that a marker was objective and not swayed by a student's ability to communicate. Pryor (2001) in her research on tertiary students ability to evaluate media reports calculated that text comprehension had twice the impact compared to graph comprehension and critical thinking in predicting students' ability to think statistically. If text comprehension and the ability to communicate the evaluation of the statistical process are related then it could be conjectured that such ability might have some effect on the level of response.

The analysis of the student responses produced a general hierarchy for evaluation of the statistical process described briefly by *specify*, *justify*, and *relate*. Gal (1997) referred to students justifying their opinions so that assessors could judge their reasonableness. This research confirmed his viewpoint in that a multistructural response was considered to be one in which the proposed improvement was justified. A relational response, however, extended the idea further by asking students to relate their opinions to the question under consideration. A similar scheme of argumentation for justifying inferences from data was also proposed by Cobb (1999). Furthermore, the analysis led to the conjecture that evaluating the statistical process for the problem, plan, and conclusion stages of an investigation might, firstly, be built on contextual knowledge, and secondly, on statistical knowledge. It was conjectured that this might be either a second UMR cycle in the concrete-symbolic mode or the beginning of the next mode, the formal mode, in the SOLO model (Biggs & Collis, 1982). Whatever the mode the next level is in, it is hypothesized that the multistructural level may occur when the student is able to give a statistical justification for the improvement mentioned. The relational level might occur when there is full integration of the contextual and statistical justification or critique when related to the original question. Presumably at this level the student would also demonstrate a fluent use of statistical language and ideas. The conjectured integration of these two knowledge bases at the relational level is supported by Wild and Pfannkuch (1999) who identified it as one of their five fundamental statistical thinking elements. It would seem that for learners contextual knowledge would be more prominent at first, a facet also found by other researchers (e.g., Watson, Collis, Callingham, & Moritz, 1995).

When evaluating media reports, Gal (2002) suggested students should have a list of critical questions in their heads while Wild and Pfannkuch (1999) suggested that students should have thinking tools at their disposal for all stages of the investigative cycle. From the student responses and the task prompts the critical questions for evaluation of the statistical process are shown in Figure 2. These questions were derived from two tasks and might be limited but they seem to be a suitable subset of critical questions at this stage for the Year 11 students.

All seven factors considered above may have affected the level and type of student response. At this stage, the teaching factor is the one area that can be targeted for improvement in the second year of the project. The analysis of the student responses has allowed some insight into their patterns of thought and the hierarchical descriptors have provided a possible structure for teachers to foster students' ability to evaluate the statistical process.

- Could improvements be made to the question? Are the measures used relevant to the question posed?
- Could improvements be made to the method of data collection? Has a reasonable sampling method been used?
- What other data should be considered or collected before making a decision?
- Are there better graphs that could be drawn or other statistics that could be calculated? (If you believe that you have made the best choice of graph(s) and statistics, explain why.)
- Could improvements be made to the categorization of the data?
- Is the conclusion valid? Does my conclusion make sense with what I know about the real world? What are some possible alternative explanations?
- Has the conclusion been drawn about the sample data under consideration?

Figure 2. Proposed judgement criteria for the evaluation of the stages of the statistical process

5. CONCLUSION

From a teaching perspective this analysis with the resultant hierarchical descriptors enabled the writing of model solutions to the evaluation questions of both tasks, which will be used by the teachers. The hierarchical descriptors have explicitly revealed the type of responses sought and will direct their teaching to scaffold students to higher levels of thinking. From an assessment perspective the hierarchical descriptors will enable teachers to be sure that a high quality response is awarded 'excellence' and they will explicitly know why a response is high quality. From a curriculum perspective the analysis raised some questions about the learning experiences and conceptual development at Year 11 for the evaluation of the statistical process. Statisticians, educators, and researchers need to work together on developing a teaching pathway that gradually builds up more critical questions or judgement criteria for evaluating a statistical process, which is directly linked to the prescribed curriculum content. From the statistics discipline perspective the current approach to evaluation is largely unstructured and is reliant on a statistician's experience in the field. Statisticians should begin to develop thinking tools for the problem, plan, and conclusion stages of the investigative cycle for the general statistics discipline not only for the enhancement of problem-solving but also for the evaluation of their own and others' investigations (Wild & Pfannkuch, 1999). These tools will require a synthesis of contextual and statistical understanding.

Gal (2002) claimed that critical evaluation of media reports was predicated on the joint activation of a knowledge component and a dispositional component. This research suggests that critical evaluation of the statistical process may be predicated on such joint activation but is enacted through communication and evaluative skills. These skills need to be explicitly taught and fostered in instruction with specific attention paid to justifying opinions and relating such justifications to the question under consideration. For the problem, plan, and conclusion stages of a statistical investigation instruction needs to focus on scaffolding students thinking to consider not only contextual but also statistical justifications and specifically explaining those justifications. Such thinking will present real challenges for teaching and the curriculum.

Wild and Pfannkuch (1999) suggested there were four dimensions in statistical thinking: the investigative cycle, types of thinking, the interrogative cycle, and dispositions. The interrogative cycle can be thought of as operating at the micro and macro level whereby the thinker evaluates the statistical process by: generating possibilities, seeking or recalling information, interpreting and connecting ideas, criticizing ideas against contextual knowledge, statistical knowledge, beliefs and so forth, and judging what to believe currently. Further research is needed on eliciting, understanding, and developing students' evaluative thinking. This research is based on a small sample and must be regarded as exploratory. Evaluative thinking, however, is a crucial dimension in fostering students' statistical thinking and deserves more research attention.

ACKNOWLEDGEMENT

The author thanks Jane Watson and Julia Horring for their comments and contributions to this paper.

REFERENCES

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

Bloom, B. (1956). *Taxonomy of Educational Objectives*. New York: David McKay Company.

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

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382-393.

Gal, I. (1997). Assessing students' interpretation of data. In B. Phillips (Ed.), *IASE Papers on Statistical Education ICME-8*, Spain, 1996, (pp. 49 - 57). Hawthorn, Australia: Swinburne Press.

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

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). Amsterdam, The Netherlands: IOS Press.

Gravemeijer, K. (1998). Developmental research as a research method. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 277-295). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Holmes, P. (1997). Assessing project work by external examiners. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 153-164). Amsterdam, The Netherlands: IOS Press.

Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis*. Thousand Oaks, CA: Sage Publications.

New Zealand Qualifications Authority (2001). *Level 1 achievement standards: Mathematics*. [Online: http://www.nzqa.govt.nz/ncea/ach/mathematics/index.shtml]

Pegg, J. (2003). Assessment in mathematics: A developmental approach. In J. Royer (Ed.), *Mathematical cognition* (pp. 227-259). Greenwich, CT: Information Age Publishing.

Pfannkuch, M. (1996). Statistical interpretation of media reports. In J. Neyland & M. Clark (Eds.), *Research in the learning of statistics: Proceedings of the 47th Annual New Zealand Statistical Association Conference* (pp. 67-76). Wellington, New Zealand: Victoria University.

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

Pfannkuch, M., & Horring, J. (in press). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill (Ed.), *Curricular development in statistics education: Proceedings of the 2004 International Association for Statistical Education Round Table Conference, Lund University, Sweden, 28 June-3 July, 2004*.

Pfannkuch, M., & Wild, C. J., (2003). Statistical thinking: How can we develop it? In *Bulletin of the International Statistical Institute 54th Session Proceedings* [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Pryor, H. (2001). *Assessment of the statistical literacy ability of some tertiary students using media reports.* Unpublished Masters Thesis, The University of Auckland.

Skovsmose, O., & Borba, M. (2000). *Research methodology and critical mathematics education.* Publication No. 17 Roskilde, Denmark: Centre for Research in Learning Mathematics, Roskilde University.

Starkings, S. (1997). Assessing student projects. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 139-152). Amsterdam, The Netherlands: IOS Press.

Watson, J. M. (1997). Assessing statistical thinking using the media. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam, The Netherlands: IOS Press.

Watson, J. M., Collis, K., & Moritz, J. (1994). *Authentic assessment in statistics using the media*. Report prepared for the National Center for Research in Mathematical Sciences Education – Models of Authentic Assessment Working Group (University of Wisconsin). Hobart, Australia: University of Tasmania, School of Education.

Watson, J. M., Collis, K., Callingham, R., & Moritz, J. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation*, *1*, 247-275.

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

Wittmann, E. (1998). Mathematics education as a ‘design science’. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 87-103). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

APPENDIX A

Task One: Holiday Temperatures **Name: ________________**
Maths Teacher _________

Part A

Every year you and a friend argue about where to go on your January summer holiday. You both enjoy outside activities and really enjoy the beach. Next year you will either go to Napier or Wellington for your holiday.

Since you both want to go to the warmest place you decide to analyse the maximum temperatures in Napier and Wellington. Your friend has found a stack of last summer's newspapers. She has gone through them and recorded the maximum temperatures in both places. These are shown in the tables below. Note that the temperatures are not in order.

What *statistical* question or hypothesis could you answer using this data?

Maximum Temp Napier °C
25.2
24.5
22.0
24.5
21.7
22.8
22.9
24.6
24.1
25.2
23.8
20.2
23.9
19.9
23.6
25.8
21.2
22.7
23.4
28.7
21.4
27.6
22.8
22.8
22.9
23.0
26.4
25.8
27.3
20.5
28.9
29.6
33.1

Maximum Temp Wellington °C
21.6
21.5
20.9
22.0
23.5
18.8
18.0
22.2
19.2
24.0
24.6
19.5
24.6
25.0
22.2
21.6
20.5
21.4
19.9
16.1
18.6
19.7
16.0
20.2
21.8
25.6
25.5
27.4
23.6
23.1
Not given
Not given
Not given

Task One: Holiday Temperatures **Name:______________________**

Part B

Analyse the data in order to answer your question.

Use the data for Napier and Wellington to *answer your question* or *test your hypothesis*. The following instructions will help you do this.

1. Calculate statistics for Napier and Wellington. These must include
   at least one measure of central tendency
   at least one measure of spread.

2. On the graph paper provided draw appropriate graphs(s) that allow you to answer the question or test the hypothesis you posed.

3. *Respond to your question or hypothesis.* Refer to your statistics and the features of the graph(s). Use these to support your conclusion. Make 3 statements that justify your conclusion.

4. Write an evaluation of the statistical process. Aim to make 3 statements. If you make more than 3 statements *select* the 3 statements which you think are the best.

Your statements could refer to some of the following aspects:
- Other data that could be collected to improve the analysis.
- Improvements to the method of data collection.
- Better graphs that could be drawn, or other statistics that could be calculated. (If you believe that you have made the best choice of graph(s) and statistics explain why).
- The validity of your conclusions.
- Improvements to the question posed.

APPENDIX B

Task Two: Cloud Blanket

Part B

Jason read a European magazine about how clouds act as a warm blanket. The article said that the clouds keep the heat in and therefore prevented the temperature dropping very much. He decided to analyse the data he collected from the newspaper to see if this was true.

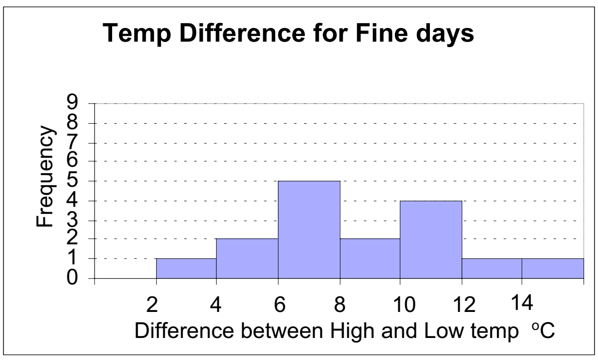
He thought that if the ***clouds keep the heat in*** then the ***difference*** between the high and low temperatures will be ***less*** on cloudy days compared to fine days.

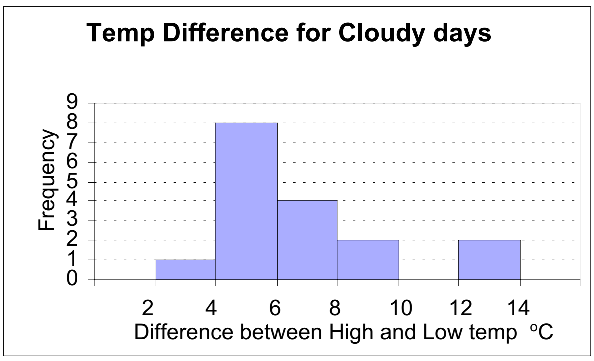
Jason took all the data for the Pacific/Australia and New Zealand and categorised them into 'Fine' and 'Cloudy'. He then took the high and subtracted the low to give the difference. His tables are shown below.

	Clear Sky Temp		
	High	Low	Difference
clear	23	11	12
fine	32	23	9
fine	31	24	7
fine	20	15	5
fine	26	18	8
fine	22	15	7
fine	29	24	5
fine	28	24	4
fine	25	13	12
clearing	13	1	12
fine	18	11	7
fine	19	12	7
fine	17	7	10
fine	18	7	11
fine	18	4	14
fine	16	1	15

	Cloudy/Showers Temp		
	High	Low	Difference
rain	16	7	9
showers	16	11	5
showers	17	12	5
showers	17	11	6
showers	13	4	9
showers	11	6	5
cloudy	16	10	6
showers	31	25	6
showers	27	20	7
showers	14	10	4
showers	19	13	6
showers	31	23	8
showers	30	22	8
showers	31	24	7
showers	29	24	5
cloudy	25	11	14
cloudy	18	5	13

Jason drew the following graphs with this data. He also calculated a few statistics.





Fine Days	°C
Mean difference	9.1
Median difference	8.5

Cloudy Days	°C
Mean difference	7.2
Median difference	6.0

1. Comment on Jason's theory that the clouds keep the heat in. Justify all comments using features of the graphs and/or statistics. (Make 3 statements)

2. Write an evaluation of the statistical process that Jason used for his theory that the clouds keep the heat in. Aim to make 3 statements about how his analysis can be improved. If you make more than 3 *select* the best 3 statements.

Your statements could refer to some of the following aspects:

- Other data that could be collected to improve the analysis.
- Improvements to the method of data collection.
- Better graphs that could be drawn, or other statistics that could be calculated. (If you believe that he has made the best choice of graph(s) and statistics explain why)
- The validity of your conclusions.

PRE-SERVICE ELEMENTARY SCHOOL TEACHERS' METAPHORS FOR THE CONCEPT OF STATISTICAL SAMPLE

RANDALL E. GROTH
Salisbury University
regroth@salisbury.edu

JENNIFER A. BERGNER
Salisbury University
jabergner@salisbury.edu

ABSTRACT

The study describes the nature of pre-service teachers' idiosyncratic metaphors for the concept of statistical sample. These metaphors were investigated because of their potential to provide insight about individuals' content knowledge and how that content knowledge is enacted during teaching. Personal metaphors were elicited from 54 pre-service teachers through writing prompts. The writing prompt responses revealed seven different categories of thinking. In some instances, pre-service teachers struggled to construct a metaphor for the concept of sample. In the majority of cases, they constructed a metaphor for sample and discussed its relationship to their knowledge of the concept. The categories of thinking highlight some of the aspects of the concept of sample that teacher educators need to attend to over the course of instruction, and they also point out directions for further research related to metaphorical thinking about statistical content and its interaction with teaching practice.

Keywords: *Statistics education research; Sample; Metaphor; Teacher education*

1. INTRODUCTION

The concept of sample is foundational to the practice of statistics. In many cases, statisticians are interested in drawing conclusions about a large population when it is impractical to gather data from each of its members. Well-designed samples lead to viable conclusions about the population of interest. Because of the importance of samples in statistical reasoning, various curriculum documents recommend that students begin to develop understanding of the concept during elementary school (e.g., National Council of Teachers of Mathematics, 2000; American Association for the Advancement of Science, 1993).

Teachers' content knowledge impacts their abilities to carry out curriculum recommendations (Ball, Lubienski, & Mewborn, 2001; Shulman, 1987). Because of this, the development of content knowledge is an important part of many university-level mathematics courses for pre-service elementary teachers (Conference Board of the Mathematical Sciences, 2001; National Research Council, 2001). Teacher education

Statistics Education Research Journal, 4(2), 27-42, http://www.stat.auckland.ac.nz/serj

efforts can be highly effective when curricula are designed to be responsive to teachers' thinking (Mewborn, 2003). Curricula designed without the benefit of a research base are likely to be ineffective (Battista & Clements, 2000). The present study sought to inform the construction of curricula aimed at building pre-service teachers' statistical content knowledge by examining their metaphorical thinking in regard to the concept of sample.

2. METAPHOR AS A WINDOW ON TEACHERS' THINKING

Metaphor can be defined as, "A figure of speech in which a word or phrase literally denoting one kind of object or idea is used in place of another to suggest a likeness or analogy between them" (Mish, 1991). Under this definition, "metaphor is a window" is itself an example of a metaphor. Any metaphor one constructs will have both ground and tension (Presmeg, 1998). Similarities between the objects involved in a metaphor constitute the ground, while dissimilarities constitute the tension. For instance, the ground of the metaphor "metaphor is a window" includes the aspect that both objects allow one to observe phenomena that might otherwise remain unseen. The tension of the metaphor includes the aspect that a window is a physical object, while a metaphor is not. The reader can likely identify various other similarities and dissimilarities between "windows" and "metaphors."

2.1. METAPHOR IN MATHEMATICAL THINKING

Metaphor plays a prominent role in mathematical thinking. Presmeg (1992) illustrated how personal metaphors can influence problem-solving behavior. She gave an example of a high school student named Alison who thought of the x-axis as a "water level" in working with trigonometric ratios. Her problem-solving behavior was described in the following manner:

> Alison…used the metaphor of a ship sailing on a water level (x-axis) to remind her to use 180 degrees and 360 degrees in obtaining the acute angle she requires for the trigonometric ratios rather than 90 degrees or 270 degrees which would be contrary to her metaphor and would give an incorrect acute angle for the trigonometric ratio she is using (Presmeg, 1992, p. 600).

Alison's self-invented, idiosyncratic metaphor played a significant role in her mathematical activity. It should be noted that this is not simply an isolated case of metaphor playing a role in mathematical problem-solving. Other common mathematical metaphors include thinking of a function as a machine (Cuevas & Yeatts, 2001) and thinking of the arithmetic mean as a balance point for a set of numerical data (Mokros & Russell, 1995).

Reliance upon metaphors to guide mathematical thinking is not a phenomenon restricted to students. Sfard (1994) discussed how the thinking of mathematicians is guided by metaphors. She argued that commonly accepted and used mathematical words such as "increasing and decreasing," "closed and open," and "saturated and stable" all "clearly have their roots in the world of material objects" (Sfard, 1994, p. 47). Further, the mathematicians she interviewed stressed the use of analogy and metaphor in their work. For example, some spoke of using analogies mapping elements from one domain of mathematics to another to guide conjecture formation. Lucid analogies played a role in providing evidence that a result from one mathematical situation was likely to carry over to another similar situation.

Given the prominent role of metaphor in mathematical thinking, Lakoff and Nunez (2000) have gone so far as to argue that the very substance of mathematical thinking is metaphorical in nature. While a comprehensive critique of their argument is beyond the scope of this paper, it is important to note that Lakoff and Nunez (2000) have been cited for their lack of attention to the role that idiosyncratic metaphors play in individuals' mathematical thinking and learning (Presmeg, 2002). Such metaphors can provide instructors with valuable insights about students' construction of knowledge (Presmeg & Bergner, 2002). Therefore, the investigation of individuals' idiosyncratic metaphors for the concept of sample was the focus for the present study.

2.2. METAPHOR IN THINKING ABOUT STATISTICAL SAMPLES

In studies concerned with the nature of students' thinking in regard to statistical samples, elements of metaphorical thinking have been revealed even when the researchers were not explicitly looking for them. For example, when Jacobs (1997) posed the question, "What is a sample?" to middle school students, one response was that a sample was "a piece of food or carpet [that] gives you an idea of what the real thing is" (p. 28). This particular student appeared to make sense of the idea of sample metaphorically, by relating it to concrete objects. There were other similar instances, in which samples were thought of in terms of grocery store food samples, samples of products sent in the mail, and student writing samples (Jacobs, 1997). Jacobs (1999) subsequently recommended that teachers need to be aware of and build on children's informal thinking about samples during instruction.

In another study, when Watson and Moritz (2000a) posed the question, "If you were given a 'sample,' what would you have?" to a group of students, the students also related the concept to concrete objects. One student, for example, stated, "I'd have, if it was for clothes, a small piece of material" (Watson & Moritz, 2000a, p. 119). A unique aspect of the Watson and Moritz (2000a) study is that they used the SOLO Taxonomy (Biggs, 1992; Biggs & Collis, 1982) to place a hierarchical structure on the categories of children's definitions. Four distinct levels were identified. At the lowest level, students' responses did not capture any relevant aspects of the concept of sample. At the second level, responses included the idea that a sample is a "bit," or "part," but made no reference to a whole. Responses at the third level clearly characterized a sample as a part of a whole. The highest level observed included responses that characterized a sample as a representative part of a whole. Watson (2004) found that students generally moved to higher levels of response about the concept of sample as they progressed through school, suggesting that the hierarchy may map a developmental sequence.

The Jacobs (1997, 1999) and Watson and Moritz (2000a) studies suggest that metaphorical thinking plays a role in the development of understanding of the statistical idea of sample. Personal metaphors for "samples" reveal aspects of an individual's thinking about them. Therefore, in the present study, we elicited idiosyncratic metaphors from pre-service teachers in order to better understand their content knowledge in regard to the concept of sample, and also compared the content knowledge displayed against the hierarchy described by Watson and Moritz (2000a).

2.3. IMPACT OF METAPHOR ON TEACHING ACTIONS

While we expected pre-service teachers' idiosyncratic metaphors to provide a window on the structure of their content knowledge, we also expected the metaphors to provide some insight about the role their content knowledge might play in teaching. Educators in the fields of mathematics, science, and statistics education have recommended the use of metaphor and other types of analogy as teaching devices (delMas, 2004; English, 1999; Glynn, 1991; Martin, 2003). When teachers share well-formed concrete analogies with students it can enhance their learning (Mayer, 1987). Gentner and Holyoak (1997) provided some insight about why analogies can be helpful to learning. They stated,

> One basic mechanism [for learning] is analogy – the process of understanding a novel situation in terms of one that is already familiar. The familiar situation – often termed the base or source analog – provides a kind of model for making inferences about a particular situation – the target analog (p. 32).

By examining the metaphors for sample constructed by pre-service teachers, we expected to gain some insight about their ability to construct and use metaphors in their own classrooms. Therefore, while learning about teachers' content knowledge, we also expected to gain insight about their pedagogical content knowledge (Shulman, 1987), which enables teachers to make subject matter understandable to students.

It should also be noted that even if teachers do not consciously construct and share metaphors and other analogies during the course of teaching, students' thinking can be influenced by the spontaneous metaphors implicit in teachers' presentations of subject matter. Presmeg (1992) argued that although students ultimately must construct their own individual meanings for mathematics concepts, ideas encountered within classrooms have an impact on the sort of meanings that are constructed. She provided an example of a student who had a higher level of mathematics achievement than other students in his class because his thinking was influenced by his teacher's operating metaphor of "pure logic is beauty." The metaphor was bound together with the mathematical notions of generalization, abstraction, and patterns. Hence, even if teachers are not aware that they are sharing metaphors in teaching, the metaphors they unknowingly convey and operate under can have an impact on the meanings their students construct for concepts, because students are likely to construct meanings at least partially from their interactions with teachers in classroom communities.

2.4. SUMMARY

Metaphor is a useful tool for examining individuals' mathematical and statistical knowledge. Individuals' idiosyncratic metaphors for concepts provide a window on their understanding of them. In addition, the personal metaphors held by teachers are related to the manner in which they are able to convey information to students, and also to their abilities to help develop students' understanding. Hence, the purpose of the present study was to elicit and analyze idiosyncratic metaphors generated by pre-service elementary teachers in regard to the concept of statistical sample. We expected this analysis to provide insights about factors teacher educators should consider in developing instructional programs to enhance pre-service elementary teachers' statistical content knowledge. Just as teachers should build on students' informal conceptions of sample during instruction (Jacobs, 1999), teacher educators need to become aware of and build on teachers' conceptions of sample, because successful models of teacher education use teachers' thinking as a starting point for instructional design (Mewborn, 2003).

3. METHODOLOGY

In the present study, we collected and analyzed qualitative data in order to explore the nature of pre-service teachers' idiosyncratic metaphors for sample. We used a qualitative design because the study sought to investigate and describe thinking patterns among a given group of pre-service teachers rather than to make statistical generalizations to a larger population (Bogdan & Biklen, 1992; Strauss & Corbin, 1990). The specific components of that design are outlined in this section.

3.1. PARTICIPANTS

Idiosyncratic metaphors were elicited from 54 pre-service elementary teachers. The pre-service teachers were distributed among three different sections of a university-level mathematics teaching methods course at a university in the Eastern United States. They comprised approximately 75% of the population of students enrolled in all of the sections of the course. Four of the participants were male and 50 were female. This distribution of gender is not at all atypical in the U.S., as currently just 9% of all elementary school teachers in the country are male (National Education Association, 2005).

At the time of the study, statistical content and teaching methods had not yet been topics of study in the participants' teaching methods course. Each participant had, however, taken and passed an introductory statistics course at the college level before taking the methods course that provided the setting for the present study. Some had taken the introductory statistics course at the same university where the methods course took place, while others had transferred the courses from other institutions. Regardless of where it had been taken, it was required to satisfy the following course description: "descriptive and inferential analysis of raw data, emphasizing appropriate assumptions, computer use, and interpretation; consideration of parametric and non-parametric methods and comparison of their powers." While this course description leaves considerable room for interpretation by individual instructors, the essence of the required course mirrored what Cobb and Moore (1997) characterized as the traditional introductory statistics course, where heavy emphasis is placed on selection and use of formal probability models. In such a course, the idea of sample is implicit in concepts such as random samples and sampling distributions, but it may or may not be explicitly defined.

3.2. DATA GATHERING PROCEDURE

During a session of the participants' teaching methods course, they were introduced to the idea of using metaphors in teaching mathematical concepts. Following Glynn (1991), the first author, who was the instructor of the course, recommended to the participants that when using metaphors to teach concepts, the teacher should discuss both the ground of the metaphor and its tension. If such discussions do not take place, students may be more likely to make invalid conclusions about which aspects of a concept a metaphor captures and which it does not. To illustrate the application of the Glynn (1991) model for using metaphor in teaching, the methods course instructor asked the participants to consider the ground and the tension inherent in the metaphor "a function is a machine." Participants identified several aspects of the ground, such as the fact that there is an output for any given input in a function and also in some types of machines. They also identified several aspects of the tension, such as the fact that a machine is a physical object, while a function is not. This was the first point during the methods course when the idea of metaphor was discussed.

Immediately following the initial discussion of the function machine metaphor, participants were given a writing prompt (Figure 1) in which they were asked to write a metaphor for the concept of statistical sample, and to identify the ground and tension inherent in the metaphor they had written. Writing prompts were chosen as a data collection method because of their empirically demonstrated ability to reveal aspects of students' thinking related to mathematical concepts (Aspinwall & Aspinwall, 2003; Miller, 1992). It should be noted that one of the limitations of using writing prompts to elicit thinking is that students sometimes do not express everything they know about any given piece of content in writing (Aspinwall & Aspinwall, 2003). In order to minimize the impact of this limitation, we informed participants that it was important for their writing to be as reflective of their knowledge as possible, because the writing prompts would be used to design instruction for the methods course. Participants were also told in advance that with their permission, the material in the writing prompts would be used for this study. While there may still have been aspects of the participants' thinking that were not revealed in their writing, we chose to use writing prompts because of their demonstrated ability to enhance teachers' knowledge of students' thinking and in turn influence teaching (Miller, 1992). The use of writing prompts also facilitated the collection of data from a fairly large number of study participants.

Metaphor Construction Exercise
Background and Instructions

Definition

Metaphor: A figure of speech in which a word or phrase literally denoting one kind of object or idea is used in place of another to suggest a likeness or analogy between them (Mish, 1991).

Example of a metaphor

The world is a stage.

Discussion

Consider the metaphor "the world is a stage." There are many reasons why the world can be thought of as a stage (people are actors, life is a drama, etc.). However, there are some characteristics of the world that the idea of "stage" does not capture (its size, chemical composition, etc.). This is true of metaphors in general. Any given metaphor will capture some of the characteristics of any given idea while missing others.

Task

In this exercise, I am asking you to write a metaphor for a statistical idea. After writing the metaphor, discuss how and why the metaphor works (which characteristics of the statistical idea are captured). Also discuss how and why it does not work (which characteristics of the statistical idea are missed or perhaps misrepresented). Please do the exercises in writing, individually.

Statistical idea: Sample

(1) Write a metaphor for the statistical idea.
(2) Explain how and why your metaphor works.
(3) Explain how and why your metaphor does not work.

Figure 1. Metaphor Writing Prompt Given to Study Participants

3.3. DATA ANALYSIS

Both authors read the writing prompt responses generated by the participants. The first author used a clustering procedure (Miles & Huberman, 1994) to group similarly-structured responses into categories. The research pertaining to students' understanding of sample discussed earlier (Jacobs, 1997; Watson, 2004; Watson & Moritz, 2000a) highlighted relevant cognitive issues to which to attend in judging whether or not a group of responses was similarly-structured. For example, some participants' responses conceptualized a sample as a representative part of a whole, and since this sort of response was similar to a level of thinking documented by Watson and Moritz (2000a), the responses were clustered together. The second author then independently analyzed the responses and attempted to place them into the categories formed by the first author. At the conclusion of the independent analyses, the authors agreed on the placement of 30 out of the 54 responses.

After analyzing writing prompt responses independently, the authors met in order to determine the source of the 24 disagreements in coding. Ten of the 24 disagreements were caused by responses that were ultimately judged to be pedagogically awkward in their characterizations of the concept of sample. These ten instances became a new category, and its characteristics along with sample responses are given in the results section. The rest of the disagreements were resolved as the authors discussed their rationale for coding. Ultimately, consensus was reached about the category in which each response was to be placed.

4. RESULTS

Seven different categories of responses to the metaphor writing prompts were formed during the data analysis process. The names of the categories formed and the numbers of responses fitting each one are shown in Table 1. The characteristics of each category along with some sample responses are discussed in this section.

Table 1. Categories of Response to Metaphor Writing Prompt

Category descriptor	Frequency
No Metaphor Provided	3
Sample as a Collection of Objects	4
Sample as Part of a Whole	23
Sample as a Representative Part of a Whole	10
Metaphor for Place of Sample within Mathematics and Statistics	2
Metaphors Describing Actions to be Taken Upon Samples	2
Pedagogically Awkward Characterizations	10

4.1. CATEGORY 1: NO METAPHOR PROVIDED

Three participants gave no metaphor in response to the writing prompt. One of the three participants wrote, “A sample is not satisfactory” in response to the request to write a metaphor for the statistical idea of sample. Another of the three wrote, “A sample is” and did not finish her thought. The last of the three did not write anything in response to the prompt. The fact that three participants did not write any metaphor appears to have been caused by the cognitive complexity of thinking metaphorically rather than unwillingness to cooperate with the writing prompt exercise. None of the students handed in a paper without first spending time thinking through the process of writing an appropriate metaphor, but several remarked that they found it difficult to generate one.

This first category of response strongly resembles the lowest category in the hierarchy described by Watson and Moritz (2000a). In that category, students' definitions did not capture any of the relevant aspects of the concept of sample. Likewise, in this first category, responses were off-target. It should be noted, however, that the reason for the off-target responses in this category in the present study may well be due to cognitive difficulties in producing a metaphor rather than with lack of statistical content knowledge.

4.2. CATEGORY 2: SAMPLE AS A COLLECTION OF OBJECTS

Four of the writing prompt responses characterized a sample as a collection of objects with no mention of a larger population from which the objects were drawn. One participant provided the metaphor, "A sample is a forest," explaining that, "a forest is a group of trees." Another wrote the metaphor, "A sample is a car lot," because "a sample is like a group." The third of the students in the first category wrote "A sample is pizzas in a pizza parlor." She justified the metaphor by saying, "This shop has different sizes and types of pizza, just like a sample is dynamic. There are different aspects that come together to make the sample a whole." The final response in the first category started with the metaphor, "A sample is a bag of bagels," because "it includes more than one item from a wide range of possibilities." While each of the metaphors in the second category related the idea that samples often consist of diverse members or objects, they did not characterize samples as parts of larger populations.

The second category of response in the present study connects to the second level in the Watson and Moritz (2000a) hierarchy. In both instances, a sample is characterized as a part, but at the same time, the whole from which the part is drawn is not described.

4.3. CATEGORY 3: SAMPLE AS PART OF A WHOLE

The most common type of response characterized a sample as part of a whole. Twenty-three responses fit this category. Some examples include:

> Metaphor: A sample is one toy off of a toy shelf. Ground: My metaphor works because one toy is a sample from the toy shelf. Tension: My metaphor does not work because a sample does not always have to be a toy off of a toy shelf.

> Metaphor: A sample is the sugar in a cookie. Ground: To make a cookie, sugar is one of the ingredients. Likewise, a sample is one part of a whole set. Tension: You need more than just sugar to make a cookie and more than just one number to complete a set of data.

> Metaphor: A sample is a piece of the pie. Ground: A sample is one piece of data from an entire experiment. A piece of pie is only one part of the whole. Tension: A sample is only supposed to be one person's data, not the whole.

In the discussions of ground and tension for responses in this category, there was no mention of a sample being a representative part of a whole. While some of the metaphors provided, such as "a sample is a piece of a pie," had the potential to be used to illustrate the idea of representativeness (e.g., a piece of pie could be thought of as representative of the taste of the entire pie), that idea was not touched upon in accompanying discussion of ground and tension. The idea that a sample, as part of a whole, can serve to represent the whole, was absent from responses in the third category.

The third category of response in the present study is structurally similar to the third level in the Watson and Moritz (2000a) hierarchy. In the third level of their hierarchy, students conceptualized a sample as part of a whole, and explicitly referred to the population from which the sample was drawn. This pattern resonates with pre-service teacher responses in the third category, since in each instance the whole population was an integral part of the metaphors that were constructed (e.g., in the case of the toy shelf metaphor, the toy was thought of as a sample, and the toy shelf was referred to as the population from which it was drawn).

4.4. CATEGORY 4: SAMPLE AS A REPRESENTATIVE PART OF A WHOLE

Ten writing prompt responses characterized a sample as a representative part of a whole. This group of responses generally contained the richest descriptions and most vivid ideas. Some typical responses include:

> Metaphor: A sample is a handful of cereal out of the cereal box. Ground: My metaphor works because it is an equal representation of the box of cereal. It is taking a part of the whole and examining it instead of the whole. Tension: My metaphor doesn't work because every population is not a box of cereal.

> Metaphor: A sample is a handful of M&M's out of the candy dish. Ground: The metaphor shows that the sample is a smaller group than what is being studied. A handful of M&M's represents the whole dish of M&M's like a small group can represent a whole group. You don't need to take all the M&M's because a handful is the same. Tension: It doesn't work because all samples are not candy or bright colored.

> Metaphor: A sample is one WalMart store. Ground: A sample is a way of looking at a large amount of data by examining a small amount that can represent it. When you've been in one WalMart store – you've been in them all. Therefore – one WalMart can be a representative of all of them. Tension: A sample isn't really a store.

The responses in this category included the idea of sample as a part of a whole that was present in the previous category. However, they also incorporated a unique dimension in characterizing a sample as representative of the population from which it is drawn.

Since category 4 in the present study involved the conceptualization of sample as a representative part of a whole rather than simply part of a whole, the category strongly resonates with the fourth level in the Watson and Moritz (2000a) hierarchy. It is a significant conceptual leap to think of a sample as representative of the population from which it was drawn. This idea was missing from the previous category (e.g., the "toy on the toyshelf" metaphor and its accompanying discussion said nothing about how or if the toy selected might tell something about the larger population of toys on the shelf).

4.5. CATEGORY 5: METAPHOR FOR PLACE OF SAMPLE WITHIN THE FIELD OF STATISTICS

Two writing prompt responses were very different from the others in that they presented metaphors for the place held by the concept of sample within statistical data analysis. One of these students wrote, "A sample is a pathway on a long journey." Her discussion of ground for the metaphor stated, "Taking a sample is only part of the process of data analysis, so it is like a journey because at some pathway you can choose to turn left or right, and, just like a sample, this determines the end result." The second student in this category wrote that "Statistics is one small fish in a big ocean." The explanation of the ground for the metaphor read, "It works because statistics is only part of the math world. There is also algebra, trig, calculus, geometry, and so on. A sample is an even smaller part of math because it is a part of statistics." These two metaphors mark a sharp break from those in the previously discussed categories, since they seek to situate the concept of sample within the disciplines of statistics and mathematics rather than to illustrate intrinsic characteristics of the concept itself.

These two writing prompt responses appear to fall outside the levels described in the Watson and Moritz (2000a) hierarchy. None of the students in that study were reported as describing the place of sample within the field of statistics. Part of the reason for this may be that the students in the Watson and Moritz (2000a) study were younger, and hence had less opportunity to think about how various branches of mathematics fit together. Another part of the reason for this new category of response may be that the task was slightly different. Whereas Watson and Moritz (2000a) requested a definition of "sample," in the present study a metaphor was requested. Asking students to write a metaphor may trigger different cognitive processes than asking for a definition.

4.6. CATEGORY 6: METAPHORS DESCRIBING ACTIONS TO BE TAKEN UPON SAMPLES

In two instances, participants characterized sample as a collection of data from which statistics are to be calculated. In one such instance, a student stated that "A sample is a phone bill." She justified the metaphor by stating,

> A sample is a bunch of data that you get and perform computations with to get more data. Once you have the data, you might find the mode, median, mean, range, etc. of the data to draw further conclusions. Similarly, when you receive a phone bill you are receiving a whole bunch of data…Then, from this info. one might calculate mode, median, mean, etc., to figure out who they made the most calls to, what was the length of their longest call, what time most of their calls were made, etc.

The other response in this category began by characterizing a sample as, "a small part of the big picture." The discussion of the tension for the metaphor included the thought that the metaphor "doesn't describe any parts of the math process involved," again implying that mathematical calculations are keys to defining the concept of sample. This sort of characterization of sample was similar to that conveyed by textbook problems which present a sample of numerical data and ask students to perform mathematical operations upon it.

This sixth category appears to be another that is not related to the Watson and Moritz (2000a) hierarchy. As with the fifth category, it is possible that both the structure of the task and the ages/experiences of the students involved played roles in this occurrence.

Experiences of students may play an especially pronounced role, since it is conceivable that years of doing nothing but performing computations on numerical samples would foster a resilient belief that samples exist for the purpose of being acted upon with arithmetic algorithms. This sort of "number crunching" characterization is not an uncommon portrayal of the field of statistics in traditional school curricula. Scheaffer (2002) noted that, "Statistics is often presented as a collection of techniques and tools rather than as a process for quantitative reasoning and problem solving" (p. 6). The responses in category six may possibly reflect a strong indoctrination into that particular view of statistics.

4.7. CATEGORY 7: PEDAGOGICALLY AWKWARD CHARACTERIZATIONS

In ten cases, it was judged that writing prompt responses were pedagogically awkward in their characterization of sample. For example, one participant in this category used the metaphor of "one in a million" for the concept of sample, which suggests that samples are generally not representative of the populations from which they are drawn, since the phrase generally is taken to mean that one has encountered something or someone unique. However, the same participant who constructed the "one in a million" metaphor stated that the tension of the metaphor was "the one sample does not accurately represent the population." Hence, the metaphor that was constructed appeared to obscure the participant's actual conception of sample to a large extent. Therefore, such responses were considered pedagogically awkward because of their inhibited potential to communicate meaning.

Other responses placed in the seventh category offered metaphors that were not clearly written or explained. For instance, one participant wrote, "a sample is like an example" without explaining why that would be an apt metaphor. Another wrote, "a slice is a piece of a pie," leaving the authors to speculate that she may have meant to write that a sample is a slice of pie or a piece of pie. However, those meanings could not be directly inferred from the writing. Since the authors agreed that these types of responses were highly ambiguous, they were considered likely to be awkward in conveying meaning in a teaching situation.

Since teachers' content knowledge regarding the concept of sample was not clearly displayed in responses fitting the seventh category, it is not possible to say where these participants may fall in relation to the Watson and Moritz (2000a) hierarchy. While it is difficult to gain any insight about the statistical content knowledge of teachers whose responses fit this seventh category, some information is gained about their pedagogical content knowledge (Shulman, 1987). It appears that the ten teachers whose responses fit the seventh category needed further learning experiences before they could incorporate metaphor as a teaching tool for helping students understand the important statistical aspects of the concept of sample.

5. DISCUSSION

The present study, exploratory in nature, provides directions for further research and some pedagogical implications. To conclude, both areas will be discussed.

The present study holds implications for instruction designed to help teachers use metaphors and other analogies explicitly as tools for articulating statistical ideas in everyday language. In discussing analogy as a teaching tool in the field of science education, Glynn (1991) suggested that teachers identify the ground and the tension in the analogy for students after introducing it. Participants' responses highlight that it is also important for teachers to attend closely to the initial construction of the metaphor. If a metaphor with a great deal of tension is used, the power of the metaphor may well work against the content that the teacher desires to convey, even if the tension inherent in the metaphor is explicitly discussed with students. For instance, using the metaphor of "one in a million" for sample creates a strong initial impression that samples identify a unique aspect of a group rather than serving the purpose of representing it. Although the teacher may not mean to convey this meaning, the power of the initial metaphor has the potential to obscure the meaning the teacher wishes to convey. Teacher educators might encourage pre-service teachers to share their idiosyncratic metaphors with one another and with young students to obtain a sense of what types of images the metaphors tend to invoke in others' minds.

The categories of participants' responses also suggest areas for teacher educators to focus upon in developing pre-service teachers' statistical content knowledge. Some participants exhibited very limited notions of the concept of sample by characterizing it simply as a group of objects or a collection of numerical data to perform mathematical procedures upon. Teacher educators must be conscious of designing instruction to expand and revise the personal metaphors of pre-service teachers holding such limiting conceptions. If limited personal metaphors are not replaced with richer ones, those metaphors may well constrain the types of tasks in which they have their future students engage, since one's personal metaphors guide teaching actions (Chapman, 1997; Sfard, 1998).

While it is beyond the scope of the present study to decisively state which teaching methods best foster understanding of the concept of sample, directions for further research in this area are illuminated by participants' responses to the metaphor task. Of particular interest is the fact that less than 20% of the pre-service teachers mentioned representativeness in discussing samples. Further research might focus on investigating the extent to which various teaching methods are able to improve on this figure. Various pedagogical methods for introducing the concept of sample are suggested in existing published texts. Many of the participants in this study had taken a course using a text which first formally defined a sample as a subset of units from a population, and then later gave a separate formal definition for "representative" sample (McClave & Sincich, 2003). Other authors take the approach of beginning with a formal definition for sample which encompasses the idea that information gathered from a sample is used to draw conclusions about the population from which it is drawn (e.g., Moore, 1997). Of course, these two approaches do not exhaust all possible patterns of teaching the concept, as individual instructors are likely to have their own pedagogical patterns. Even those who use the aforementioned texts in their courses may deviate from the text's presentation, as the interactions between intended and implemented curricula are often complex (Schmidt, McKnight, & Raizen, 1997). It would be beneficial to have future studies that focus on the effectiveness of various teaching approaches in building understanding of the representative functions of statistical samples and the role that formal definition plays in developing students' understanding.

Another possible direction for further research would be to attempt to determine how widespread the documented categories of metaphorical thinking are among pre-service teachers. Although no statistical generalizations were sought in the present study, we do expect some overlap between the characteristics of the group studied and other groups of pre-service elementary teachers, therefore making the study valuable for informing pre-service teacher instruction in other settings (Eisner, 1998; Glesne, 1999). Researchers desiring statistical generalizability could take the categories described in the present study as a working framework for larger-scale studies.

While the methodology used to ascertain teachers' thinking patterns in the present study yielded some useful data, it is not without its limitations. Eliciting written idiosyncratic metaphors yielded a wide range of categories of response, but even richer data could be gathered in future studies by coupling written metaphors with interviews in which participants are asked to elaborate further upon their personal metaphors. Such interviews would allow researchers to further unpack reasons for impoverished and ill-formed metaphors. Also, while personal metaphors provide a window on an individual's thinking about a mathematics concept, it is important for future studies to also investigate the relationship between elicited and spontaneous metaphors. It may be in some cases the spontaneous metaphors that occur in the course of instruction do not match the metaphors that are constructed upon the request of a researcher. It is important to understand teachers' spontaneous metaphors that occur during the course of instruction, given that they can influence students' thinking (Presmeg, 1992). As we begin to understand the types of spontaneous metaphors used in instruction and how they relate to elicited metaphors, we can become more effective in helping teachers re-think and re-construct the metaphors that guide their instruction of content.

Finally, it should be noted that the present study deals largely with describing teachers' knowledge of the concept of sample, but not with its use in problem solving situations. Therefore, we primarily learned about teachers' thinking within what Watson (1997) called "tier 1" of statistical literacy, which involves understanding of statistical terminology. Tier 2 involves "An understanding of statistical language and concepts when they are embedded in the context of wider social discussion" (Watson, 2000a, p. 54) and tier 3 involves, "A questioning attitude that can apply more sophisticated concepts to contradict claims made without proper statistical foundation" (Watson, 2000a, p. 54). Studies that delve into tiers 2 and 3 with an eye toward the concept of sample have, however, taken place. For example, Metz (1999) explored children's conceptions of sampling within the context of designing their own studies, and found that many of them were not convinced of the power of sampling even after designing and executing a research project. Also, Watson (2000b) has investigated pre-service teachers' strategies for solving the famous "hospital problem" (Kahneman & Tversky, 1972) and found that some strategies were based on intuition, some on mathematical principles, and others on both. Finally, the research of Watson and Moritz (2000a, 2000b) provides more examples of tasks that can be used to begin to understand the nature of thinking in regard to sample within tiers 2 and 3. Partially replicating these studies (Metz, 1999; Watson, 2000b; Watson & Moritz, 2000a, 2000b) while maintaining an eye toward the spontaneous metaphors teachers construct in such problem solving situations could yield rich insights about their understanding of sample within tiers 2 and 3 of statistical literacy.

6. CONCLUSION

Personal metaphors impact the manner in which individuals function on a day-to-day basis. This small-scale study brought to light a variety of personal metaphors that pre-service teachers hold for the concept of statistical sample. In some cases, the metaphors were rich and captured a number of important attributes of the concept. In other cases, the metaphors were impoverished and in need of further development. The impoverished patterns of thinking suggest that pre-service teachers need to experience problem-solving activities, such as those that require drawing samples for the purpose of making an inference to a larger population (e.g., Morita, 1999), that allow them to begin to conceive of a sample as more than just a collection of objects or a set of numerical data from which statistics are to be calculated. Teacher educators also need to construct learning trajectories (Simon, 1995) that aid the construction of richer personal metaphors. The present study provides a starting point in the construction of such trajectories, since it provides an overview of the patterns of thinking one might expect to encounter among pre-service teachers at the beginnings of interventions that include the goal of enhancing knowledge of the concept of statistical sample. This ultimately can improve statistics education for teachers by allowing pre-service curricula to have their foundation built on knowledge of teachers' statistical thinking.

REFERENCES

American Association for the Advancement of Science (1993). *Benchmarks for scientific literacy*. New York: Oxford University Press.

Aspinwall, L., & Aspinwall, J. S. (2003). Investigating mathematical thinking using open writing prompts. *Mathematics Teaching in the Middle School, 8*, 350-353.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 433-456). New York: Macmillan.

Battista, M. T., & Clements, D. H. (2000). Mathematics curriculum development as a scientific endeavor. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 737-760). Mahwah, NJ: Erlbaum.

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

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

Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.

Chapman, O. (1997). Metaphors in the teaching of mathematical problem solving. *Educational Studies in Mathematics, 32*, 201-228.

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801-823.

Conference Board of the Mathematical Sciences (2001). *The mathematical education of teachers*. Providence, Rhode Island: American Mathematical Society.

Cuevas, G. J., & Yeatts, K. L. (2001). *Navigating through algebra in grades 3-5*. Reston, VA: National Council of Teachers of Mathematics.

delMas, R. (2004). Mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield, (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-95). Dordrecht, The Netherlands: Kluwer.

Eisner, E. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, NJ: Merrill.

English, L. D. (1999). Reasoning by analogy: A fundamental process in children's mathematical learning. In L.V. Stiff & F.R. Curcio (Eds.), *Developing mathematical reasoning in grades K-12* (pp. 22-36). Reston, VA: National Council of Teachers of Mathematics.

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *American Psychologist, 52*, 32-34.

Glesne, C. (1999). *Becoming qualitative researchers: An introduction* (2nd ed.). New York: Longman.

Glynn, S. M. (1991). Explaining science concepts: A teaching-with-analogies model. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 219-240). Hillsdale, NJ: Erlbaum.

Jacobs, V. R. (1997). *Children's understanding of sampling in surveys*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

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

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3*, 430-454.

Lakoff, G., & Nunez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York: Basic Books.

Martin, M. A. (2003). It's like…you know: The use of analogies and heuristics in teaching introductory statistical methods. *Journal of Statistics Education*, *2*. Retrieved February 14, 2005 from http://www.amstat.org/publications/jse/v11n2/martin.html.

Mayer, R. E. (1987). *Educational psychology*. Reading, MA: Addison Wesley.

McClave, J. T., & Sincich, T. (2003). *A first course in statistics* (8th ed.). Upper Saddle River, NJ: Pearson.

Metz, K. E. (1999). Why sampling works or why it can't: Ideas of young children engaged in research of their own design. In F. Hitt & M. Santos (Eds.), *Proceedings of the 21st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 492-498). Cuernavaca, Mexico: PME.

Mewborn, D. S. (2003). Teaching, teachers' knowledge, and their professional development. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 45-52). Reston, VA: National Council of Teachers of Mathematics.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Miller, D. L. (1992). Teacher benefits from using impromptu writing prompts in algebra classes. *Journal for Research in Mathematics Education, 23*, 329-340.

Mish, F. C. (Ed.) (1991). *Webster's ninth new collegiate dictionary*. Springfield, MA: Merriam-Webster.

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

Moore, D. S. (1997). *Statistics: Concepts and controversies* (4th ed.). New York: W.H. Freeman.

Morita, J. G. (1999). Capture and recapture your students' interest in statistics. *Mathematics Teaching in the Middle School, 4*, 412-418.

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

National Education Association (2005). Wanted: More male teachers. Accessed online June 24, 2005: http://www.nea.org/teachershortage/03malefactsheet.html

National Research Council (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Presmeg, N. C. (1992). Prototypes, metaphors, metonymies and imaginative rationality in high school mathematics. *Educational Studies in Mathematics, 23,* 595-610.

Presmeg, N. C. (1998). Metaphoric and metonymic signification in mathematics. *Journal of Mathematical Behavior, 17*, 25-32.

Presmeg, N. C. (2002). Mathematical idea analysis: A science of embodied mathematics – A review of where mathematics comes from: How the embodied mind brings mathematics into being. *Journal for Research in Mathematics Education, 33*, 59-63.

Presmeg, N. C., & Bergner, J. A. (2002). Cognitive elements of students' use of metaphors in a college geometry class. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 2* (pp. 755-766). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Scheaffer, R. L. (2002). Data analysis in the K-12 curriculum: Teaching the teachers. *New England Mathematics Journal, 34*(2), 6-25.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (Eds.) (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Dordrecht, The Netherlands: Kluwer.

Sfard, A. (1994). Reification as the birth of metaphor. *For the Learning of Mathematics, 14* (1), 44-55.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4-13.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*, 114-145.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: ISO Press and the International Statistical Institute.

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

Watson, J. M. (2000b). Preservice mathematics teachers' understanding of sampling: Intuition or mathematics. *Mathematics Teacher Education and Development, 2*, 121-135.

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

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

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

RANDALL E. GROTH
119 Emily Dr.
Salisbury, MD 21804
USA

STATISTICS IN THE WORKPLACE: A SURVEY OF USE BY RECENT GRADUATES WITH HIGHER DEGREES

JOHN A. HARRAWAY
Department of Mathematics and Statistics
University of Otago, New Zealand
jharraway@maths.otago.ac.nz

RICHARD J. BARKER
Department of Mathematics and Statistics
University of Otago, New Zealand
rbarker@maths.otago.ac.nz

ABSTRACT

A postal survey was conducted regarding statistical techniques, research methods and software used in the workplace by 913 graduates with PhD and Masters degrees in the biological sciences, psychology, business, economics, and statistics. The study identified gaps between topics and techniques learned at university and those used in the workplace, and points to deficiencies in statistical preparation for employment. Courses requested include multivariate statistics, generalized linear models, research design and power analysis taught with minimal emphasis on probability and mathematics. Recommendations are presented, such as expanding statistical service courses to eliminate gaps, the development of intensive workshops for postgraduate students and for workplace retraining, or involving staff from other departments to provide context for statistics teaching.

Keywords: *Statistics education research; Survey; Curriculum development; Data specialist major; Workplace needs*

1. INTRODUCTION

Like other academic topics, the teaching of statistics at the university level has to be informed, at least in part, by what graduates will have to do with their acquired statistical knowledge in their respective fields of occupation after graduation. Workplaces have unique demands and some graduates need to learn further at work as part of continued lifelong learning experience (Holmes, 1998). Yet information about workplace demands and about the actual statistical activities which graduates encounter in the workplace can be useful for planning academic curricula that are aligned with primary workplace needs. The present study was designed in light of the limited knowledge in this area.

The MEANS project (Holmes, 1998) surveyed employers and universities and sought opinion on training programmes which involved a major in statistics. However, a majority of students at the university level learn statistics as a service subject (i.e., outside departments of statistics). Indeed Holmes (1998) noted that a further survey would be needed to assess adequately the service teaching of statistics and the subsequent match with employment needs.

Statistics Education Research Journal, 4(2), 43-58, http://www.stat.auckland.ac.nz/serj

One way to characterize the statistical needs of different workplaces is by analysing the nature of statistical techniques used by researchers who operate in different fields. Harraway, Manly, Sutherland and McRae (2001) surveyed 2927 research papers published in 16 high impact journals from botany, ecology, food science, marine science and nutrition during 1999. The analysis of the results established that published research in the different subjects used a wide range of statistical procedures. The procedures used varied both between and within subject specialties. But in addition to knowing what researchers use we need to know what use is made of statistics in the workplace, especially by those who have only taken service courses in statistics at university.

An ability to manage data and use statistics software is also crucial in the workplace. Jolliffe and Rangecroft (1997) reported results from two surveys of statistical software use, one involving academic degree courses with substantial statistical content, another involving statistics consultancy organisations in the United Kingdom. The respondents to the academic survey cited the use of 21 different packages, with Minitab being by far the most popular followed by GLIM, SPSS, Genstat, SAS, S-Plus and Matlab. The respondents to the survey of consultancy firms cited SAS as the most popular package in the field, followed by Minitab, GLIM, and then equal numbers of responses for Genstat, SPSS and Statgraphics. But these surveys need to be augmented as they did not break down usage by speciality nor did they examine techniques used in the workplace.

Finally, information about the uses of statistics in the workplace can also help to design teaching in context, which may be important for success in learning statistics. The Secretary's Commission on Achieving Necessary Skills in the USA (SCANS, 1989) has concluded that an effective way of learning any skill, including quantitative skill, is "in context", placing learning objectives within the real environment rather than insisting that students learn in abstract what they will be expected to apply. We hypothesize that the real workplace environment is different for different specialties which are taught at universities, but this has so far not been studied in detail.

Given the limitations of extant knowledge of the actual usage of statistics in the workplace, this study reports on a survey of university graduates (primarily graduates of Masters and Doctoral programmes) which was designed with three goals in mind. First, we wanted to identify statistical techniques, research methods, and statistical software used in the workplace. Second, we wanted to identify deficiencies in the training received at university in these techniques, methods and software. Finally, a further goal of our study is to summarize similarities and contrasts in statistics training for several different specialties. Our expectation was that such a survey can contribute to the creation of an evidence basis which is needed for informing the design of statistical curricula. In addition, as teachers of statistics to undergraduates and as providers of consulting services in statistics to a wide range of researchers, we hoped that such a survey can also help us find ways to teach statistics in a more meaningful and motivating way.

The paper is organized in four parts. First we describe the target population and the methods employed in surveying our sample, which involved initially graduates in seven specialties subsequently reduced to five, statistics, economics and finance, marketing, the biological sciences and psychology, after deleting nutrition and food science. Next, results of the survey are reported and deficiencies with the teaching of statistics at university in the different subjects are summarized. Finally, the discussion focuses on recommendations for improved teaching in the light of our results and on suggestions for further research.

2. METHODS

2.1. SUBJECTS

The target population for our survey comprised students from the seven New Zealand universities who had completed PhD and Masters degrees in statistics, economics and finance, marketing, the biological sciences, psychology, nutrition and food science in the period 1995-2000. In the case of statistics there were few PhD or Masters graduates in New Zealand, hence it was decided to include those who had completed a Bachelor degree only as well. The period 1995 to 2000 was chosen to ensure most surveyed would have had the opportunity to be in employment while also able to recall what they had been taught at university; their relatively recent graduation should ensure accurate long term recall of topics that they had been taught. Brennan, Lyon, Schomburg and Teichler (1994) comment: "Around two years after graduation, most graduates will be settled into their first proper jobs, transitional shocks will be largely over, and graduates will be more dispassionate about their higher education but will still remember it." A number of those who had completed a Masters degree were currently enrolled for a PhD at a university in New Zealand or overseas, and these students were viewed as being in employment.

Because there were insufficient graduates in our chosen specialties at any one university, and there were doubts about address accuracy that could result in lower response rate, it was decided to approach all New Zealand universities in order to increase the number of potential responders and as a consequence the sample size. Six of the seven New Zealand Universities agreed to post out a questionnaire on our behalf and at our expense to their graduates in the nominated subjects, after we approached the alumni offices at each institution for permission to access the addresses of their graduates. Final permission was in each case granted by the Registrar or Vice Chancellor at each university. Participation by the major New Zealand universities guaranteed coverage of the target population. The seventh university which did not participate has only 3% of total student roll in New Zealand.

Ethical approval for the survey was obtained from the University of Otago Ethics Committee and other universities accepted this. A condition was that all information would be confidential and recorded by identification number only. Each university wrote a covering letter explaining the purpose of our survey and its importance for the training in statistics at each institution. This we felt would improve the response rate as many graduates are loyal to their own institution. But it did complicate the organisation associated with subsequent follow-up of non-respondents as we had to rely on the commitment of the various alumni offices to fit our project around their own work.

2.2. QUESTIONNAIRE CONSTRUCTION

An eight-page questionnaire was developed. General questions on academic background, attitudes to university teaching and the nature of employment were based on the Graduate Opinion and Employers' Survey of the University of Otago. We also designed questions on 46 statistical techniques and research methods taught in statistics programmes. These were compiled from techniques and research methods listed in the papers and reports discussed in the literature cited in the introduction, together with a report by Moore (1997) on studies of the use of statistics in several fields. We also consulted four statisticians and 15 other university teachers from departments that apply statistics or require statistics as part of their major requirements. These teachers were generally happy with the structure and clarity of our questionnaire. The questionnaire was then tested on a group of 20 graduates from the target population. Some minor changes were made as a consequence to remove ambiguities and the average response time of 15 minutes noted. This information was included in a covering letter sent out with the survey document in order to encourage response.

The questionnaire comprised six sections. Sections 1 and 2 collected details of university degrees, major subject and main field of employment. The crucial section 3 aimed to establish the frequency of use in the workplace of the 46 statistical techniques and research methods as well as six computer packages, and whether these techniques, methods and the use of each package had been taught at university. Section 4 investigated the types of statistical activity prevalent in the workplace. Section 5 surveyed course experiences in relation to the statistical methods which the respondents had been taught at university and invited respondents to nominate topics for short retraining courses and workshops that would support their current work. Section 6 asked for written comments, first on the relevance of the statistics training and education received at university for the subsequent employment of each respondent, and second on how statistics training at university could be improved to better prepare graduates for the workplace.

2.3. PROCEDURE

The questionnaire was administered via a mail survey in April-May 2002. All the recommendations made by Edwards, Roberts, Clarke, DiGuiseppi, Pratap, Wentz and Kwan (2002) for increasing response rates to postal questionnaires were used except for incentives. Either national or international post paid reply envelopes were sent with the questionnaire depending on whether the survey was posted to an address within New Zealand or overseas. There were 2758 questionnaires distributed to graduates in our target group and 721 completed questionnaires were returned over the next two months. Non-respondents were followed up by further mailings of the questionnaire between October 2002 and March 2003 after progressively compiling the addresses of the non-respondents from each University. This resulted in a further 256 completed questionnaires. We were not permitted to follow-up by phone, email addresses were not available and we did not have facilities for on-line response. A total of 977 responses were received and 353 questionnaires were returned undelivered or found to have been sent to graduates not in the target population. The effective response rate was therefore 39%.

A preliminary comparison of responses from the initial and follow-up phases showed no important differences in the calculated proportions, and for this reason all 977 completed questionnaires were used initially to produce results. All 977 were manually checked for incompleteness and inconsistency. All records entered in the database were checked and mistakes corrected. Data were entered in an excel spreadsheet but SPSS was used for analysis.

3. RESULTS

Of the 977 respondents in the survey, 93% were either employed or engaged in PhD study having already completed a Masters degree within our target time frame of 1995 to 2000, 5% were unemployed having been previously employed, and 2% only were unemployed. Opinions expressed therefore represent the views of employed graduates with higher degrees in subjects that use statistics.

There were 172 PhD respondents, 759 Masters or equivalent respondents and 46 Bachelor degree respondents who had completed their degree in statistics and who were added to the statistics PhDs and Masters respondents. Nutrition (34 respondents) and food science (30 respondents) were omitted from further analysis because these small numbers were insufficient for reliable conclusions. As a consequence only 913 respondents are included in the results reported for the five remaining specialties.

3.1. TYPE OF STATISTICAL ACTIVITY IN THE WORKPLACE

The dominant statistical activities carried out in the workplace by the respondents in the five specialties are listed in Table 1. Several of the activities are performed by many of the respondents, hence the proportions of the activities performed in each specialty can sum to more than one. The most common activity across all specialties involves the carrying out of data analyses. Next in importance is reading the results of published research. Only 11% of respondents report they have no need of statistics. Review of proportions which exceed one half in each speciality in Table 1 shows that biological sciences and psychology graduates design studies, carry out data analyses and have a need to read published work. Graduates in all other specialities also carry out data analyses. But, in addition, the statisticians are involved with report writing, the economics/finance graduates carry out, as expected, financial analyses and read published work while the marketing graduates, as expected, carry out market research and write reports.

Table 1. Numbers and proportions for statistical activities in the workplace

Activity	Subject Specialty				
	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psych (283)
Own data analyses	80 (0.67)	66 (0.78)	47 (0.57)	265 (0.77)	190 (0.67)
Reading published research	38 (0.32)	45 (0.53)	34 (0.41)	213 (0.62)	199 (0.70)
Report writing	49 (0.41)	36 (0.42)	45 (0.55)	160 (0.47)	118 (0.42)
Designing Studies	24 (0.20)	23 (0.27)	27 (0.33)	183 (0.53)	144 (0.51)
Financial analysis	38 (0.32)	52 (0.61)	27 (0.33)	28 (0.08)	28 (0.10)
Understanding a consultant	13 (0.11)	8 (0.09)	6 (0.07)	80 (0.23)	32 (0.11)
Market research	19 (0.16)	11 (0.13)	54 (0.66)	12 (0.03)	22 (0.08)
Quality control	13 (0.11)	5 (0.06)	7 (0.09)	30 (0.09)	14 (0.05)
None needed	18 (0.15)	4 (0.05)	5 (0.06)	42 (0.12)	29 (0.10)

3.2. STATISTICAL TECHNIQUES USED IN THE WORKPLACE

The 46 statistical techniques and research methods nominated in the questionnaire are listed in Table 2. For the five specialties in the study, numbers and proportions of graduates who use the techniques and methods are listed. To interpret pattern in the responses we define a technique with proportional use > 0.35 as frequently used, a technique with proportional use from 0.10 to 0.35 as moderately used and a technique with proportional use < 0.10 as seldom used.

Combining responses across the five specialties identifies frequent use of graphical procedures, basic tests, analysis of variance and simple linear regression; moderate use of contrasts, multiple regression, survey design, nonlinear and logistic regression, factorial and repeated measures designs, multivariate analysis of variance, principal components and statistics theory; some of the techniques seldom used are ordinal and nominal regression, survival analysis, cross over designs, path analysis, confirmatory factor analysis, correspondence analysis, Bayesian statistics, randomisation testing, state space and lag models, adaptive sampling and bioinformatics.

Table 2. Numbers and proportions using the techniques and methods in the workplace

Technique or method	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
	Subject Specialty				
1. Graphs	84 (.71)	70 (.82)	49 (.60)	247 (.72)	196 (.69)
2. Basic tests, t, χ^2	50 (.42)	43 (.51)	24 (.29)	201 (.58)	145 (.51)
3. ANOVA	34 (.29)	28 (.33)	13 (.16)	178 (.52)	115 (.41)
4. Contrasts	22 (.18)	10 (.12)	15 (.18)	108 (.31)	99 (.35)
5. Simple reg.	63 (.53)	55 (.65)	28 (.34)	182 (.53)	127 (.45)
6. Multiple reg.	44 (.37)	41 (.48)	12 (.15)	78 (.23)	71 (.25)
7. Nonlinear reg.	22 (.18)	19 (.22)	5 (.06)	64 (.19)	33 (.12)
8. Nonparametric reg.	7 (.06)	11 (.13)	3 (.04)	49 (.14)	29 (.10)
9. Mixed models	14 (.12)	10 (.12)	2 (.02)	43 (.13)	19 (.07)
10. Logistic reg.	30 (.25)	13 (.15)	6 (.07)	53 (.15)	30 (.11)
11. Ordinal/Nom reg.	8 (.07)	6 (.07)	3 (.04)	17 (.05)	10 (.04)
12. Loglinear models	17 (.14)	9 (.11)	3 (.04)	37 (.11)	8 (.03)
13. Survival analysis	10 (.08)	4 (.05)	1 (.01)	18 (.05)	8 (.03)
14. Factorial designs	11 (.09)	2 (.02)	7 (.09)	80 (.23)	58 (.20)
15. Blocking	10 (.08)	0 (.00)	3 (.04)	70 (.20)	23 (.08)
16. Repeated measures	10 (.08)	0 (.00)	3 (.04)	75 (.22)	56 (.20)
17. Crossover designs	7 (.06)	1 (.01)	1 (.01)	22 (.06)	23 (.08)
18. Clinical trials	7 (.06)	2 (.02)	0 (.00)	13 (.04)	42 (.15)
19. MANOVA	6 (.05)	4 (.05)	8 (.10)	74 (.22)	52 (.18)
20. Principal compts	16 (.13)	5 (.06)	10 (.12)	72 (.21)	31 (.11)
21. Factor analysis	9 (.08)	3 (.04)	15 (.18)	16 (.05)	33 (.12)
22. Path analysis	7 (.06)	3 (.04)	6 (.07)	13 (.04)	22 (.08)
23. Confirmatory F.A.	5 (.04)	1 (.01)	6 (.07)	7 (.02)	23 (.08)
24. Cluster analysis	15 (.13)	3 (.04)	14 (.17)	63 (.18)	16 (.06)
25. Correspondence A.	3 (.03)	0 (.00)	2 (.02)	32 (.09)	3 (.01)
26. Discriminant A.	10 (.08)	0 (.00)	5 (.06)	40 (.12)	15 (.05)
27. Scaling/Ordination	3 (.03)	0 (.00)	3 (.04)	42 (.12)	7 (.02)
28. Canonical Corr.A.	3 (.03)	0 (.00)	2 (.02)	35 (.10)	3 (.01)
29. Statistics theory	27 (.23)	24 (.28)	9 (.11)	55 (.16)	34 (.12)
30. Estimation theory	21 (.18)	17 (.20)	5 (.06)	39 (.11)	8 (.03)
31. Bayesian statistics	11 (.09)	2 (.02)	3 (.04)	16 (.05)	7 (.02)
32. Jackknifing	5 (.04)	1 (.01)	1 (.01)	19 (.06)	3 (.01)
33. Simulation	26 (.22)	12 (.14)	1 (.01)	42 (.12)	0 (.00)
34. Randomisation test	5 (.04)	3 (.04)	0 (.00)	31 (.09)	7 (.02)
35. ARMA	27 (.23)	23 (.25)	5 (.06)	7 (.02)	7 (.02)
36. Forecasting	25 (.21)	28 (.33)	19 (.23)	3 (.01)	7 (.02)
37. Markov chains	7 (.06)	6 (.07)	0 (.00)	11 (.03)	0 (.00)
38. State-space models	2 (.02)	5 (.06)	0 (.00)	3 (.01)	0 (.00)
39. Lag models	1 (.01)	8 (.09)	0 (.00)	3 (.01)	3 (.01)
40. Mark-recapture	6 (.03)	0 (.00)	0 (.00)	54 (.16)	0 (.00)
41. Survey design	27 (.23)	8 (.09)	28 (.34)	95 (.28)	64 (.23)
42. Adaptive sampling	2 (.02)	0 (.00)	3 (.04)	14 (.04)	3 (.01)
43. Power analysis	13 (.11)	2 (.02)	2 (.02)	62 (.18)	34 (.12)
44. Meta analysis	5 (.04)	2 (.02)	1 (.01)	14 (.04)	27 (.10)
45. Data mining	15 (.13)	6 (.07)	8 (.10)	13 (.04)	7 (.02)
46. Bioinformatics	2 (.02)	0 (.00)	0 (.00)	20 (.06)	3 (.01)

There are differences in the proportions across the subject specialties, however. A hierarchical average linkage clustering based on the proportions identifies three clusters. The biological sciences and psychology form the first cluster, statistics and economics/finance form the second cluster, while marketing has its own unique pattern of statistical activity in the third cluster.

In addition to the frequently used techniques, the biological sciences and psychology graduates together display moderate use of nonlinear, nonparametric and logistic regression, factorial and repeated measures designs, multivariate analysis of variance and principal components, survey design and power analysis. The biological sciences graduates further show moderate use of mixed models, loglinear models, cluster analysis, discrimination, ordination, canonical correlation analysis, estimation theory, mark recapture and simulation while psychology graduates show moderate use of clinical trials, factor analysis and meta analysis.

The statistics and economics/finance graduates together display moderate use of mixed models, nonlinear and logistic regression, statistics theory and estimation theory, auto regressive moving averages, forecasting and simulation, while statistics graduates in addition have moderate use for principal components, cluster analysis, data mining, survey design and power analysis.

Marketing graduates have moderate use for multivariate analysis of variance, principal components, factor analysis, cluster analysis, forecasting, survey design and data mining in addition to the frequently used basic procedures.

3.3. TECHNIQUES IN USE BUT NOT TAUGHT AT UNIVERSITY

Those surveyed were asked whether statistical techniques and research methods used in their employment had been included in their training at university. Responses summarized in Table 3 give, for each subject specialty, the numbers and proportions of graduates who use each technique or method but who have received *no* instruction in that technique or method. Inspection of the numbers greater than 46 (5% of 913 responders) when summing across the five specialties indicates training deficiencies overall in multiple, nonlinear, nonparametric and logistic regression, mixed models, multivariate analysis of variance, principal component analysis, cluster analysis, discriminant analysis, simulation, survey design and power analysis.

Table 3. Numbers and proportions not taught but using techniques or methods in the workplace

	Subject Specialty				
Technique or method	Statistics (119)	Econ/ Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Graphs	3 (.03)	2 (.02)	0 (.00)	19 (.06)	6 (.02)
2. Basic tests, t, χ^2	3 (.03)	0 (.00)	0 (.00)	13 (.04)	5 (.02)
3. ANOVA	2 (.02)	0 (.00)	0 (.00)	21 (.06)	6 (.02)
4. Contrasts	2 (.02)	0 (.00)	1 (.01)	19 (.06)	8 (.03)
5. Simple regression	2 (.02)	1 (.01)	0 (.00)	18 (.05)	11 (.04)
6. Multiple reg	4 (.03)	1 (.01)	0 (.00)	29 (.08)	21 (.07)
7. Nonlinear reg	8 (.07)	3 (.04)	3 (.04)	34 (.10)	13 (.05)
8. Nonparametric reg	1 (.01)	4 (.05)	3 (.04)	24 (.07)	16 (.06)
9. Mixed models	8 (.07)	3 (.04)	2 (.02)	22 (.06)	11 (.04)
10. Logistic reg	8 (.07)	2 (.02)	3 (.04)	28 (.08)	18 (.06)
11. Ordinal/Nom reg	7 (.06)	0 (.00)	2 (.02)	11 (.03)	6 (.02)

12. Loglinear models	5 (.04)	1 (.01)	1 (.01)	22 (.06)	6 (.02)
13. Survival analysis	4 (.03)	2 (.02)	0 (.00)	13 (.04)	7 (.02)
14. Factorial designs	0 (.00)	0 (.00)	2 (.02)	15 (.04)	4 (.01)
15. Blocking	0 (.00)	0 (.00)	2 (.02)	20 (.06)	1 (.00)
16. Repeated measures	2 (.02)	0 (.00)	2 (.02)	22 (.06)	6 (.02)
17. Crossover designs	3 (.03)	1 (.01)	0 (.00)	6 (.02)	3 (.01)
18. Clinical trials	5 (.04)	2 (.02)	0 (.00)	5 (.01)	8 (.03)
19. MANOVA	2 (.02)	1 (.01)	1 (.01)	41 (.12)	17 (.06)
20. Principal compts	4 (.03)	2 (.02)	1 (.01)	45 (.13)	18 (.06)
21. Factor analysis	4 (.03)	2 (.02)	1 (.01)	10 (.03)	17 (.06)
22. Path analysis	5 (.04)	2 (.02)	5 (.06)	10 (.03)	19 (.07)
23. Confirmatory F.A.	3 (.03)	1 (.01)	3 (.04)	5 (.01)	17 (.06)
24. Cluster analysis	3 (.03)	2 (.02)	2 (.02)	41 (.12)	11 (.04)
25. Correspondence A.	1 (.01)	0 (.00)	2 (.02)	22 (.06)	2 (.01)
26. Discriminant A.	3 (.03)	0 (.00)	3 (.04)	30 (.09)	10 (.04)
27. Scaling/Ordination	1 (.01)	0 (.00)	2 (.02)	29 (.08)	3 (.01)
28. Canonical Corr A.	2 (.02)	0 (.00)	2 (.02)	29 (.08)	2 (.01)
29. Statistics theory	0 (.00)	1 (.01)	1 (.01)	16 (.05)	4 (.01)
30. Estimation theory	3 (.03)	1 (.01)	2 (.02)	27 (.08)	3 (.01)
31. Bayesian statistics	3 (.03)	0 (.00)	2 (.02)	14 (.04)	4 (.01)
32. Jackknifing	2 (.02)	0 (.00)	1 (.01)	13 (.04)	2 (.01)
33. Simulation	7 (.06)	8 (.09)	1 (.01)	36 (.10)	0 (.00)
34. Randomisation test	2 (.02)	1 (.01)	1 (.01)	21 (.06)	3 (.01)
35. ARMA	1 (.01)	1 (.01)	0 (.00)	1 (.00)	3 (.01)
36. Forecasting	2 (.02)	5 (.06)	6 (.07)	4 (.01)	2 (.01)
37. Markov chains	0 (.00)	2 (.02)	0 (.00)	9 (.03)	1 (.00)
38. State-space models	0 (.00)	2 (.02)	0 (.00)	4 (.01)	1 (.00)
39. Lag models	0 (.00)	1 (.01)	0 (.00)	2 (.01)	3 (.01)
40. Mark-recapture	2 (.02)	0 (.00)	0 (.00)	29 (.08)	0 (.00)
41. Survey design	4 (.03)	7 (.08)	3 (.04)	21 (.04)	11 (.04)
42. Adaptive sampling	1 (.01)	0 (.00)	2 (.02)	6 (.02)	1 (.00)
43. Power analysis	5 (.04)	2 (.02)	2 (.02)	30 (.09)	17 (.06)
44. Meta analysis	5 (.04)	2 (.02)	1 (.01)	9 (.03)	22 (.08)
45. Data mining	14 (.12)	4 (.05)	7 (.09)	12 (.03)	5 (.02)
46. Bioinformatics	2 (.02)	0 (.00)	0 (.00)	17 (.05)	3 (.01)

Deficiencies within each specialty are also apparent. Proportions greater than 0.10 (reflecting the smaller numbers in each specialty) indicate some lack of training in the biological sciences for nonlinear regression, multivariate analysis of variance, principal component analysis, cluster analysis and simulation. There is a lack of training among statistics graduates in data mining. In addition, 0.07 to 0.09 of the graduates in the biological sciences have received no training in logistic and multiple regression, the multivariate techniques of ordination, discrimination and canonical correlations, power analysis and mark-recapture. At an equivalent level there are gaps in training in multiple regression, path analysis and meta analysis for psychology graduates; mixed models, nonlinear and logistic regression for statistics graduates; survey design and simulation for economics/finance graduates; forecasting and data mining for marketing graduates.

3.4. COMPUTING IN THE WORKPLACE

Much statistical activity involves use of computer software for both data management and statistical analyses. Table 4 lists Excel, Access, SPSS, SAS, Minitab and S+/R as packages used by over 5% of the sample. Several other packages are used to a lesser extent as follows: Statistica (34 users); Shazam (33); Systat (29); Matlab (22); Datadesk (16); Statview (16); Stata (15); Sigmastat (13); Genstat (10).

The results in Table 4 show widespread use of Excel. Access is frequently used overall but to a lesser degree by graduates in the biological sciences and psychology. SPSS is frequently (0.37) used by psychology graduates and moderately used by marketing and biological sciences graduates. SAS on the other hand is mainly used by statistics and economics/finance graduates. Minitab is seldom used in the workplace except by biological sciences graduates where usage is moderate (0.12). S+ and R are seldom used except by statistics graduates. The full implications of the free to use package, R, are only now being realised, and this is likely to result in greater use of R in the next few years.

Table 4. Numbers and proportions using statistical packages in workplace

	Subject Speciality				
Package	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Excel	91 (0.76)	72 (0.85)	62 (0.76)	279 (0.81)	178 (0.63)
2. Access	49 (0.41)	33 (0.39)	43 (0.52)	111 (0.32)	66 (0.23)
3. SPSS	10 (0.08)	5 (0.06)	14 (0.17)	63 (0.18)	106 (0.37)
4. SAS	41 (0.34)	18 (0.21)	2 (0.02)	51 (0.15)	25 (0.09)
5. Minitab	10 (0.08)	4 (0.05)	0 (0.00)	42 (0.12)	2 (0.01)
6. S+/R	22 (0.18)	4 (0.05)	2 (0.02)	23 (0.07)	2 (0.01)

Table 5 reports the numbers and proportions of graduates using each package with no prior instruction, and points to gaps in statistical package training that have implications for needed changes in training in order to cover these gaps. Overall, 45% of those using Excel and 30% of those using Access have had no prior instruction. There are deficiencies in the teaching of SPSS to psychology and biological sciences graduates and deficiencies in the teaching of SAS to statistics and economics/finance graduates.

Table 5. Numbers and proportions not taught but using packages in workplace

	Subject Speciality				
Package	Statistics (119)	Econ/Finance (85)	Marketing (82)	Bio Science (344)	Psychology (283)
1. Excel	50 (0.42)	40 (0.47)	22 (0.27)	161 (0.47)	133 (0.47)
2. Access	43 (0.36)	29 (0.34)	33 (0.40)	105 (0.31)	63 (0.22)
3. SPSS	8 (0.07)	4 (0.05)	1 (0.01)	57 (0.17)	56 (0.20)
4. SAS	15 (0.13)	16 (0.19)	0 (0.00)	32 (0.09)	16 (0.06)
5. Minitab	1 (0.01)	4 (0.05)	0 (0.00)	15 (0.04)	2 (0.01)
6. S+/R	4 (0.03)	4 (0.05)	1 (0.01)	16 (0.05)	2 (0.01)

3.5. RETRAINING

Respondents were invited to nominate courses or intensive workshops they thought would help reinforce the statistics techniques and research methods needed for their employment. Some respondents expressed an interest in attending retraining courses but did not list topics. But for the 396 respondents who listed courses, Table 6 summarises the preferred workshop requests in order of most mentions. Some respondents made multiple requests. The three most requested topics are multivariate methods, generalized linear models and survey design including power analysis.

Table 6. Preferred workshops with numbers and proportions of requests

Workshop Topic	Number of respondent requests	Proportion of respondent requests
Multivariate methods	99	0.25
Regression/generalized linear models	88	0.22
Survey design and power analysis	84	0.21
Statistical software developments	75	0.19
Introductory statistical methods	75	0.19
Stochastic processes	39	0.10
Psychology topics/clinical trials	35	0.09
Forecasting/Time Series	32	0.08
Theory including Bayesian methods	20	0.05
Experimental designs	19	0.05
Marketing related topics	16	0.04
New methods e.g. meta analysis, data mining, bioinformatics	12	0.03
Computer intensive statistics	8	0.02

Table 7 summarizes the three most cited topics for retraining, by specialty. The table also records the number of times each course was nominated in that subject area. Many of the respondents noted that these topics could be taught alternatively in the undergraduate programmes at university as part of the requirements for majors.

Table 7. Preferred courses requested by specialty

Statistics (119)	Regression/Generalized linear models (14 nominations) Forecasting/Time Series (12) Statistical software developments (8)
Econ/Finance (85)	Forecasting/Time Series (20) Regression/Generalized linear models (13) Multivariate methods (7)
Marketing (82)	Marketing related topics (10) Regression/Generalized linear models (7) Multivariate methods (6)
Bio Science (344)	Survey design and power analysis (55) Multivariate methods (48) Regression/Generalized linear models (31)
Psychology (283)	Multivariate methods (31) Statistical software developments (28) Psychology topics/clinical trials (27)

The course nominations match moderately or frequently used techniques as listed in Table 2 or the training deficiencies noted from Table 3. For example, for the graduates in the biological sciences, training deficiencies were identified initially in aspects of regression, multivariate methods, data mining, power analysis and mark recapture; these topics bear a close resemblance to the three most frequently nominated requests, survey design and power analysis (55), multivariate methods (48) and regression/generalized linear models (31).

3.6. COURSE APPRECIATION

Respondents were invited to provide free-form comments on the relevance of their statistical training for their employment and to make recommendations for improving university teaching in statistics to better prepare them for their workplaces. These responses have not been analysed quantitatively, instead being used to provide anecdotal examples to help gain deeper appreciation for the linkage (or lack thereof) between studying statistics and research methods at university and the actual demands of the workplace with consequential training gaps.

One theme raised in the written comments about the work environment cited the importance of statistics which had in many cases not been realised while students were studying at university. One graduate wrote about statistics:

> *I would be impotent without it. Keep statistics compulsory – I never would have done it otherwise – and it is now the most incisive tool in my arsenal.*

A zoology and ecology MSc graduate made the following comment along similar lines:

> *I am not currently employed in the field for which I studied but hope to be eventually. There is little requirement for stats understanding in my local government position but I know my knowledge of statistics will be insufficient if/when I begin an ecological career, which was the focus of my study. I did not understand the importance of stats in ecology when I began my BSc so only took one first year course, as that was all that was compulsory. Biological Science students should be forced to do at least second year stats courses. When I came to do honours research in my 4^{th} year I found my statistical understanding to be poor and relied heavily on outside help and did not really understand the computations behind the statistical software I was using. I introduced myself to SPSS software and quoted results of the analyses in a paper that was eventually published but would have benefited greatly from more guidance on the results' interpretation.*

It was clear from the comments that only after finishing their study do many students come to realise that they need more statistics. A further common view expressed was that the disciplines, especially those in science, should promote statistics more. Many respondents thought that more statistics at each level of university study would have been beneficial. By the time they begin their research a lot of what they have learnt has been forgotten. A few said they had found it hard to fit extra statistics courses into their major study and therefore statistics courses should have been more a part of their major requirements. One graduate commented:

> *Stats teaching needs to be more integrated with the discipline in which it is being applied so that students think of it as another important tool within that discipline rather than perceiving it as scary numerical stuff outside the discipline they're studying and thus able to be avoided.*

Another advocated teaching in context by suggesting that:

> *Courses targeting analyses commonly used in a specific discipline (e.g. psychology, marketing) would increase student interest.*

Respondents also suggested that there should be a more co-ordinated approach between departments with greater collaboration between major research disciplines and departments of mathematics and statistics in order to provide courses that suit. Many psychology students in particular wanted statistics subjects to be relevant to their work. Some thought that their statistics training was too general and recommended that statistics should be taught within their discipline so that topics would be applicable. One graduate maintained that courses on statistics taught within psychology departments were "*excellent*". Yet on the other hand, another two graduates who obviously had had different experiences supported the view that statistics departments should teach statistics by stating:

> *Psychology and education departments should be forced to have the statistics department teach their students – as psychology and education departments do a bad job.*

> *Statisticians teaching stats, not biologists teaching stats.*

These diverging views need further research to ascertain the best location for the teaching of statistics and to identify how to incorporate statistics in each of the disciplines which use statistics.

4. DISCUSSION

This study was initiated because we sought ways to improve training in statistics to better prepare graduates for employment. We surveyed graduates, currently employed, who had completed Masters and PhD degrees mainly in a range of specialties that have statistics pre-requisites at universities. Holmes (1998) only surveyed employers and universities and concentrated on majors in statistics. We have placed the emphasis on statistics teaching in service courses.

Our survey has identified statistical activity in the workplace and shortcomings in the statistical training received for this work. The respondents in each specialty surveyed have nominated key courses for retraining or, equivalently, key statistical techniques and research methods that should be included in coursework at university.

The SCANS report (1989) identified skills that those who have completed either high school or university should master if they are to be effective in the workplace. Table 1 summarises the main statistical activities in the workplace and this shows that carrying out data analyses, reading published research and report writing are important for 71%, 58% and 45% respectively of those surveyed. These skills relate to understanding mathematics, reading and writing which are emphasised in the SCANS report.

When they surveyed psychology departments, Aitken, West, Sechrest and Reno (1990) discovered that PhD graduates were able to handle traditional statistics techniques but were deficient in newer procedures. We show in Table 2 that the elementary techniques and topics are used frequently in the workplace and are well covered in the training of psychology students according to the numbers in Table 3. But our study has, in addition, identified three more advanced retraining workshops important for psychology graduates in the workplace. This is in agreement with the opinions of Howard, Pion, Gottfredson, Flattau, Oskamp, Pfafflin, Bray and Burstein (1986) who claim that many psychology graduates obtain employment but are poorly prepared for some aspects of their work.

Higgins (1999) argues for a concentration on the service aspects of statistics teaching. He suggests that planning and management of scientific studies and communication skills are important for graduates. This is consistent with the evidence in Table 1 that designing studies, report writing and reading published work are statistical activities occurring frequently in the workplace. What emerges is a need for an approach where service courses on the above areas and on statistical methodology are taught in a way that reduces dependence on mathematics and probability. The focused workshops nominated in our survey are consistent with the spirit of this approach.

Both Cobb (in Higgins, 1999) who describes a liberal arts major in statistics at Mt Holyoke College, and Moore (2001), also support this approach. Moore remarks:

> While the discipline of statistics is healthy, its place in academe is not. Our future there depends strongly on achieving a more prominent place in undergraduate education beyond the first methods course. To this end we must offer undergraduate programs that are popular with students … the primary intent of such programs cannot be to prepare students for graduate study in statistics, but to equip them for employment with a bachelor's degree or for further study in a wide variety of areas. Finally, success requires greater co-operation between statistics and other disciplines, ….

4.1. STUDY LIMITATIONS

This survey, involving seven subject specialties, was ambitious yet not wholly successful. The overall response rate was 39% which is reasonable for a mail survey but not as good as one would hope. There were small response numbers in some of the categories in our target population. This resulted in omitting both nutrition and food science with 34 and 30 respondents respectively from our analyses.

The sample sizes in the biological sciences and psychology on the other hand were satisfactory adding strength to conclusions reached for these subjects. The sample sizes in the business areas were not as good but we felt it was reasonable to keep these subjects in the analyses.

There were few PhD and Masters graduates in statistics at the six universities surveyed. For this reason we also mailed our questionnaire to graduates who had completed Bachelor level degrees only in statistics. At one university 38 students who had completed a Bachelor degree had also completed a second Bachelor degree in economics/finance. For our analyses these graduates were not viewed as economics/finance graduates because they only had Bachelor degrees in these subjects. One consequence of this policy on classification of respondents is that the techniques used by the statistics graduates in the workplace will be biased towards the statistics techniques used by Masters or PhD graduates in economics/finance. Care is therefore needed when generalizing the results for statistics graduates. It is interesting that the cluster analysis reported in Section 3.2 showed similarity between statistics and economics/finance.

There is a question about the generalizing of our results to other countries. Even though 100 of our respondents (about 10%) were found to be working overseas at the time of our survey, the academic education system, including in statistics service courses, is not the same in all countries and therefore conclusions should be generalized with caution to countries whose academic education system differs from that in New Zealand. It would be interesting to carry out a similar survey in other countries.

It is also possible that our survey time frame 1995 to 2000 has missed recent developments in computing like the expanding popularity of the free-to-use statistical package R. Similarly, statistical procedures and research methods used in recent advances in the area of genetics will have been missed in our investigation.

4.2. RECOMMENDATIONS

About 89% (815 respondents) of the graduates surveyed in this study used some aspect of statistics in their employment. To overcome the deficiencies in statistical training identified for almost half of them, we make five recommendations. These relate to the content of advanced university courses at either the undergraduate or the postgraduate level, instructional style, and the content of retraining statistics modules for lifelong learning while in employment.

Recommendation 1: expand statistical methods taught in the biological sciences to include advanced topics in regression and generalized linear modelling, multivariate methods, power analysis, mark-recapture and data mining; expand the statistical content of psychology to include regression and generalized linear modelling, survey design, multivariate methods especially factor analysis, the design of clinical trials and meta analysis; expand the statistical content of economics, finance and marketing to include regression modelling and multivariate methods. These advanced topics should follow from introductory first year statistical methods courses but should be taught without mathematics and probability prerequisites in order to make them accessible to the graduates in the specialties surveyed.

Recommendation 2: university statistics departments should develop short courses or intensive workshops for postgraduate students. Such courses will also serve as retraining modules for those in employment. Recently we had success with a 3-day workshop on multivariate statistics techniques for 25 ecologists enrolled for PhD or Masters degrees. Data sets relevant to the work of the participants were used, some of these being generated by those attending. The workshop was therefore taught in the context of ecology. It was successful and requests have been received to mount a second workshop on regression and generalized linear modelling as well as repeating the first.

Recommendation 3: seek support from staff in departments teaching the specialties we have surveyed. They should be asked to encourage their senior undergraduate students to enrol for more undergraduate statistics courses. The staff in these departments could be used as guest or visiting lecturers. They could also provide interesting data sets which would place the statistics in the context of the specialty. It could even be possible to have statisticians located in the other departments. A consequence of these approaches could be the inclusion of critical appraisal of some recently published research articles and the development of group project work using unstructured real data generated from consulting or work placement. These instructional styles should equip students with the essential skills for the workplace identified by the SCANS report (1989).

Recommendation 4: university statistics or mathematics departments should investigate the development of a data specialist undergraduate major along the lines of Higgins (1999). This could be studied in conjunction with a major in another specialty and employment prospects for graduates with this type of qualification would be enhanced.

Recommendation 5: check carefully the statistical software that is used when teaching students in the different specialties surveyed in order to meet the deficiencies noted in Section 3.4.

These five recommendations go a long way towards meeting the views expressed by many respondents who advocated that statistics must be part of their major requirement. This study provides compelling evidence in support of teaching more statistics while paying close attention to workplace needs. The support is coming not from statisticians but from recent PhD and Masters graduates, now in the workplace, who have majored in disciplines other than statistics. Conveying these attitudes to students studying at university can only encourage them to include as much statistics in their courses as possible.

4.3. FUTURE RESEARCH

The limitations caused by low response rates in nutrition and food science could be addressed by raising response rates in these areas. The use of on-line replies and phone interviews could increase sample size. There would also be benefit in focusing on statistics graduates in a further survey. The list of techniques and research methods used for the specialties in our survey could be adapted for a statistics graduate survey by including more specialised statistical procedures.

Another survey could assess the relative importance of techniques and the difficulties students encounter when they have to cope with statistics in the workplace.

Reasons why students involved in research at universities allow gaps to develop in their statistics knowledge should be investigated given the overwhelming need for extended training which we have identified. We propose therefore a survey of research students at university together with an analysis of study design, research methods and statistical procedures, if any, used in their theses. The thesis results could be cross referenced to the student opinion. We are currently attempting this exercise ourselves. Use of on-line response is giving fuller information than that obtained on written response forms. We are attempting to establish at what stage statistics help was sought, for instance at the early planning stage of the study or when the data were to be analysed.

ACKNOWLEDGEMENTS

This work was supported by a Research into University Teaching Grant from the University of Otago and a grant from the Campbell bequest to the New Zealand Statistical Association. We are grateful for this support. We also thank the four anonymous referees whose extensive comments have significantly improved this paper.

REFERENCES

Aiken, L.S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate Training in Statistics, Methodology and Measurement in Psychology. *American Psychologist, 45*, 721-734.

Brennan, J., Lyon, E. S., Schomburg, H., & Teichler, U. (1994). The experiences and views of graduates messages from recent surveys. *Higher Education Management*, *6*, 275-304.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*, *324*, 1183-1185.

Harraway, J. A., Manly, B. F. J., Sutherland, H., & McRae, A. (2001). Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in The Use of Statistics* (pp. 177-195). Voorburg, The

Netherlands: International Association for Statistical Education and International Statistical Institute.

Higgins, J. J. (1999). Nonmathematical statistics: a new direction for the undergraduate discipline. *The American Statistician*, *53*(1), 1-6.

Holmes, P. (1998) Matching education, assessment and employment needs in statistics. *Report on MEANS project.*
[Online: http://science.ntu.ac.uk/rsscse/means/means.html]

Howard, A., Pion, G. M., Gottfredson, G. D., Flattau, P. E., Oskamp, S., Pfafflin, S. M., Bray, D. W., & Burstein, A.G. (1986). The changing face of American Psychology: A report from the committee on employment and human resources. *American Psychologist*, *41*, 1311-1327.

Jolliffe, F., & Rangecroft, M. (1997). Statistical software use by student and consultant statisticians in the UK-a comparative study.
[Online: http://www.stats.gla.ac.uk/cti/activities/reviews/97_08/software_su]

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*, 254-261.

Moore, D. S. (2001). Undergraduate programs and the future of academic statistics. *The American Statistician*, *55*(1), 1-6.

Secretary's Commission on Achieving Necessary Skills (SCANS) (1989). US Department of Labor, Employment and Training Administration [Online: http://www.tvi.cc.nm.us/html/scans.html]

J. A. HARRAWAY
Department of Mathematics and Statistics
University of Otago
PO Box 56, Dunedin
New Zealand

LOS OBSTÁCULOS EN EL APRENDIZAJE DEL CONOCIMIENTO PROBABILÍSTICO: SU INCIDENCIA DESDE LOS LIBROS DE TEXTO

OBSTACLES IN THE LEARNING OF PROBABILISTIC KNOWLEDGE: INFLUENCE FROM THE TEXTBOOKS

ANA SERRADÓ
Profesor E.Secundaria
ana.serrado@uca.es

JOSÉ Mª CARDEÑOSO
Universidad de Granada
josem@ugr.es

PILAR AZCÁRATE
Universidad de Cádiz
pilar.azcarate@uca.es

RESUMEN

En este trabajo reinterpretamos en términos de obstáculos (epistemológicos, cognitivos y didácticos) algunos de los sesgos, heurísticos, falacias y paradojas que surgen al asignar probabilidades a fenómenos aleatorios y que han sido descritos por las investigaciones en el campo. En segundo lugar, se presenta un análisis de las unidades dedicadas al "Tratamiento del Azar" en una muestra de libros de texto españoles de Educación Secundaria Obligatoria (de 12 a 16 años) con la finalidad de mostrar algunos posibles obstáculos que puede inducir la presentación de este conocimiento en dichos textos. Todo ello con el objetivo de aportar criterios en la elaboración de los libros de texto que tengan en cuenta las investigaciones sobre didáctica de la probabilidad. Nota: Un resumen extenso en Inglés se incluye al comienzo de este artículo, que está escrito en Español.

Palabras clave*: Investigación en educación estadística; Educación secundaria Obligatoria; Obstáculos probabilísticos; Libros de texto*

ABSTRACT

In this work we first reinterpret, using the idea of (epistemological, cognitive and didactical) obstacle some biases, heuristics, fallacies and paradoxes that arise in assigning probabilities to random phenomena, and have been described in previous research. We then reflect on the way obstacles should be taken into account in the process of teaching and learning probability. Thirdly, we analyse some didactic units related to "Dealing with chance" in a sample of Compulsory Secondary Education Spanish's textbooks (12 to 16 year-old students) to show some possible obstacles that might be induced by the presentation of this *knowledge in the books. The final aim is providing some criteria to elaborate new textbooks that take into account research on probability education. Note: An extended summary in English is provided at the beginning of this paper, which is written in Spanish.*

Keywords: *Statistics education research; Compulsory Secondary Education; probability obstacles; textbooks*

EXTENDED SUMMARY

In this paper we present and analyse some possible epistemological, ontogenical and didactical obstacles, which might be introduced to children from the way some textbooks develop the probabilistic content.

In the first section of the paper, starting from Brousseau's (1983) description of obstacles, we reinterpret some of the difficulties found in the historical development of probability or that have been described in previous research on probabilistic reasoning in terms of three types of obstacles:

1. Epistemological obstacles, which are usually identified from the historical analysis, since they coincide with difficulties that arose in the development of the subject. In our work we are focussing on some epistemological obstacles related to the ideas of chance, randomness and probability.
2. Ontogenetic obstacles, which are related to children's cognitive development. We have reinterpreted, in terms of obstacles, some well known research on the understanding of the following basic probabilistic notions: simple and compound events, equiprobable outcomes and independent events. From our point of view ontogenetic obstacles are also reflected in the reasoning strategies that children use in uncertain situations, such as the "outcome approach", the representativeness heuristic; the gambler's fallacy and the equiprobability bias.
3. Didactic obstacles, which are related to the way a topic is taught. In our research we focus on two types of didactic obstacles: inappropriate use of probabilistic language, and the contexts used for exemplification and experimentation.

In the second section of the paper we describe the research design. We carried out a content analysis of the lessons dealing with Chance and Probability in a sample of textbooks aimed at pupils in Spanish Compulsory Secondary Education (12 to 16 year-old students). This sample included the full set of textbooks (through the different educational levels from 12 to 16 years) of four Spanish series of wide diffusion (20 books in total). For each of the textbooks we analysed the definitions, explanations, examples and activities included for two main topics: a) chance and randomness and b) probability.

Results are presented in the third section of the paper, where we characterise some possible epistemological, ontogenical and didactical obstacles, giving examples of definitions and examples presented in the textbooks that we analysed in our research and that support our analysis. These results are summarised in Tables 1 and 2. We identify the four series with the letters B, S, G, M in these tables.

The analysis and comparison of these textbooks led us to conclude that chance is modelled in these textbooks basically as synonymous with luck and randomness, and it is related with the uncertainty of the event only. These characterizations are insufficient to completely understand the meaning of probability. The opposition to determinism emphasizes the cause and effect relationships that, in turn, might produce biases in the correct interpretation of dependent and independent events.

Table 1. Treatment of chance and randomness and possible obstacles induced

Concept	Presentation in the textbooks	Obstacles that might be induced by the way the topic is presented	Type of obstacle
Chance	Chance is presented as synonymous of unknown cause (B, S, G, M). There is only one characterisation.	Restricting the type of phenomena that can be studied in probability	Epistemological
Random Experiment	Presenting random experiment as anything unpredictable (B, S, G, M).	Restricting the wide meaning of random experiment	Epistemological
Deterministic Experiment	Presenting determinist experiment as anything predictable (B, S, G, M).	Inappropriate understanding of uncertainty and cause-effect	Epistemological
	Emphasis in cause-effect relationship (B, S, G, M).	No distinction between chance and causality	Ontogenic
Process / Event	No distinction between random process and outcome (B, S, G, M).	No distinction between compound experiment and repeated trials of the same experiment Non distinction between simple/compound event and experiment	Ontogenical
Sample space / outcome	Introducing sample space from the set notion, which is not clarified (B, S, G, M-3º-4º).	Misunderstanding of reasons for some assignments of probability	Ontogenical
	Presenting only activities and examples associated with random generators in the context of chance games (S, G).	Blocking transference to other contexts	Didactical
Random sequence	Lack of experience with random sequences, (B, G).	Favouring the use of heuristics	Didactical

Secondly, the examples and activities included in the textbooks are most of the times associated with random generators for finite sample spaces and equally likely outcomes. These examples facilitate the determination of the sample space, but they can block later generalization to sample spaces in daily life contexts and also hide the multiple applications of probability calculus. Moreover, in spite of the fact that these examples and activities provide a good opportunity to discuss the equiprobability of events, this is not requested in the textbooks. This lack of discussion favours the identification of equiprobability with equi-ignorance and may constitute an obstacle in favouring equiprobability bias.

Thirdly, these textbooks do not clarify the meaning of terms such as unforeseeable, set, certain, impossible, convergence, etc., to which children might assign inappropriate meaning and thus block the construction of some probabilistic notions which are described by the said terms, such as: random experiment, event and process, sample space, random sequence, stability of relative frequencies.

Table 2. Treatment of probability and possible obstacles induced

Concept	Presentation in the textbooks	Obstacles that might be induced by the way the topic is presented	Type of obstacle
Probability (Frequential view)	No reflection on the meaning of stochastic convergence in analysing the stability of the relative frequencies (B, S, G, M).	Expecting convergence in small samples	Ontogenical
	Introduction, stability of relative frequencies only at theoretical level, (B, S).	Representativeness heuristics	Didactical
Comparison of Probabilities	Quantifying and assigning probabilities only from comparison of specific random generator (roulettes) (B, S).	Associating the generator to the strategy to assign probabilities	Didactical
Equiprobability	Circularity in defining equiprobality (Bº, G) Taking equiprobability for granted (B).	Equiprobability bias	Epistemological
Dependence and independence	Activities and examples associated to sampling with /without replacement, (B, S, G, M).	Non transfer to other contexts	Didactical

Finally, the introduction of the frequential notion of probability in the textbooks is based on the idea of stochastic convergence, which is not discussed at these educational levels; however, an intuitive understanding is taken for granted. These textbooks include neither examples nor activities that could enable the students to become aware of their misconceptions about the occurrence of chance in random series, which appear in form of heuristics (the gambler's fallacy, representativeness of the sample misconception, "outcome approach", etc.).

LOS OBSTÁCULOS EN EL APRENDIZAJE DEL CONOCIMIENTO PROBABILÍSTICO: SU INCIDENCIA DESDE LOS LIBROS DE TEXTO

1. INTRODUCCIÓN

El análisis de los libros de texto es una forma de aproximarse al estudio de los difíciles y complejos procesos de enseñanza y aprendizaje del conocimiento probabilístico, como muestran los trabajos de Malara (1989), Ortiz (1999), Ortiz y Serrano (2001), Serradó (2003), Serradó y Azcárate (2003). Trabajos que nos aportan antecedentes sobre la descripción de los conceptos probabilísticos propuestos en los libros de texto y las actividades que facilitan el proceso de enseñanza y aprendizaje. En este artículo completamos dichas investigaciones con relación a los posibles obstáculos (en el sentido de Brousseau), a los que se puede enfrentar el alumno durante este proceso de enseñanza y aprendizaje.

En la primera sección del artículo, desde la descripción de la noción de obstáculo (Brousseau, 1983), reinterpretamos en términos de obstáculo epistemológico, ontogénico o didáctico, algunas de las dificultades descritas en los estudios históricos o investigaciones sobre razonamiento probabilístico, en relación con el aprendizaje de las diferentes nociones que se desarrollan en los proyectos curriculares españoles. A continuación, tras una breve descripción de las características del diseño metodológico utilizado, presentamos un análisis de las unidades dedicadas al "Tratamiento del Azar" en una muestra de libros de texto españoles de Educación Secundaria Obligatoria (de 12 a 16 años) con la finalidad de mostrar algunos posibles obstáculos que puede inducir la presentación de este conocimiento en dichos textos. Finalizamos con unas reflexiones en torno al contraste entre los obstáculos en la construcción del conocimiento probabilístico y las propuestas de los libros de texto.

2. OBSTÁCULOS EN LA CONSTRUCCIÓN DEL CONOCIMIENTO PROBABILÍSTICO

La noción de obstáculo utilizada en este artículo se basa en las consideraciones aportadas por Brousseau (1983), para quien esta noción se relaciona con aquel conocimiento que ha sido, en general, satisfactorio durante un tiempo para la resolución de ciertos problemas, y que por esta razón se fija en la mente de los estudiantes, pero que posteriormente resulta inadecuado, y difícil de adaptarse, cuando el alumno se enfrenta a problemas nuevos.

Un conocimiento puede ser considerado como obstáculo cuando es elaborado como fruto de la interacción del alumno con su medio y, en esa situación, produce un resultado "interesante" y "útil", aunque no válido ni adecuado. Brousseau (1983) considera que los obstáculos que se presentan en el sistema didáctico pueden tener diferentes orígenes: epistemológico, didáctico u ontogénico.

Los obstáculos *epistemológicos* se identifican a partir de las investigaciones sobre Historia de la Ciencia y los procesos de construcción de los conocimientos, los *ontogénicos* a partir de la revisión de las investigaciones sobre el aprendizaje de estas nociones y los *didácticos*, se extraen de las investigaciones sobre el "Tratamiento del Azar" en el aula. Aunque Brousseau describe estas tres categorías separadas, a veces es difícil determinar si un cierto obstáculo de tipo epistemológico no tiene también un carácter ontogénico o bien los obstáculos epistemológicos y ontogénicos se reproducen siempre en un sistema didáctico y, por tanto, también se pueden configurar como obstáculos de carácter didáctico.

A continuación tratamos de aplicar estas ideas describiendo algunos posibles obstáculos asociados a las nociones probabilísticas que se recogen en el currículum escolar español para alumnos de 12 a 16 años (Decreto 106/1992) y que, a su vez, se incluyen en los libros de texto de la muestra.

2.1. OBSTÁCULOS EPISTEMOLÓGICOS

El análisis de los estudios de evolución histórica del conocimiento probabilístico llevados a cabos por diferentes autores nos ha permitido identificar los obstáculos que se han ido superando a la hora de caracterizar las nociones de azar, aleatoriedad y probabilidad, que resumimos a continuación.

La noción de azar Son muchos los análisis filosóficos y didácticos de la noción de azar y su relación con la aleatoriedad (por ejemplo, Azcárate, 1995 o Bennett, 1993; 2000). La clarificación del significado de la noción de azar ha pasado por diferentes etapas significativas, en las que se buscaban diferentes explicaciones para los fenómenos indeterminados (Discursos del azar, según Azcárate, 1995).

Desde el *Discurso del orden,* característico de las primeras civilizaciones, el azar es entendido como causa desconocida, es el que ocasiona sucesos inesperados o "extraños", y el que se asocia con el desorden inicial (Caos) que a veces surge a través de fuerzas incontroladas de origen mágico o divino.

Ya en la época grecorromana, impera el *Discurso del Azar/Necesidad,* según el cual el azar, que sigue siendo algo desconocido, es explicado como un simple reflejo del cruce inesperado de un conjunto de hechos que son producto de series causales independientes.

La existencia de la incertidumbre en otros momentos de la historia se explicó también desde el poder de la *Providencia*, que garantiza el orden y la armonía del Universo. El "azar" surge como reflejo de la voluntad de la Divina Providencia, como un hecho inexplicable para el ser humano, idea que se mantiene hasta nuestros días.

La progresiva separación de las explicaciones religiosas y científicas de los fenómenos hace concluir la aparición de un nuevo discurso del Azar, el *Discurso de la Ignorancia*. El "azar" era producto de la ignorancia humana a la hora de analizar científicamente ciertos acontecimientos de la Naturaleza través de leyes causales y deterministas. Según esta concepción no hay "azar" realmente, no existe el azar en sí mismo, es nuestra ignorancia lo que nos hace recurrir a él.

Poincaré (1979), en una búsqueda de una noción de "azar" que significase algo más que la ignorancia humana, describió tres tipos de sucesos cuyo comportamiento era atribuido al azar. En primer lugar, sucesos que pueden estar producidos por causas insignificantes, que se nos escapan perceptivamente, pero que determinan un efecto considerable. En segundo lugar, sucesos en que lo importante no es la pequeñez de las causas que lo provocan, sino la complejidad de todas las interacciones entre ellas. En tercer y último lugar, la limitación de los sujetos para describir todas las partes del Universo, que obliga a razonar de forma aislada, considerando sólo los aspectos directamente implicados.

El reconocimiento de estas posibles categorizaciones del azar hace que se conforme un último *Discurso de la Complejidad*, en el que caracteriza el "azar" como elemento provocador de la complejidad existente en la realidad, donde su significación puede estar ligada a un carácter más ontológico.

La noción de aleatoriedad La aleatoriedad, en sí misma, es un concepto complejo; Ayton, Hunt y Wright (1989) describen la variedad de criterios que utilizan los individuos para determinar si una cierta secuencia es aleatoria o no. Nosotros utilizaremos la visión de Kyburg (1974), quien propone la caracterización de la noción de aleatoriedad a partir de cuatro elementos independientes: el objeto de estudio, el conjunto que lo acoge, la proposición que determina dicha pertenencia del acontecimiento de la clase, y un cuerpo de conocimiento como referente.

El *objeto de estudio* es el acontecimiento que hay que enmarcar en el estudio e interpretación probabilística y ha tenido una cierta evolución histórica, dependiendo de las diferentes concepciones sobre la probabilidad. En un primer periodo histórico se consideran como objetos los acontecimientos lúdicos. En un segundo periodo, los objetos considerados son los acontecimientos naturales que susciten el interés de los científicos. En un tercer periodo, los objetos son los acontecimientos de la vida cotidiana de orden social, como unas situaciones a estudiar y modelizar.

El paralelismo entre la evolución histórica del significado de los objetos aleatorios, y la capacidad del sujeto en discriminarlos como tales sugiere una posible relación entre los obstáculos epistemológicos en la identificación de los fenómenos (aleatorios), y los obstáculos ontogénicos asociados a la capacidad de clasificar del sujeto. La superación de dichos obstáculos ha de surgir de un tratamiento didáctico de la noción de fenómeno aleatorio que tenga en consideración la explicación del acontecimiento y su clase de referencia.

La noción de probabilidad Numerosos autores han analizado los obstáculos epistemológicos asociados a la construcción de la noción de probabilidad (Hacking, 1975). En Azcárate (1995) se analizan cuatro etapas diferenciadas en su construcción. En una primera etapa del desarrollo de la idea, que reconocemos como la Prehistoria, donde la idea de probabilidad surge asociada a la noción de juegos de azar, Cardano fue uno de los primeros matemáticos en realizar un argumento teórico para calcular las posibilidades de los distintos resultados relacionados con los juegos de dados. Paralelamente surgieron ideas intuitivas relacionadas con el grado de posibilidad, con algunos cálculos de frecuencias teóricas, sin tener consideración del cuerpo de conocimientos.

En una segunda etapa de *iniciación al cálculo de probabilidades*, se comienza su estudio sistemático, a través, también, de su aplicación al estudio de situaciones de juego ligadas a contextos empíricos. En los textos históricos se suele presentar como la primera sistematización del cálculo de probabilidades de los sucesos la aportada por Pascal y Fermat, relacionándola, muy a menudo, con la *combinatoria*.

Los cálculos probabilísticos empiezan a adquirir consistencia a partir de las aportaciones de Bernoulli. Este autor, estudiando el aparente desorden que presentaban los resultados obtenidos en situaciones de juego, observó una cierta regularidad en la aparición de dichos resultados, y demostró el teorema que se conoce en la actualidad como la *Primera ley de los Grandes Números*. La comprensión incorrecta del significado de la estabilidad de las frecuencias, común en los adultos, se refleja en una transferencia de dicha ley a muestras pequeñas, esperando que se cumpla con un número limitado de pruebas (Kahneman, Slovic y Tversky, 1982). Dicho razonamiento se basa en la creencia de que el azar funciona como un mecanismo auto-correctivo en el que una desviación en una dirección es rápidamente equilibrada por una desviación en la dirección contraria (Falk y Konold, 1992; Bennett, 2000). Esta consideración incorrecta supone un obstáculo epistemológico y ontogénico que dificulta la comprensión de la noción de probabilidad.

En una tercera etapa aparecen dos perspectivas diferenciadas. Una primera perspectiva, la frecuencial, que atenderá el estudio de las frecuencias relativas, y otra, la bayesiana, que se presentará más relacionada con los grados de credibilidad y sus necesarios ajustes con la realidad. Y una última etapa, correspondiente a la introducción de una *Teoría Axiomática*, la construcción de la medida se realiza con independencia de la interpretación real a la que se preste cada situación, aportando un soporte axiomático-deductivo a la teoría matemática (Kolmogorov, 1950). Se entra en la fase de asimilación de los avances teóricos y el posterior desarrollo de aplicaciones.

El análisis de la evolución histórica del cálculo de probabilidades ha permitido identificar los obstáculos que se han superado hasta convertirse en ciencia, así como la dificultad de caracterizar las nociones de azar, aleatoriedad y probabilidad.

2.2. OBSTÁCULOS ONTOGÉNICOS

En las siguientes páginas, presentamos una interpretación de las investigaciones más relevantes sobre la comprensión de las nociones básicas probabilísticas y las estrategias de razonamiento que, en términos de obstáculos ontogénicos, utilizan los sujetos al otorgar significado a dichas nociones.

Obstáculos asociados a la comprensión de las nociones básicas Los obstáculos ontogénicos son descritos en algunas investigaciones sobre desarrollo del razonamiento probabilístico. Por ejemplo, Hoemman y Ross (1982) establecen que los niños, antes de los seis años, no tienen bien definida la relación causa-efecto y, en consecuencia, no diferencian las nociones de azar y causalidad. Es más, Inheler y Piaget (1985, p. 91) argumentan que:

> *...en el nivel preoperatorio, ante el azar los sujetos presentan una actitud paradójica: esperan que frente a condiciones semejantes los fenómenos se repitan de modo idéntico.*

En un principio el niño se desorienta ante lo inesperado o fortuito, pero luego, progresivamente, busca causas que justifiquen "más o menos" las fluctuaciones encontradas, lo que les lleva a buscar razones ocultas para los hechos de un cierto orden oculto; por ejemplo, cuando esperan una cierta compensación en la aparición de los posibles resultados esperados, como una autorregulación de los resultados.

A este respeto, Bennett (2000, p. 13) indica que:

> *Las ideas intuitivas sobre el azar pueden preceder a las ideas formales y, si son correctas, pueden ser de gran ayuda en el aprendizaje; pero en caso contrario, pueden llegar a dificultar la correcta comprensión de los conceptos.*

En consecuencia, pensamos que la no-inclusión de esta noción en la enseñanza puede suponer un obstáculo didáctico en la construcción de la noción de probabilidad, ya que ésta se sustentará en ideas intuitivas que no permitirán comprender la naturaleza de los fenómenos a los que se enfrentan.

Fischbein, Nello y Marino (1991), destacan las dificultades de los sujetos en la comprensión de las nociones de suceso simple y compuesto, que puede ser un obstáculo para la posterior comprensión de las nociones de sucesos equiprobables, sucesos contrarios y sucesos independientes.

La distinción entre sucesos equiprobables o no, se puede presentar de dos formas diferentes. Una primera puede provenir de una hipótesis que establezca el sujeto sobre la simetría del azar, igualdad de oportunidades o posibilidades de ocurrencia del suceso. Una segunda, que apela a un juicio no cuantitativo de la equiprobabilidad, en relación exclusivamente con su carácter fortuito. La subjetividad asociada a la determinación de la simetría o equidad del azar puede ser un obstáculo para la determinar si dos sucesos son equiprobables. Es más, dicha subjetividad favorece la aparición del llamado sesgo de la equiprobabilidad (Lecoutre y Duran, 1988), en el que los sujetos consideran los posibles resultados de cualquier fenómeno equiprobables porque son materia del azar.

Los sujetos también presentan dificultades en la determinación de si dos sucesos son independientes y, en consecuencia, en la construcción del significado de la noción de independencia estocástica. Truran y Truran (1996) analizan el concepto de independencia, indicando que es diferente considerar sucesos aleatorios que pruebas aleatorias e indican la necesitad de dos caracterizaciones diferentes con la finalidad de no generar un obstáculo didáctico. Por una parte, se debe introducir el significado de independencia clásica, mediante la regla del producto, y por otra, reflexionar sobre la independencia de la prueba, que refleja que los resultados de una prueba aleatoria no están influenciados por los resultados de cualquier otra prueba del mismo generador aleatorio. El reconocimiento de la dependencia clásica necesita de la distinción de sucesos excluyentes. En el mundo real los sucesos mutuamente excluyentes no son necesariamente sucesos complementarios, lo cual provoca una confusión en el sujeto (Sánchez, 1999). El obstáculo podría surgir en el momento en que el sujeto debe realizar argumentaciones para determinar las posibles pertinencias a ciertas clases que generan los sucesos aleatorios.

Obstáculos asociados a las estrategias de razonamiento Entre las investigaciones relacionadas con las estrategias de razonamiento que utilizan los sujetos ante situaciones de incertidumbre citamos las realizadas por Konold (1995). Konold usa el término "outcome approach" para referirse a los razonamientos de los sujetos que obvian en el marco global de decisión la serie aleatoria y evalúan en función del resultado siguiente, cuando se encuentra bajo condiciones de incertidumbre.

Es significativo el conjunto de investigaciones relativas al heurístico de la *representatividad* (Tversky y Kahneman, 1982). Los sujetos consideran que un resultado debe ser representativo de la población que proviene, es decir, representativo del conjunto de características o cualidades sobre los que se trabaja. Según Cardeñoso (2001), la enseñanza reafirma esta creencia y no elimina sus sesgos, configurándose no sólo como un posible obstáculo ontogénico, sino como uno didáctico, que no facilita la comprensión de las situaciones dominadas por la incertidumbre.

Relacionado con el uso del heurístico de la representatividad, están las argumentaciones basadas en la *falacia del jugador*. Esta falacia se fundamenta en una idea errónea sobre la imparcialidad de las leyes del Azar. Esta falacia se puede constituir en un obstáculo en la comprensión del significado de la estabilidad de las frecuencias relativas, al considerar que la estabilidad se puede dar en series limitadas de números.

En último lugar, en las investigaciones se reafirma la importancia del *sesgo de la equiprobabilidad*, ya introducido con anterioridad. Bajo este sesgo los sujetos consideran los posibles resultados como equiprobables porque son materia del azar. Este sesgo se constituye como un obstáculo para la comprensión de noción de aleatoriedad, que reduce su significado a argumentaciones basadas en la equiprobabilidad de los sucesos.

2.3. OBSTÁCULOS DIDÁCTICOS

El análisis de los obstáculos de carácter didáctico que presentamos se centra en dos aspectos de la construcción del conocimiento: uno relacionado con el uso del lenguaje probabilístico, y otro con los contextos de ejemplificación y experimentación para la construcción del conocimiento.

Obstáculos asociados con el uso del lenguaje probabilístico La importancia del lenguaje en la construcción del conocimiento probabilístico se refleja en los diferentes diseños curriculares. En los trabajos de la Comisión Internacional IREM para la enseñanza de la probabilidad y la estadística en Francia, se sugiere que:

> *Conviene precisar el vocabulario, de forma que a cada nivel de descripción se asocien términos específicos que permitan a los alumnos tener en cuenta el punto de vista en que nos situamos: realidad, sentido, sensible, modelo, modelo probabilista* (Henry, 2001, p. 164).

Todas estas recomendaciones son importantes puesto que, cuando el alumno se inicia en el estudio de la probabilidad, ha usado en sus juegos y vida diaria términos y expresiones para referirse a los sucesos aleatorios que, con frecuencia, no tienen el mismo sentido preciso que adquieren en el "Tratamiento del Azar" (Ortiz y Serrano, 2001). Estas diferencias existentes entre el lenguaje cotidiano y el lenguaje probabilístico pueden ser un obstáculo para la construcción del conocimiento. Por ejemplo, las investigaciones de Truran (1994) sugieren que muchos niños confunden los términos “imposible” y “muy poco probable”. Lo que resulta problemático no son los términos imposible, seguro,… en sí mismos, sino los conceptos y procesos subyacentes que se están comunicando y el significado que transmiten.

Batanero y Serrano (1999) indican que la introducción de la idea de *aleatoriedad* se hace preferentemente de un modo descriptivo, cobrando un papel primordial los matices de lenguaje. La descripción de las características atribuidas a los resultados de los experimentos se realiza mediante palabras como imprevisibles, incierto, etc., con las que se pretende que se evoquen las propiedades de tales fenómenos, pero cuyo significado no suele clarificarse. La falta de clarificación de la noción de aleatoriedad deja abierta la posibilidad de interpretación ambigua, y se puede configurar como un obstáculo en la comprensión de la noción de aleatoriedad por parte de los alumnos. Esta idea indica que un tratamiento inadecuado de la forma de contextualizar y referenciar los objetos (acontecimientos, fenómenos, experimentos aleatorios,…) puede ocasionar un obstáculo didáctico en la comprensión de la noción de aleatoriedad y probabilidad.

Por otro lado, en la enseñanza de las matemáticas se pueden encontrar tres tipos diferentes de palabras (Pimm, 1987):

- Palabras técnicas, que normalmente no forman parte del lenguaje cotidiano.
- Términos que aparecen en matemáticas y en el lenguaje ordinario, aunque no siempre con el mismo significado en los dos contextos, como límite o convergencia.
- Palabras que tienen significados iguales o muy próximos en ambos contextos.

En el caso de la enseñanza de la probabilidad en niveles no universitarios, la mayoría de los vocablos pertenecen a las dos últimas categorías aunque, si el niño no está muy familiarizado por el uso, muchas de las palabras de la tercera categoría se convertirán en términos de la segunda, lo que podrá crear dificultades de comunicación en el aula.

Obstáculos asociados a la experimentación y ejemplificación Heitele (1975) indica que las posibilidades didácticas que se deducen de las experiencias empíricas son más limitadas de lo que sugieren los textos escolares. Las sucesiones aleatorias obtenidas en clase convergen lentamente. Debido, a su carácter aleatorio, puede ocurrir que no se obtenga el resultado deseado cuando se quiera mostrar con una simulación una cierta probabilidad.

Con referencia a los generadores aleatorios, que generalmente se utilizan para introducir las nociones de probabilidad, se observa que una ruleta visualiza mejor la relación parte-todo, y por tanto la Regla de Laplace, y además, permite al alumno utilizar consideraciones de tipo geométrico. Las chinchetas permiten ejemplificar situaciones de sucesos no equiprobables. Las barajas, urnas o bolsas con bolas permiten trabajar situaciones de muestreo con y sin remplazamiento, facilitando la determinación de la probabilidad para sucesos dependientes e independientes, que no es fácil ejemplificar con otros dispositivos (Ortiz y Serrano, 2001). El uso exclusivo de uno de estos instrumentos al introducir una noción, puede favorecer una asociación del dispositivo con el concepto a aplicar, sin favorecer el aprendizaje significativo de las propiedades de los mismos (Azcárate, 1995; Serradó, 2003).

La revisión anterior sobre algunos posibles obstáculos en la construcción del Conocimiento Probabilístico presentada en esta sección se configura como un referente teórico para el análisis de su tratamiento en los libros de texto de Educación Secundaria Obligatoria.

3. DISEÑO DE LA INVESTIGACIÓN

Este trabajo de investigación se enmarca en un estudio más amplio (Serradó, 2003; Azcárate, Cardeñoso, y Serradó, 2003) que analiza el contenido de las unidades dedicadas al "Tratamiento del Azar" en los libros de textos de Educación Secundaria Obligatoria en España (de 12 a 16 años). El presente artículo, presenta los resultados relacionados con el análisis de los obstáculos epistemológicos, ontogénicos y didácticos, que se pueden inducir desde los libros de texto a la hora de desarrollar los contenidos probabilísticos.

Este estudio presenta diferentes niveles de reflexión, un primer análisis del contenido de los libros de texto, centrado en la descripción de cada uno de los posibles obstáculos en relación con las nociones consideradas previamente en los libros de la muestra. Un segundo análisis explicativo que surge del contraste entre los resultados provenientes del análisis anterior y el marco teórico elaborado. Al ser una investigación cualitativa, hemos considerado una muestra intencional de libros de texto, que ha sido seleccionada desde un perfil de atributos a cumplir. Se han buscado editoriales con distribución en Andalucía, que traten el azar a lo largo de todos los cursos de la etapa y tengan desarrollado el proyecto curricular en las tres etapas de Educación no universitaria. La tabla 1 recoge los atributos que cumplen los libros de texto de la muestra.

Tabla 1. Atributos de los textos de la muestra

Atributos	*Santillana*	*Bruño*	*Guadiel*	*McGraw Hill*
Alta incidencia en el mercado nacional	SI	SI	SI	SI
Alta incidencia en el mercado andaluz	SI	SI	SI	SI
Impresión en Andalucía	NO	NO	SI	NO
Tratamiento del Azar longitudinal	SI	NO	NO	SI
Proyecto curricular de las tres Etapas	SI	SI	NO	NO

En cada uno de los textos se identificaron las explicaciones, ejemplificaciones y actividades que utilizaban para tratar cada una de las nociones consideradas (aleatoriedad y probabilidad), se organizó la información y se procedió a su análisis.

En la siguiente sección se presentan los resultados correspondientes a la caracterización de los posibles obstáculos epistemológicos, ontogénicos y didácticos que pueden inducir en su tratamiento los libros de texto, en función del marco teórico elaborado.

4. TRATAMIENTO DE LOS OBSTÁCULOS EN LOS LIBROS DE TEXTO

4.1. NOCIÓN DE ALEATORIEDAD

Con referencia a la *noción de aleatoriedad*, ninguno de los libros de texto analizados introduce una sección dedicada al estudio de dicha noción, sino que se refieren en diferentes apartados a nociones relacionadas como azar, fenómeno y/o experimento aleatorio, proceso y suceso, serie aleatoria, etc.

Azar Ninguno de los libros de texto presenta una sección dedicada al estudio de la noción de azar. Se presentan referencias implícitas al significado que adquiere dicha noción, a partir de la presentación de textos y actividades de motivación. En estos textos y actividades se caracteriza el azar como una causa desconocida, otorgándole un significado mágico, asociado a la suerte y un carácter causal asociado a la imprevisibilidad del fenómeno. Dicha caracterización identifica el azar, como el reflejo de una causa desconocida, y dificulta su caracterización como un elemento provocador de la complejidad existente en la realidad (Azcárate, Serradó y Cardeñoso, 2004). A este respecto, Bennett (2000, p. 13) indica que:

Las ideas intuitivas sobre el azar pueden preceder a las ideas formales y, si son correctas, pueden ser de gran ayuda en le aprendizaje; pero en caso contrario, pueden llegar a dificultar la correcta comprensión de los conceptos.

En consecuencia, pensamos que la no-inclusión de esta noción en la enseñanza y la falta de reflexión sobre su significado, puede favorecer la aparición de un *obstáculo epistemológico* para la comprensión de las nociones de aleatoriedad y probabilidad, ya que éstas se sustentarán en ideas intuitivas que no permitirán comprender la naturaleza de los fenómenos a los que se enfrentan. Y, en la misma línea puede reflejar suponer un *obstáculo didáctico* ya que es una situación producto de decisiones docentes.

Fenómeno y experimento Los libros de texto de las diferentes editoriales realizan un uso indiferenciado de las nociones de fenómeno y/o experimento, que se puede observar en los dos siguientes ejemplos:

Podemos predecir, con toda seguridad, lo que va a ocurrir: se trata de fenómenos o experimentos deterministas (Bruño, 4º A ESO, pág. 266).

El lanzamiento de un dado, de una moneda, el giro de la ruleta o la extracción de una bolsa son experimentos en los que no se puede predecir el resultado que se va a obtener. Se llaman experimentos aleatorios y su resultado depende del azar (Santillana, 1º ESO, pág. 248).

Las cuatro editoriales de la muestra definen el experimento aleatorio a partir de la imposibilidad de predecir el resultado. Por ejemplo:

En estos casos no podemos conocer previamente el resultado de las experiencias, que reciben el nombre de experimentos aleatorios (Guadiel, 2º ESO, pág. 128).

La presentación de una única modelización puede inducir un obstáculo para la comprensión del amplio significado de esta noción, y un obstáculo epistemológico en la posterior comprensión de la noción de probabilidad, tal como se ha argumentado en la sección anterior.

Las definiciones de *experimento determinista* propuestas en las cuatro editoriales están relacionadas todas con la posibilidad de predecir a priori el resultado, como contraposición a la definición de experimento aleatorio propuesto. La presentación de ambas nociones como antagónicas puede suponer un *obstáculo epistemológico* en la comprensión adecuada de las situaciones dominadas por la incertidumbre y en relación al significado de causa-efecto (Serradó, Azcárate y Cardeñoso, 2005). Este tipo de obstáculos se puede manifestar en forma de sesgos asociados al determinismo de los fenómenos, que surgirán en un intento de delimitar si ciertos fenómenos dependen o no de la incertidumbre.

Las editoriales Bruño y Santillana presentan básicamente actividades de carácter conceptual asociadas al reconocimiento de los experimentos aleatorios y/o deterministas. Las actividades de las editoriales Guadiel y McGraw Hill permiten analizar las diferentes propiedades definitorias de los fenómenos. Por ejemplo:

Cuando llegas a un semáforo: ¿De qué color estará? ¿De qué depende? ¿Qué es más fácil que esté: rojo, amarillo o verde? (Mc Graw Hill, 2º ESO, pág. 249, act. 7).

Tanto en el nivel de introducción de las nociones teóricas, como en el tipo de actividades propuestas, se favorece el uso descriptivo del lenguaje para incidir en la caracterización de las diferencias existentes entre los fenómenos/experimentos aleatorios o deterministas. En este sentido, la falta de un vocabulario adecuado o el uso inadecuado de éste puede ocasionar un *obstáculo ontogénico* en la comprensión del significado de la noción de experimento aleatorio.

Proceso y suceso aleatorio El análisis del contenido de los libros de texto refleja que, en general, no se enfatiza el estudio de las diferencias entre las nociones de proceso y suceso aleatorio. La falta de distinción de estas nociones se puede configurar en un obstáculo ontogénico en la construcción de las nociones de suceso en el experimento simple y en el experimento compuesto (Fischbein, Nello y Marino, 1991). A su vez, la consideración de estas diferencias puede ser esencial para la posterior comprensión de las nociones de experimento compuesto y serie aleatoria, básica para el estudio de la noción frecuencial de la probabilidad (Azcárate, Serradó y Cardeñoso, 2004).

Sucesos elementales y espacio muestral En los textos se define la noción de espacio muestral a partir del uso del término conjunto, que no se clarifica, como se observa en la siguiente definición del texto de la editorial Bruño:

El espacio muestral es el conjunto de todos los sucesos elementales, y se designa por la letra E (1º de ESO, pág. 238).

Esta falta de clarificación puede configurar un obstáculo ontogénico para la comprensión del espacio muestral, siguiendo la línea argumental de la sección anterior. Una mayor comprensión de su significado se podría obtener del análisis de las ejemplificaciones y actividades propuestas en el texto. Dichas ejemplificaciones y actividades se reducen a la identificación de los sucesos que lo componen; por ejemplo la actividad propuesta por la editorial Santillana:

Se lanzan dos dados. ¿Es éste un experimento aleatorio? En caso afirmativo escribe el espacio muestral (4º B de ESO, pág. 264).

Salvo en excepciones, los textos presentan todas las ejemplificaciones y actividades asociadas a la identificación de espacios muestrales de experimentos asociados a generadores aleatorios particulares, que facilitan su determinación y cuantificación, ya que la repetición de los lanzamientos puede imaginarse de forma más rápida (Meletiou y Stylianou, 2003). Sin embargo, la reducción de los contextos de ejemplificación y experimentación puede ser un *obstáculo didáctico* para la transferencia de dichas nociones a otros contextos, como el social, en que los espacios muestrales no son tan explícitos, y no siempre se comprenden correctamente.

Serie aleatoria En los libros de textos se utiliza el término de serie aleatoria para caracterizar las secuencias de sucesos aleatoria, por ejemplo:

Los resultados se han anotado en una tabla de recuento que te presentamos a continuación, en series de 20, 40, 60, 80 y 100 lanzamientos… (McGraw Hill, 3º).

En relación con la noción de serie aleatoria, no se proponen, en general, actividades en que los alumnos tengan que reflexionar sobre su significado, ni se enfrenta a los alumnos con situaciones en que tengan que verbalizar el uso de heurísticos como la falacia del jugador. Este uso del término serie, sin conceptualizarlo, genera una pobre imagen de la noción y puede conllevar posibles *obstáculos didácticos* al asociarlo a secuencias aleatorias muy poco representativas de la potencialidad de la noción de serie aleatoria (o/y de su convergencia), con las consecuencias posteriores para la visión frecuencial. En el caso de que los textos sean las únicas fuentes de información a la hora de planificar y desarrollar el proceso de enseñanza y aprendizaje, podría suceder que las argumentaciones de los alumnos estuviesen dominadas por el uso de heurísticos, convirtiéndose en un sesgo para la adecuada comprensión del significado de la probabilidad, y de la probabilidad condicionada.

A modo de síntesis en la Tabla 2 hemos recogido el tratamiento reflejado en los libros de las diferentes nociones y la naturaleza del obstáculo que dicho tratamiento puede inducir.

Tabla 2. Posibles obstáculos desde el tratamiento de la noción de aleatoriedad

Noción analizada	Tratamiento en los libros de texto (1)	Obstáculo que se podría promover debido a la presentación	Tipo de obstáculo
Azar	Única caracterización, como causa desconocida (B, S, G, M).	Comprensión del tipo de fenómenos que aborda el cálculo probabilístico	Epistemológico
Experimento Aleatorio	Única modelización como imposibilidad de predecir el resultado (B, S, G, M).	Comprensión del amplio significado de esta noción	Epistemológico
Experimento Determinista	Definición a partir de la posibilidad de predecir a priori el resultado, (B, S, G, M).	Comprensión inadecuada de las situaciones de incertidumbre en relación con el significado de causa-efecto	Epistemológico
	Falta de énfasis en el establecimiento de la relación causa-efecto, (B, S, G, M).	No distinción entre azar y causalidad	Ontogénico
Proceso/ Suceso	Falta de distinción de proceso y suceso aleatorio, (B, S, G, M).	No distinción entre repetición de un experimento y experimento con múltiples resultados posibles. No distinción entre suceso experimento simple/ compuesto y entre experimento compuesto y serie aleatoria	Ontogénico
Espacio muestral/ Suceso elemental	Introducción a partir de la noción de conjunto, que no se clarifica (B, S, G, M-3º-4º)	Obstaculizar argumentaciones para determinar la pertenencia a ciertas clases y asignación de probabilidades	Ontogénico
	Actividades y ejemplos asociadas a solo a un tipo de generadores aleatorios (S, G).	Obstaculizan la transferencia a otros contextos	Didáctico
Serie aleatoria	Falta de reflexión sobre la "serie" aleatoria, (B, G).	Puede favorecer el uso de heurísticos	Didáctico

(1): Bruño (B); Santillana (S); Guadiel (G) y McGraw Hill (M)

4.2. NOCIÓN DE PROBABILIDAD

El significado escolar de la noción de probabilidad se suele introducir asociado al valor numérico que se asigna a cada suceso, o como el valor que se obtiene del análisis de la estabilidad de las frecuencias relativas.

Noción frecuencial de la probabilidad En los textos de las cuatro editoriales de la muestra existen dos perspectivas diferenciadas para el análisis de la estabilidad de las frecuencias relativas, aunque ninguna tiene en consideración las posibles dificultades asociadas a la comprensión del significado de la convergencia estocástica, lo cual puede configurar un obstáculo ontogénico.

La editorial Santillana y Bruño introducen estas nociones a nivel teórico, sin proponer actividades que favorezcan la exploración de las tendencias de las series aleatorias, como se puede analizar en el siguiente ejemplo (Santillana, 3º, pág. 252):

Observa la gráfica siguiente (Figura 1) y mira como, conforme aumenta el número de lanzamientos se observa que:
**las oscilaciones de la gráfica son menos pronunciadas (menos picos)*
**las frecuencias relativas se van acercando a un determinado valor, es decir, tiende hacia un número fijo. El número es 0,5.*
Esta tendencia se observa en todos los fenómenos aleatorios y se denomina ley del azar o de estabilidad de las frecuencias, también denominada primera ley de los grandes números.

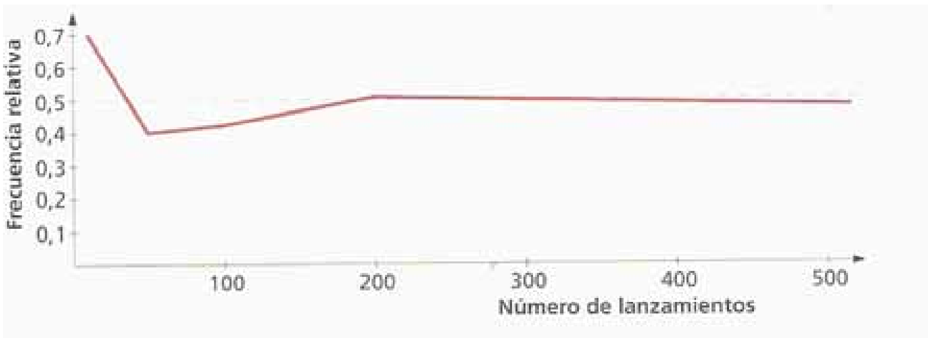


Figura 1. Evolución de las frecuencias relativas

El texto presenta una breve aclaración del significado de tender hacia un número fijo:

Cuando son pocas las veces que repites el lanzamiento de una moneda, la frecuencia relativa que esperas 0,5, y la que obtienes se parecen muy poco. Conforme vas aumentando el número de lanzamientos, esa diferencia se va reduciendo, y cuando lo repites muchísimas veces, esa diferencia se hace casi nula. Quiere esto decir que la frecuencia relativa empírica (la que obtienes al repetir el experimento) se parece a la frecuencia relativa esperada, tanto más cuanto mayor sea el número de lanzamientos, y llega a ser casi idéntica cuando el número de lanzamientos es muy grande.

La propuesta de actividades, que complementa la introducción de la noción frecuencial de la probabilidad para estas dos editoriales, se reduce al cálculo de la frecuencia relativa, a partir de una tabla que recoge la frecuencia absoluta teórica de un dado. No se pretende que los alumnos experimenten la repetición de un experimento aleatorio, y construyan series aleatorias que les permitan indagar sobre el significado de estabilidad de frecuencias aleatorias, y comprender el significado de la noción de probabilidad. Esta propuesta se puede configurar como un *obstáculo didáctico* para la construcción de la noción frecuencial de la probabilidad puesto que los alumnos, podrían basar sus argumentaciones en el uso de heurísticos como el "outcome approach" o la falacia del jugador. También se podrían convertir en un *obstáculo ontogénico* en la comprensión del significado de la estabilidad de las frecuencias relativas.

En cambio, las editoriales Guadiel y McGraw Hill proponen un conjunto de actividades para que los alumnos puedan analizar el significado de la estabilidad de las frecuencias relativas, como se puede ver en la siguiente actividad (Guadiel, 2º ESO, pág. 132, ej. 15):

Coge un dado de quinielas y efectúa 50 lanzamientos. Anota el número de veces que sale cada signo y completa la siguiente tabla con las frecuencias absolutas y relativas de los sucesos 1, X y 2.

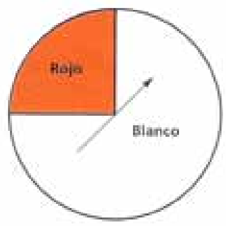
Suceso	Frecuencia Absoluta	Frecuencia Relativa
1		
X		
2		

Reúne en una sola tabla tus resultados y los de tus compañeros y analízalos.

El texto no introduce ningún comentario sobre cómo se deben analizar los resultados, dejando a la elección del profesor y/o alumno el nivel de profundización en la inferencia del significado de la estabilidad de las frecuencias, la ausencia de regularidad de las series aleatorias o la impredicibilidad del resultado. En el caso de la editorial McGraw Hill y Guadiel, las propuestas didácticas incluyen actividades que favorecen la reflexión sobre el significado de la noción frecuencial de la probabilidad, con un número reducido de experimentos. La consideración de un número pequeño de repeticiones puede, a la larga, ocasionar un *obstáculo epistemológico* en la comprensión del significado de convergencia aleatoria, y reforzar la errónea validez del heurístico de la representatividad.

Comparación de probabilidades En los textos de las editoriales Bruño y Santillana, se introduce el proceso de cuantificación de las probabilidades, y la asignación de un valor numérico a partir de la comparación entre sucesos aleatorios correspondientes a un único tipo de generadores aleatorios. Un ejemplo de esta situación se puede observar en el siguiente texto incluido en el texto de 3º de ESO de la editorial Bruño (pág. 274):

Si giramos la aguja, parece claro que la probabilidad de pararse en la zona "Blanco" es mayor que la probabilidad de pararse en la zona "Rojo". Se trata de un experimento con dos sucesos elementales que tienen probabilidades distintas.



Este tipo de generadores aleatorios permite que los alumnos indaguen sobre las comparaciones de tipo geométrico, que favorecen la visualización de la proporción a considerar. Sin embargo, aunque facilitan el establecimiento de comparaciones, pueden fomentar la asociatividad entre este tipo de generador y la estrategia a aplicar para asignar las probabilidades, pudiéndose configurar como un *obstáculo didáctico* para la realización de comparaciones mediante otros generadores.

La consideración de las expectativas del suceso, entendidas en el sentido de Huygens, ha de permitir la correcta comprensión del significado del valor esperado (esperanza matemática) y el significado de juego justo. Sólo la editorial McGraw Hill introduce en sus textos actividades en que los alumnos deben inferir el valor esperado, como por ejemplo:

Lanza 100 chinchetas y estima la probabilidad de que una chincheta caiga con la punta hacia arriba. (McGraw Hill, 4º A, pág. 212, ej. 15).

La introducción de ejemplificaciones que favorezcan la comparación de sucesos, y la asignación de un valor numérico a estos, se introduce en los textos como paso previo a las explicaciones relativas a la aplicación de la *Regla de Laplace*. Los textos de las cuatro editoriales de la muestra presentan estrategias diferentes para poder aplicar esta regla.

Equiprobabilidad y regla de Laplace Los textos no clarifican las condiciones requeridas para la equiprobabilidad. La editorial Bruño identifica en 1º de ESO (pág. 242), el significado de equiposible con equiprobable:

Cuando los sucesos de un experimento aleatorio tienen las mismas posibilidades de obtenerse se dice que son sucesos equiprobables.

En el texto se indica que, para poder aplicar la Regla de Laplace, los contextos de aplicación deben ser equiprobables, presentando como ejemplificaciones de sucesos elementales equiprobables los correspondientes al lanzamiento de dados y monedas. Sólo en 3º de ESO, presenta ejemplos de generadores aleatorios como chinchetas o ruletas en que, según el texto, visualmente se puede determinar que no son equiprobables. No se incluyen actividades en que los alumnos puedan experimentar sobre el significado de sucesos equiprobables.

La editorial Guadiel, al igual que la editorial Bruño, propone en la definición de suceso equiprobable la identificación con el significado de equiposible. La diferencia radica en que propone una actividad para comprobar si un experimento es equiprobable, previo a la explicación del significado de esta noción:

Tenemos dos dados, uno perfecto, en el que todas las caras pesan igual, y otro en el que una de las caras pesa más.
¿Crees que en los dos dados todos los sucesos tienen la misma probabilidad de producirse? (Guadiel, 4º, pág. 236).

La editorial Santillana define los sucesos elemental equiprobables a partir de que tengan la misma probabilidad de salir:

Todos los sucesos elementales tienen la misma probabilidad de salir, y se denominan equiprobables (Santillana, 4º A, pág. 187).

Dicha definición, se complementa con un ejercicio para razonar si los sucesos elementales de un cierto experimento son equiprobables:

Justifica tu respuesta. En el lanzamiento de una moneda, los sucesos correspondientes al número de caras, ¿son equiprobables? (Santillana 4º A, pág. 189).

La editorial McGraw Hill propone, previamente a la introducción de la definición de suceso equiprobable, el análisis de la estabilidad de las frecuencias relativas de sucesos elementales equiposibles y no. Indica que la aplicación de la regla de Laplace sólo puede realizarse para sucesos elementales equiprobables, indicando que, en el caso contrario, se debe estimar a partir del análisis de la estabilidad de las frecuencias relativas:

En ejemplos como los anteriores, o en algunas experiencias como las del lanzamiento de chinchetas o tabas, o la predicción de la probabilidad de que un fumador padezca una enfermedad respiratoria, etc., solamente podemos calcular la probabilidad haciendo una estimación (observando un gran número de casos y calculando la frecuencia relativa) (McGraw Hill, 4º A, pág. 208).

En cada uno de los niveles educativos se introducen actividades que tienen por finalidad analizar si ciertos sucesos elementales son equiprobables o no, asociados a la experimentación con variedad de generadores aleatorios y fenómenos aleatorios relacionados con situaciones cercanas al alumnado.

En resumen, el tratamiento de la noción de equiprobabilidad que se realiza en los textos puede ocasionar dificultades en la comprensión de dicha noción, ya que no se clarifica. Además, en los textos en que no se presentan actividades para razonar si los sucesos elementales son equiprobables, dándose por supuesta esta propiedad, se puede fomentar la aparición del *sesgo de la equiprobablidad*, que se configura como un *obstáculo ontogénico*.

Este sesgo se refuerza en los libros de diferentes formas. En primer lugar, por la ignorancia en los textos de la importancia de la comprobación de esta propiedad como paso previo a la aplicación de la Regla de Laplace. En segundo lugar, por la presencia de la equi-posibilidad, al presentar los ejercicios asociados a juegos aleatorios con sucesos elementales equiposibles como monedas, dados, etc. Y, en tercer lugar, el método dicotómico que reposa sobre las opciones de aparición y de no aparición de cada resultado aislado, con un 50% para cada posibilidad, sesgados por el uso de heurísticos como la falacia del jugador o el "outcome approach", comentados en la sección anterior.

El uso de estos heurísticos también pueden ser un obstáculo para la comprensión del significado de dependencia e independencia de sucesos aleatorios, ya que predisponen a considerar que dos sucesos consecutivos están siempre relacionados (Konold, Pollatsek, Well, Lohmeier y Lipson, 1993; Meletiou y Stylianou, 2003).

Dependencia e independencia de los sucesos aleatorios La introducción de dichas nociones difiere según el texto de cada editorial. La principal diferencia que existe entre las cuatro editoriales radica en cómo se introduce el significado de sucesos dependientes e independientes. La editorial Bruño distingue ambas definiciones a partir de identificar las diferencias existentes al realizar extracciones con o sin reemplazamiento. La editorial Santillana define la independencia o no de sucesos en función de la influencia de la ocurrencia de un suceso respecto al otro. Ambas editoriales deducen, mediante ejemplificaciones en que se aplica la regla de Laplace, las fórmulas que relacionan la probabilidad condicionada con el producto de probabilidades. En cambio, las editoriales Guadiel y McGraw Hill definen directamente la dependencia e independencia a partir del establecimiento de estas igualdades. En el caso concreto de la editorial McGraw Hill se introduce previamente el significado de la probabilidad condicionada a partir del análisis de la diferencia existente al realizar experimentos con y sin reemplazamiento. Las actividades propuestas en las cuatro editoriales reducen el estudio de la dependencia e independencia a la caracterización de las diferencias debidas al reemplazamiento o no, como por ejemplo:

Extraemos dos cartas de una baraja, una después de otra y sin devolución. Calcula la probabilidad de obtener dos figuras en los siguientes casos:

a) La primera carta es un as
b) La primera carta es un caballo (Guadiel 4º, pág. 241, ej. 28).

Los contextos básicos que se utilizan para analizar la dependencia e independencia de los sucesos aleatorios y calcular su probabilidad son generadores aleatorios como urnas y barajas, que facilitan la comprensión del significado de reemplazamiento. La introducción de estas nociones, a partir de un tratamiento que se fundamente en el análisis del reemplazamiento con estos generadores aleatorios puede ser un *obstáculo didáctico* al asociar la estrategia de cálculo de la probabilidad con el generador aleatorio, impidiendo, además, la generalización de estas nociones a otros contextos no asociados con pruebas aleatorias (Truran y Truran, 1996) o la comprensión del significado en el mundo real (Kelly y Zwiers, 1988).

A modo de resumen recogemos en la Tabla 3 el tratamiento reflejado en los libros de las diferentes nociones asociadas a la probabilidad y la naturaleza del obstáculo que dicho tratamiento puede inducir.

Tabla 3. Posibles obstáculos asociados con el tratamiento de la noción de probabilidad

Noción analizada	Tratamiento en los libros de texto (1)	Obstáculo que se podría promover debido a la presentación	Tipo de obstáculo
Frecuencial de Probabilidad	Falta de reflexión sobre la convergencia estocástica (B, S, G, M).	Transferirla convergencia a muestras pequeñas	Ontogénico
	Introducción, sólo a nivel teórico, de la estabilidad de las frecuencias relativas, (B, S).	Argumentaciones basadas en el uso de heurísticos	Didáctico
Comparación de Probabilidades	Cuantificación y asignación de probabilidades a partir de la comparación entre sucesos aleatorios generados mediante ruletas, (B, S).	Asociar el generador y la estrategia al asignar probabilidades	Didáctico
Equiprobabilidad	Circularidad al definir equiprobabilidad (B-1º, G-4º) Suponer la equiprobabilidad (B).	Sesgo de equiprobabilidad	Ontogénico
Dependencia e independencia	Actividades y ejemplos asociadas a generadores aleatorios sólo en contexto de reemplazamiento, (B, S, G, M).	No transferencia a otros contexto	Didáctico

(1): Bruño (B); Santillana (S); Guadiel (G) y McGraw Hill (M)

5. REFLEXIONES FINALES

Nuestro análisis muestra que los textos analizados, en general, no tienen en consideración diversos aspectos en el “Tratamiento del Azar”, que pueden convertirse en obstáculos en la construcción de las nociones introducidas en dichos textos y para otras nociones que se desarrollaran en cursos posteriores.

En primer lugar, los textos presentan básicamente el *azar* modelizado a partir de argumentaciones asociadas con la suerte y la *aleatoriedad* con la incertidumbre del suceso, caracterizaciones que son insuficientes para poder comprender adecuadamente el significado de las nociones probabilísticas. La comprensión adecuada de la noción de sucesión aleatoria necesita de la valoración subjetiva de la complejidad de la secuencia, que se sustenta en una caracterización de la noción de aleatoriedad que supera las modelizaciones simples propuestas en los textos. La contraposición a la noción de fenómeno determinista, enfatiza las relaciones de causa y efecto que, a su vez, pueden producir sesgos en la interpretación correcta del significado de sucesos dependientes e independientes y constituirse como un obstáculo para la posterior comprensión de la noción de probabilidad.

En segundo lugar, en los textos se incluyen básicamente ejemplificaciones y actividades asociadas a *generadores aleatorios con espacios muestrales finitos y sucesos elementales equiprobables*, que surgen de estudios de generadores aleatorios simples, como ruletas, urnas, dados, barajas o monedas, que facilitan determinar los elementos del espacio muestral, pero se pueden configurar como un obstáculo para la posterior determinación y generalización del significado de espacio muestral a contextos cotidianos y el posterior cálculo de la probabilidad. A su vez, estos ejemplos y actividades asociadas a generadores aleatorios facilitan la caracterización de la equiprobabilidad de los sucesos que, a menudo, no se solicita en el texto. La falta de comprobación de dicha propiedad favorece la identificación de la equiprobabilidad como equi-ignorancia de las propiedades de los sucesos y podrá constituir un obstáculo para la determinación de sucesos aleatorios no equiprobables.

En tercer lugar, en los libros de texto no se clarifica el significado de ciertos términos como imprevisible, colección, conjunto, seguro, imposible, converge,... lo que podría dificultar la construcción de las nociones probabilísticas que los sustentan, como fenómeno y experimento, suceso y proceso, espacio muestral, secuencia aleatoria y estabilidad de las frecuencias relativas.

En cuarto lugar, en los textos la construcción de la *noción frecuencial de la probabilidad* está sujeta a la comprensión adecuada de la convergencia estocástica, no iniciada en estos niveles educativos, y que se desarrolla a nivel intuitivo. En las ejemplificaciones y actividades de los textos no se incluye cuestiones que permitan que los alumnos sean conscientes de sus concepciones sobre la incidencia del azar en las series aleatorias, apareciendo en forma de heurísticos (falacia del jugador, representatividad de la muestra, “outcome approach”, etc.).

Todos estos resultados aportan elementos a tener en cuenta en la elaboración de libros de texto, tanto para la consideración de los posibles obstáculos que la presentación y tratamiento del conocimiento probabilístico pueden inducir, como para promover su integración en el proceso de enseñanza y aprendizaje.

REFERENCIAS

Ayton, P., Hunt, A.J., y Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, *2*, 221-238.

Azcárate, P. (1995*). El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad. Su estudio en el caso de la educación primaria.* Tesis doctoral inédita. Universidad de Cádiz.

Azcárate, P., Cardeñoso, J. M., y Serradó, A (2003). Hazard's treatment in secondary school. *Proceedings of CERME 3. Third Conference of the European Society for Research in Mathematics Education*. Belaria, Italy.

Azcárate, P., Serradó, A., y Cardeñoso, J.M. (2004). Obstáculos en el aprendizaje del conocimiento probabilístico: la noción de azar y aleatoriedad. *Comunicación presentada en el XII Congreso de Enseñanza y Aprendizaje de las Matemáticas*. Huelva, España.

Batanero, C., y Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558-567.

Bennett, D. J. (1993). *The development of the mathematical concept of randomness; educational implications.* Doctoral dissertation, New York University. Dissertation Abstracts International, 54, 449A.

Bennett, D. J. (2000). *Aleatoriedad*. Madrid: Alianza Editorial.

Brousseau, G. (1983). Les obstacles épistemologiques et les problémes en mathématiques. *Recherches en Didactique des Mathématiques*, *4.2*, 165-198.

Cardeñoso, J.M. (2001). *Las creencias y conocimientos de los profesores de primaria andaluces sobre la matemática escolar. Modelización de concepciones sobre la aleatoriedad y probabilidad.* Tesis Doctoral. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.

Decreto 106/1992 de 9 de junio por el que se establecen las enseñanzas correspondientes a la Educación Secundaria Obligatoria (BOJA nº 56 de 20 de junio de 1992). [Online: http://www.juntadeandalucia.es/educacion]

Falk, R., y Konold, C. (1992). The psychology of learning probability. En F. S. Gordon y S. P. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 151-164). Washington, DC: Mathematical Association of America.

Fischbein, E., Nello, M. S., y Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*, *22*, 523-549.

Hacking, I. (1975). *The emergence of probability.* Cambridge: University Press.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, *6*, 187-205.

Henry, M. (2001). Notion d'experiénce aleatoire. Vocabularie et modèle probabiliste. En M. Henry (Ed.), *Autour de la modélisation en probabilités* (pp. 173-186). Paris: Presses Universitarires de France.

Hoeman, H. W., y Ross, B. M. (1982). Children's concepts of chance and probability. En C. J. Brainerd (Ed.), *Children's logical and mathematical cognition*. Nueva York: Springer Verlag.

Inhelder, B., y Piaget, J. (1985). *De la lógica del niño a la lógica del adolescente*. Barcelona: Paidós.

Kahneman, D., Slovic, P., y Tversky, A. (Eds.) (1982). *Judgments under uncertainty: heuristics and biases.* Nueva York: Cambridge University Press.

Kelly, I. W., y Zwiers, F. W. (1988). Mutually exclusive and independence: Unravelling basic misconceptions in probability theory. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 1-26). Victoria, BC: International Statistical Institute.

Kolmogorov, A.N. (1950). *Foundations of Theory of Probability*. New York: Chelsea.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1).
[Online: http://www.amstat.org/publications/jse/v3n1/konold.html]

Konold, C., Pollatsek, A., Well, A.D., Lohmeier, J., y Lipson, A. (1993). Inconsistencies in student's reasoning about probability. *Journal for Research in Mathematics Education, 24*(5), 392-414.

Kyburg, H. E. (1974). *The logical foundations of statistical inference*. Boston: Reidel.

Lecoutre, M. P., y Duran, J. L. (1988). Judgements probabilisties et modèles cognitifs: étude d'une situation aléatorie. *Educational Studies in Mathematics*, *19*, 357-368

Malara, N. (1989): *Probabilitá e statistica nella scuola media. Analisi di alcuni libri di testo*. [Probability and statistics in secondary school. Analysing some textbooks]. Modena, Italia: Universitá di Modena.

Meletiou, M., y Stylianou, D. (2003). On the formalist view of mathematics: impact on statistics instruction and learning. *Proceedings of CERME 3. Third Conference of the European Society for Research in Mathematics Education*. Belaria, Italy.

Ortiz, J. J. (1999). *Significados de los conceptos probabilísticos en los libros de texto de Bachillerato.* Tesis doctoral inédita. Universidad de Granada.

Ortiz, J. J., y Serrano, L. (2001). Reflexiones sobre el lenguaje probabilístico en los libros de texto de Educación Secundaria. Comunicación presentada en las *Jornadas Europeas de Estadística.* Mallorca; Instituto Balear de Estadística.
[Online: http://www.caib.es/ibae/esdeveniments/jornades_10_01]

Pimm, D. (1987). *Speaking mathematically*. New York: Routledge&Kegan Paul.

Poincaré, J. H. (1979). El azar. En J. Newman (Comp). *Sigma. El mundo de las Matemáticas,* (Vol III, pp. 68-82). Barcelona: Grijalbo.

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

Serradó, A. (2003). *El tratamiento del azar en educación secundaria obligatoria*, Doctoral Dissertation, University Microfilms Incorporated's Proquest Digital Dissertations, Michigan. AAT 3126908 Number.

Serradó, A., y Azcárate, P. (2003). Estructura de las unidades dedicadas al "Tratamiento del Azar" en los libros de texto de Educación Secundaria Obligatoria. *Educación Matemática*, *15*(1), 67-98.

Serradó, A., Azcárate, P., y Cardeñoso, J.M (2005). Randomness in textbooks: the influence of deterministic thinking. In M. Bosch (Ed.), *Proceedings of CERME 4: Fourth Conference of the European Society for Research in Mathematics Education*. Barcelona, Spain: Ramon Llull University.

Truran, J. M. (1994). Childen's understanding of random generators. In J. Garfield (Ed.), *Research Papers from the Fourth International Conference on Teaching Statistics*. University of Minnesota.

Truran, J. M., y Truran, K. M. (1996). Statistical independence - one concept or two? Implications for research for classroom practice. Comunicación presentada en el *Eighth International Conference on Mathematics Education (ICME 8).* Sevilla.

Tversky, A., y Kahneman, D. (1982). Judgements under uncertainty: Heuristics and biases. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgments under Uncertainty: Heuristics and Biases* (pp. 3-22). New York: Cambridge University Press.

PILAR AZCÁRATE GODED
Verderón, 9
11100 San Fernando
Cádiz, Spain

APÉNDICE

Editorial Bruño:

Miñano, A., y Ródenas, J. A. (1998). *Matemáticas 1º.* Madrid: Editorial Bruño.

Miñano, A., y Ródenas, J. A. (1998). *Matemáticas 2º.* Madrid: Editorial Bruño.

Miñano, A., y Ródenas, J. A. (1998). *Matemáticas 3º.* Madrid: Editorial Bruño.

Miñano, A., y Ródenas, J. A. (1998). *Matemáticas 4º A.* Madrid: Editorial Bruño.

Miñano, A., y Ródenas, J. A. (1998). *Matemáticas 4º B.* Madrid: Editorial Bruño.

Editorial Santillana:

Almodovar, J. A., Corbalán, F., García. P., Gil, J., y Nortes, A. (1999). *Matemáticas. Curso 1º ESO; Matemáticas. Curso 2º ESO.* Madrid: Grupo Santillana de Ediciones.

Almodovar, J.A., García. P., Gil, J., y Vázquez, C. (1999). *Matemáticas. Curso 3º ESO.* Madrid: Grupo Santillana de Ediciones.

Almodovar, J.A., Gil, J., y Nortes, A. (1998). *Matemáticas Opción A. Curso 4º ESO.* Madrid: Grupo Santillana de Ediciones.

Almodovar, J.A., García. P., Gil, J., y Vázquez, C. (1999). *Matemáticas Opción B. Curso 4º ESO.* Madrid: Grupo Santillana de Ediciones.

Editorial Guadiel:

Guasch, M., Merino, R., Solsona, J., y Equipo Guadiel (1996). *Matemáticas 1*. Sevilla: Guadiel-Grupo Edebé.

Fuster, M., Martín Hernández, F., y Equipo Guadiel (1997). *Matemáticas 2*. Sevilla: Guadiel-Grupo Edebé.

Doménech, M. A., Doménech, M., Jimeno, M., Morató, M. A., Suñé, M. M., Tomás, J., y Equipo Guadiel (1995). *Matemáticas 3.* Sevilla: Guadiel-Grupo Edebé.

Fuster, M., Jimeno, M., Martín, F., Martínez, E., Morató, M. A., Tomás, J., y Equipo Guadiel (1996). *Matemáticas 4 (B).* Sevilla: Guadiel-Grupo Edebé.

Editorial Mc Graw Hill:

Pancorbo, L., Becerra, M. V., Martínez, R., y Rodríguez, R. (1995). *Matemáticas 1.* Madrid: McGraw-Hill/ Interamericana de España, S.A.U.

Becerra, M. V., Martínez, R., Pancorbo, L., y Rodríguez, R. (1996). *Matemáticas 2.* Madrid: McGraw-Hill / Interamericana de España, S.A.U.

Amigo, C., Peña, P., Pérez, A., Rodríguez, A., Sivit, F., Asencio, Mª. J., y Vicente, E. (1996). *Matemáticas 3.* Madrid: McGraw-Hill /Interamericana de España, S.A.U.

Amigo, C., Peña, P., Pérez, A., Rodríguez, A., Sivit, F., Asencio, Mª. J., y Vicente, E. (1996). *Matemáticas 4 Opción A.* Madrid: McGraw-Hill /I Interamericana de España, S.A.U.

Amigo, C., Peña, P., Pérez, A., Rodríguez, A., Sivit, F., Asencio, Mª. J., y Vicente, E. (1997). *Matemáticas 4 Opción B.* Madrid: McGraw-Hill / Interamericana de España, S.A.U.

TROIS PROBLÈMES SEMBLABLES DE MOYENNE PAS SI SEMBLABLES QUE ÇA ! L'INFLUENCE DE LA STRUCTURE D'UN PROBLÈME SUR LES RÉPONSES DES ÉLÈVES

THREE SIMILAR MEAN PROBLEMS: ARE THEY REALLY THAT SIMILAR? RESEARCH ON THE INFLUENCE OF THE STRUCTURE OF THE PROBLEM ON STUDENTS' RESPONSES

CLAUDINE MARY
Université de Sherbrooke
Claudine.Mary@Usherbrooke.ca

LINDA GATTUSO
Université du Québec à Montréal
gattuso.linda@uqam.ca

RÉSUMÉ

Les résultats sont tirés d'une étude plus large sur les stratégies de résolution que des élèves de 2^e^, 3^e^ et 4^e^ secondaires (14 à 16 ans) utilisent pour résoudre des problèmes de moyenne. Dans ce texte, les résolutions de trois problèmes seront analysées. Ces problèmes ont été composés de telle sorte que nous puissions distinguer entre la capacité des élèves de calculer une moyenne et celle de saisir les liens qu'il y a entre la modification de l'effectif et d'une donnée et celle de la moyenne. Les problèmes visaient aussi à tester l'influence d'une donnée égale à zéro, influence signalée dans les études précédentes. Les résultats de la présente étude font voir que, dans le contexte choisi, le type et le sens des modifications effectuées influencent les comportements des élèves et que des conceptions inadéquates ou des glissements de sens apparaissent dans certaines situations et pas dans d'autres. Note: Un long resume en anglais precede l'article qui est ecrit en francais.

Mots-clés: *Recherche en enseignement de la statistique; Moyenne arithmétique; Enseignement secondaire*

ABSTRACT

The results are taken from a much larger study on the strategies that pupils in the 2nd, 3rd and 4th stages at secondary school (ages 14-16) use for solving problems concerning the mean. In this paper the solutions of three problems are analysed. These problems have been formulated to be of such a kind that we can distinguish between the ability of pupils to calculate a mean, and that of realising the effect of a change in the number of observations or in the value of an observation, on the mean. The problems were also seen to test the influence of a value equal to zero on the

Statistics Education Research Journal, 4(2), 82-102, http://www.stat.auckland.ac.nz/serj

mean, drawn attention to in earlier research studies. The results of the current study show us, in the chosen context, the type and sense of the modifications exerting influence on the manipulations of the pupils, and that inadequate conceptions or a change of meaning appeared in certain situations and not in others. Note: An extended summary in English is provided at the beginning of this paper, which is written in French.

Keywords: *Statistics education research; Arithmetic mean; High school*

EXTENDED SUMMARY

The results presented in this paper are part of a larger study on the strategies used to solve problems of averages by children through their high-school years. In this study we focus on the understanding of the links between a data set and the mean, to see if the pupils perceive that changing even one of the data values affects the mean. We examine also the difficulties encountered by the introduction of a data value equal to zero.

A survey of previous research shows that most pupils and students can understand the mean as a computational construct, but have more difficulty seeing it as a representative value (Cai, 1995; Garfield & Ahlgren, 1988; Gattuso & Mary, 1996; Lappan & Zawojewski, 1988; Leon & Zawojewski, 1990; Pollatsek, Lima & Well, 1981; Strauss & Bichler, 1988) However, some improvement of comprehension with age is observed (Watson & Moritz, 1999, 2000).

Studies on the conceptions of average reveal that existing conceptions (or misconceptions) of representativeness may interfere with the ones introduced in class. For example, some children confuse mode, median and mean (Mokros & Russell, 1995; Johnson, 1985 (see Garfield & Ahlgren, 1988). Watson and Moritz (1999, 2000) show that the concept of mean is far more difficult than mode or median. In the case where one value is zero, pupils or students have problems taking into account the zero value in the computation of the average. Some consider it as the neutral element and assert that it does not change the mean; this fact seems to persist with age (Mevarech, 1983; Strauss & Bichler, 1988).

In this paper, we will analyze in detail the students' responses to three problems that have the same formulation and context but are different as regards the type and direction of the modification of a value equal to zero.

The tasks were framed as three different situations involving modifying the data set by: 1) adding a value equal to zero, 2) replacing a value equal to zero by another and 3) removing one value equal to zero. Two questions were asked: a) Does the average increase, remain unchanged or decrease? b) What is this average? Even though the first question calls for a qualitative answer and the second one does not require complex computations, the numbers were chosen so as to make the calculations easy if ever the pupil found the need to do some.

The three problems:

(1) Adding one value of 0

A group of seven friends empty their pockets. They have an average of $12 each. Jean-Philippe joins the group and does not have a sou, what happens with the average?

(2) Replacing one value by 0

A group of eight friends empty their pockets. They have an average of $11 each. Peter decides to take back his money and go to work. At the same time, Jean-Philippe joins the group but he does not have a sou, what happens to the average?

(3) Removing one value of 0

A group of eight friends empty their pockets. They have an average of $12 each. Peter decides to go to work and leave the group. Anyway he did not have a sou. What happens to the average now that he has left the group?

Approximately one hundred students answered each question (40% grade 8, 40% grade 9 and 20% grade 10). They were 14 to 16 years old.

RESULTS

A good proportion of the students did anticipate the effect of the modification on the mean and they used various procedures. When asked to anticipate the effect of adding or removing a value of zero, the students who answered correctly concentrated on the change in the number of values, while when answering the question where a value different from zero is replaced by another equal to zero, they tended to look at the total. When they are asked to calculate the mean (b) they mostly use the formula.

On the other hand, the incorrect responses show different procedures and different degrees of difficulty in the problems. In problems 2 and 3, the relation between data, the number of values and the mean seems to be difficult to rebuild. Problem 3 was more difficult than the others and exposed misconceptions like “zero does not change anything” and “everyone has $12.” Considering an equal sharing may be useful in certain situations but it obviously caused a shift of sense here - $12 becomes the amount each person has, and not a made-up amount as a result of a supposed equal sharing. These conceptions don’t appear in the answers to problem 1. In problem 2, some confusion between the total and the mean was found.

Misconceptions and shifts (of sense) seem to decrease with age. However, the older students are not better at evaluating the mean when the information is more complicated and seems to be overcome by the grade 8 students. When they are asked to calculate the new mean in version (2) Replacing one value by 0, the grade 8 students (47,4%) succeed better that the grade 9 (30,77%) and even grade 10 (38,46%).

Our results emphasize the difference in behavior following the grade and the problems. These findings suggest that diverse strategies should be confronted so as to explore the link between some more primitive but adequate strategies and other more sophisticated ones. It is necessary to vary the structures of the problems and in this, problems 2 and 3 are particularly interesting for evaluating the understanding. Further research should compare these findings with others in different contexts and different values.

TROIS PROBLÈMES SEMBLABLES DE MOYENNE PAS SI SEMBLABLES QUE ÇA ! L’INFLUENCE DE LA STRUCTURE D’UN PROBLÈME SUR LES RÉPONSES DES ÉLÈVES

1. INTRODUCTION

La moyenne arithmétique est un concept très fréquemment utilisé, depuis longtemps, et non seulement dans le contexte de la statistique. On en retrace la source dans les essais des Babyloniens (500-300 AC) qui prenaient plusieurs mesures du même phénomène. Cependant, ce n’est pas avant les travaux de Tycho Brahé (16e siècle) que l’on voit clairement apparaître la moyenne (Lavoie & Gattuso, 1998). Toutefois, et plusieurs études le montrent, encore aujourd'hui, les élèves ne réussissent pas facilement à maîtriser ce concept.

Cette situation semble vraie à tout âge. Leon et Zawojewski (1990) ont établi que la plupart des élèves peuvent comprendre le calcul de la moyenne mais ont plus de difficultés à la voir comme une valeur représentative. Les enfants ont au point de départ une assez bonne représentation de la moyenne mais elle ne semble pas se développer en une compréhension profonde du concept et les difficultés apparaissent quand le problème demande autre chose qu'un simple calcul de moyenne. Cai (1995) souligne que bien qu'environ 90% d'un échantillon d'élèves de sixième année ait réussi à calculer une moyenne, seulement la moitié a fait preuve de ce que Cai a appelé une compréhension conceptuelle. Dans une situation où l'on cherchait une donnée, la moyenne et les autres données étant connues, ces sujets tentaient d'appliquer directement l'algorithme de calcul au lieu de l'inverser. Selon Gattuso et Mary (1996) ce phénomène se retrouve également chez des élèves plus âgés.

Ces résultats s'accordent avec ceux obtenus par Pollatsek, Lima et Well (1981), avec des étudiants de collège. Parmi 37 collégiens, seulement quatorze ont pu calculer un problème de moyenne pondérée. Les résultats de plusieurs études (Garfield & Ahlgren, 1988; Lappan & Zawojewski, 1988) confirment qu'il y a une prédominance du numérique, de la procédure algorithmique de calcul de la « moyenne, les élèves et même les adultes ayant tendance à répondre en procédant par un processus de broyage de nombres », plaçant des quantités dans une formule sans vraiment saisir le concept ou produire un raisonnement juste. Pollatsek suggère de présenter les problèmes dans différents contextes et différents formats alors que d'autres suggèrent de travailler l'analogie du modèle de balance (Hardiman, Well & Pollatsek, 1984; Mokros & Russell, 1995).

D'autre part, certaines études se sont concentrées sur les propriétés que l'on attribue à la moyenne. Mevarech (1983) relève que les collégiens attribuent à la moyenne les propriétés des opérations arithmétiques (associativité, commutativité, élément neutre, etc.). De plus, le zéro comme dans le cas de la division par 0 et dans le cas de l'exposant 0 semble poser ici encore problème. Si une des données est égale à 0, certains élèves la traitent comme un élément neutre et disent que la moyenne ne change pas. Une fois encore, ce fait se manifeste a differents ages. Strauss et Bichler (1988) dans une étude menée auprès d'enfants de 8 à 14 ans ont noté qu'en plus de ne pas comprendre que la somme des écarts à la moyenne est nulle et que la moyenne est représentative de l'ensemble des valeurs, les élèves ont de la difficulté à tenir compte du zéro dans le calcul de la moyenne.

Dans le cas de la moyenne, on ne peut pas dire, comme pour le score centré réduit $\left(Z = \frac{(x-\mu)}{\sigma}\right)$ ou la variance, que ce soit la nouveauté du concept qui en explique la difficulté. Au contraire, des conceptions sont déjà présentes, et ce avant tout enseignement, et peuvent interférer avec les concepts qu'on cherche à introduire. On confond souvent la moyenne avec le mode, la valeur la plus fréquente, ou encore on choisit ce qui paraît être la valeur du milieu, une sorte de médiane. Mokros et Russell (1995) dans une étude menée auprès de jeunes enfants de 4^e , 6^e et 8^e années et de leurs professeurs ont décelé chez les sujets cinq sens de la moyenne *(mean)*: le mode, un algorithme, une valeur raisonnable, un point milieu et un point d'équilibre. D'autres encore voient la moyenne comme une valeur habituelle ou typique (Johnson, 1985 voir Garfield & Ahlgren, 1988). Ces conceptions semblent répandues, persistantes et difficiles à changer. À la suite de l'administration de questionnaires et d'entrevues Watson et

Moritz (1999, 2000) proposent un modèle de développement à plusieurs niveaux (Tableau 1) des concepts associés aux mesures de tendance centrale (moyenne, médiane et mode). Ils montrent que le concept de moyenne est de loin le plus difficile. Les concepts de médiane et mode peuvent eux s'appuyer sur les idées intuitives de « *milieu »* et de « *le plus souvent», « le plus courant ».* Ces idées apparaissent fréquemment chez les élèves pour interpréter un énoncé où une mesure de tendance centrale est donnée ou suggérée. Par ailleurs, la flexibilité dans l'utilisation de l'algorithme de moyenne et l'idée de la moyenne comme mesure représentative d'un ensemble de données se développent au fur et à mesure de la scolarité et selon les niveaux proposés en hypothèse. Ainsi une étude longitudinale (Watson & Moritz, 2000) réalisée sur une quarantaine de sujets montre que la plupart des élèves y arrivent. Toutefois, dans l'étude précédente (Watson & Moritz, 1999), pour un problème portant spécifiquement sur la moyenne, seulement 37% des élèves de 11e année (n=164) ont démontré une compréhension relationnelle des différents aspects associés à la mesure et jugés essentiels par les auteurs pour la résolution de problèmes plus complexes (avec inférences). Lors de cette étude sur la compréhension des mesures de tendance centrale, le pourcentage d'élèves de 11^{e} année qui ont atteint un niveau supérieur (R) ne dépasse pas 54%. Ceci fait dire aux auteurs qu'une grande partie des élèves n'ont pas construit l'idée de mesure représentative d'un ensemble de données et n'ont donc pas atteint le niveau suffisant pour passer à des problèmes plus complexes.

Tableau 1. Modèle de développement des concepts de mesures de tendance centrale selon Watson et Moritz (1999)

Niveau	Description en bref
P-*préstructural responses*	les élèves ont vraisemblablement entendu le mot mais n'y accordent pas de signification spécifique
U-*unistructured responses*	une idée de ce que signifie le mot mais sans référence à un ensemble de données
M-*multistructural responses*	une description présentant plusieurs aspects pertinents mais des incohérences subsistent
R-*relational responses*	une description montrant une compréhension des relations entre les différents aspects associés à la mesure, construction de l'idée de mesure représentative d'un ensemble de données

Dans l'étude que nous présentons, nous nous attardons spécifiquement à la relation que les élèves établissent entre les données et la moyenne et dans ce but, nous interrogeons les élèves sur les effets que produit sur la moyenne la modification d'une donnée ou de l'effectif. Avant de présenter le contexte de l'étude et ses objectifs particuliers, il paraît important de dire quelques mots sur la formulation des questions. Plusieurs chercheurs ont montré l'influence du contexte, de l'histoire (dans un problème à texte) ou de la formulation sur le comportement des élèves (Stern & Lehmdorfer, 1992; Cerquetti-Aberkane, 1987). Certains contextes vont donner une signification particulière à l'opération (partage ou groupement pour la division, par exemple) (Bell, Fischbein & Greer, 1984; Semadeni, 1984). L'influence du contexte, notamment des grandeurs impliquées, se fait sentir également dans les problèmes de moyenne (Mary & Gattuso, 2003; Pollatsek, Lima & Well, 1981). Au-delà de l'histoire qui accompagne un problème ou de sa formulation, ce sont les relations impliquées dans le problème, la nature des données et des nombres, la place de l'inconnue dans une équation, etc. qui vont rendre plus ou moins complexe la résolution d'un problème. En particulier, Vergnaud (1996) a montré que des problèmes à structure semblable sur le plan mathématique, donnant lieu à la même équation par exemple, sont traitées de façon très différente par les élèves. Ainsi, dans le champ conceptuel des problèmes à structure additive (Vergnaud, 1996), il identifie des classes de problèmes auxquels sont associés des schèmes de fonctionnement, invariants opératoires pour cette classe de problèmes.

Dans notre étude, nous avons choisi de proposer trois problèmes semblables de moyenne aux élèves, semblables de par la formulation, le contexte mais aussi dans la mesure où l'équation qui leur était associée s'établissait de façon similaire: total initial auquel on ajoute ou retranche une donnée pour diviser le résultat par le nouvel effectif. Certaines expériences individuelles antérieures nous laissaient penser que les élèves ne traitaient pas l'ajout d'une donnée, par exemple, de la même façon que le retrait d'une donnée. C'est ce que confirment les résultats présentés.

2. CONTEXTE DE L'ETUDE ET OBJECTIFS

Les résultats qui seront présentés sont issus d'une recherche plus large sur les stratégies de résolution de problèmes de moyenne. Six cent trente huit élèves de 2^e^, 3^e^ et 4^e^ secondaires (14 à 16 ans) d'écoles secondaires de la région de Montréal ont participé à cette étude. Le choix s'est fait selon la disponibilité des enseignants. Ils répondaient à 5, 6 ou 7 questions respectivement selon leur niveau (Gattuso & Mary, 1998, 2001). Il y avait six versions du questionnaire de sorte que tous ne répondaient pas aux mêmes questions, mais les questions étaient toujours présentées dans le même ordre.

Le questionnaire a été présenté par l'enseignant aux élèves qui avaient environ une heure pour y répondre. Il y avait une seule question sur chaque page et les consignes étaient de ne pas revenir en arrière sur les problèmes déjà faits et de ne pas utiliser la calculatrice. Les enseignants étaient avertis de ne donner aucune réponse aux élèves au sujet des problèmes. Dans les groupes de 3^e^ secondaire, les élèves avaient vu la moyenne en classe comme le prévoit le programme scolaire.

Dans cet article, les trois problèmes présentés ont été conçus avec l'intention spécifique de vérifier la capacité que les élèves ont de calculer une moyenne et celle de saisir les liens qu'il y a entre la modification d'une donnée ou de l'effectif et celle de la moyenne. Ils visaient aussi à tester l'influence d'une donnée égale à zéro, influence signalée dans certaines études précédentes. Les résultats de la présente étude font voir comment, dans le contexte des problèmes donnés, le type et le sens des modifications influencent les comportements des élèves et comment des conceptions inadéquates ou des glissements de sens apparaissent dans certaines situations et pas dans d'autres. Soulignons que chaque élève ne répondait qu'à un seule des trois problèmes.

3. LES PROBLEMES

Les trois problèmes (Tableau 2) sont formulés semblablement en reprenant la même histoire: des amis qui veulent faire une sortie se vident les poches. Dans les trois cas, la question porte sur le lien entre la moyenne et la modification effectuée. Toutefois, le type et le sens de cette modification varient d'un problème à l'autre.

L'énoncé précise combien les amis ont en moyenne pour la sortie. Puis, une modification de l'effectif et des données survient. Une première question (a) demande d'anticiper l'effet de cette modification sur la moyenne. Trois alternatives sont proposées aux élèves: 1) Elle diminue, 2) Elle demeure inchangée, 3) Elle augmente. Ensuite, que l'élève pense que la moyenne change ou non, une deuxième question (b) lui demande d'évaluer la nouvelle moyenne. Pour chacune des questions posées, l'élève devait expliquer comment ou par quel raisonnement il trouvait sa réponse. Dans les trois problèmes, une donnée égale à zéro est impliquée. Les trois problèmes sont semblables mais, ils sont aussi différents si l'on considère le type et le sens des modifications qui surviennent. Pour le problème 1, on ajoute une donnée égale à zéro car un ami se joint au groupe mais il n'a pas d'argent; donc l'effectif augmente mais le montant total ne change pas. Dans le problème 2, l'effectif reste le même (un ami part et un autre arrive), une donnée non nulle est remplacée par une autre égale à zéro; ce changement entraîne une modification du montant total. Dans le problème 3, on enlève une donnée égale à zéro car un ami part, mais comme il n'a pas d'argent, le montant total reste inchangé. Comme nous allons le montrer, ces différences entraînent des comportements différents chez les élèves.

Tableau 2. Les trois problèmes

(1) Ajouter une donnée = 0

Un groupe de sept amis voulant faire une sortie se vident les poches. Ils ont en moyenne 12$ chacun. Jean-Philippe se joint au groupe et n'a pas un sou, qu'arrive-t-il maintenant à la moyenne?

a) 1) Elle diminue 2) Elle demeure inchangée 3) Elle augmente
Écris pourquoi tu obtiens cette réponse.

b) Quelle est-elle?
Écris comment tu obtiens ce résultat.

(2) Remplacer une donnée = 0

Un groupe de huit amis voulant faire une sortie se vident les poches. Ils ont en moyenne 11$ chacun. Pierre après avoir réfléchi reprend les 4$ qu'il avait donné car il doit aller travailler. À ce moment, Jean-Philippe se joint à eux pour la sortie mais il n'a pas un sou, qu'arrive-t-il maintenant à la moyenne?

a) 1) Elle diminue 2) Elle demeure inchangée 3) Elle augmente
Écris pourquoi tu obtiens cette réponse.

b) Quelle est-elle?
Écris comment tu obtiens ce résultat.

(3) Enlever une donnée = 0

Un groupe de neuf amis voulant faire une sortie se vident les poches. Ils ont en moyenne 12$ chacun. Pierre après avoir réfléchi dit qu'il doit aller travailler et quitte le groupe. De toutes façons, il n'avait pas un sou. Qu'arrive-t-il à la moyenne maintenant qu'il a quitté le groupe?

a) 1) Elle diminue 2) Elle demeure inchangée 3) Elle augmente
Écris pourquoi tu obtiens cette réponse.

b) Quelle est-elle?
Écris comment tu obtiens ce résultat.

Nous allons maintenant présenter l'analyse de chacun des problèmes plus en détail ainsi que les résultats pour la question (a), anticipation de l'effet sur la moyenne, et ceux pour la question (b), évaluation de la nouvelle moyenne. Précisons d'abord que 310 élèves ont participé à l'expérimentation, que pour chacune de ces questions, environ cent élèves ont été interrogés et que dans chaque cas on retrouve environ 40% d'élèves de 2e secondaire, 40% d'élèves de 3e secondaire et 20% d'élèves de 4e secondaire. Le tableau 3 présente le nombre d'élèves ayant répondu à chaque problème selon le niveau académique.

Tableau 3. Nombre d'élèves ayant répondu à chaque problème selon le niveau académique

Niveau	(1) Ajouter une donnée = 0	(2) Remplacer une donnée = 0	(3) Enlever une donnée = 0	Total
sec. 2	40	39	39	118
sec. 3	39	39	41	119
sec. 4	22	26	25	73
TOTAL	101	104	105	310

4. ANTICIPATION DE L'EFFET SUR LA MOYENNE (question a)

4.1. ANALYSE DES PROBLEMES A PRIORI

Dans le problème 1, la modification consiste à ajouter une donnée égale à 0: « Jean-Philippe se joint au groupe et n'a pas un sou ». On a alors plus de personnes pour le même montant d'argent ce qui entraîne une diminution de la moyenne. Dans le deuxième problème, la modification consiste à remplacer une donnée non nulle par une autre égale à zéro: « Pierre après avoir réfléchi reprend les 4$ qu'il avait donné car il doit aller travailler. À ce moment, Jean-Philippe se joint à eux pour la sortie mais il n'a pas un sou ». L'effectif ne change pas; seul le total change; il diminue et par conséquent la moyenne aussi. Dans le troisième problème, la modification consiste à enlever une donnée égale à zéro: «Pierre après avoir réfléchi dit qu'il doit aller travailler et quitte le groupe. De toutes façons, il n'avait pas un sou». On a alors moins de personnes pour le même montant d'argent, la moyenne augmente.

Dans les problèmes 1 et 3, les variations de la moyenne sont en sens inverses des effectifs. En effet, puisqu'on ajoute (1) ou enlève (3) une donnée égale à zéro, le total est inchangé. L'effectif augmentant (1) ou diminuant (3) provoque alors respectivement une diminution (1) ou une augmentation (3) de la moyenne. Pour résoudre ces problèmes, aucun calcul n'est nécessaire. De plus, un raisonnement sur l'effectif exclusivement, permet de donner une bonne réponse puisque la donnée ajoutée ou supprimée égale zéro.

Dans le problème 2, il y a deux changements qui surviennent. Envisager l'effet sur la moyenne après chaque changement peut être compliqué: une personne part avec 4$, un montant inférieur à la moyenne (11$), donc la nouvelle moyenne augmente; puis une autre personne arrive sans argent, la moyenne baisse; la question qui se pose alors est: est-ce que la nouvelle moyenne est supérieure ou inférieure à l'ancienne moyenne (11$)? Estimer le résultat de la hausse de moyenne suivie de la baisse demande un calcul des moyennes à chaque étape à moins qu'on ne raisonne sur l'effet que produit la *combinaison* des deux modifications (-4 + 0 = -4) sur le total. Dans ce cas comme le total est diminué de 4$ et que l'effectif reste le même, on a alors moins d'argent pour le même nombre de personnes et la moyenne diminue.

Dans tous les cas, on peut répondre sans avoir besoin d'utiliser la formule classique pour calculer la nouvelle moyenne: somme des données divisée par l'effectif. Notons que la différence entre les nombres utilisés dans chacun des problèmes n'est pas très grande; les nombres ont été choisis de manière à simplifier les calculs éventuels.

4.2. RESULTATS

Les résultats pour chacun des problèmes sont présentés dans le tableau 4.

Tableau 4. Réussite aux trois problèmes (question a)

Niveau	(1) Ajouter une donnée = 0	(2) Remplacer une donnée = 0	(3) Enlever une donnée = 0	Total
	n (%)	n(%)	n(%)	n(%)
sec. 2	36 (90,00)	28 (71,79)	23 (58,97)	87 (73,73)
sec. 3	35 (89,74)	31 (79,49)	26 (63,41)	92 (77,31)
sec. 4	20 (90,91)	16 (61,54)	19 (76,00)	55 (75,34)
TOTAL	91 (90,1)	75 (72,12)	68 (64,76)	234 (75,48)

Les résultats montrent i) que les élèves réussissent relativement bien à anticiper le changement sur la moyenne, avec au-dessus de 74% de réussite, quel que soit le niveau; ii) que la réussite diffère toutefois selon les versions. Le problème 1 apparaît beaucoup mieux réussi (90% de réussite) que le problème 3 (65% de réussite) (test de différence de proportions: z=4,583; p=0,000). L'impact sur la moyenne de l'ajout d'une donnée égale à zéro apparaît donc plus facile à envisager que celui du retrait d'une donnée égale à zéro. De même, l'ajout d'une donnée égale à zéro se révèle plus facile que le remplacement d'une donnée différente de zéro par une autre égale à zéro (z=3,388; p=0,001)). Pour chacun des problèmes, les résultats ne se distinguent pas selon le niveau de scolarité.

4.3. EFFET SUR LA MOYENNE: EXAMEN DES STRATEGIES UTILISEES PAR LES ELEVES

Procédures de réussite Pour le problème 1 (ajout d'une donnée égale à 0), les élèves qui réussissent raisonnent, comme prévu, majoritairement sur l'effectif qui change: il y a le même montant d'argent divisé en plus de personnes (53,5% des élèves).

C'est le cas aussi pour le problème 3 (retrait d'une donnée égale à 0), il y a autant d'argent à répartir en moins de personnes ou simplement il faut diviser par 8 au lieu de par 9 (41,0% des élèves).

Quant au problème 2 (remplacement d'une donnée par 0) les élèves qui réussissent raisonnent majoritairement sur le total qui diminue en mentionnant ou non que l'effectif reste le même (56,8% des élèves): il y a le même nombre de personnes mais moins d'argent au total ou simplement, le total diminue. Une majorité d'élèves semble donc avoir combiné les effets des deux transformations sur le total.

Ces problèmes permettent de centrer l'attention tantôt sur le total, tantôt sur l'effectif, puisqu'une seule de ces variables change.

Procédures dans le cas de non-réussite Considérons maintenant les procédures les plus utilisées dans le cas où les élèves n'ont pas réussi les problèmes. Elles sont présentées dans le tableau 5. Il s'agit de procédures fausses à moins que l'élève n'ait pas répondu ou n'ait pas donné d'explication. Les pourcentages expriment le rapport du nombre d'élèves ayant donnée la réponse sur l'ensemble des élèves à qui a été soumis le problème à chacun des niveaux et au total.

Tableau 5. Procédures de non réussite (question a)

Effet sur la moyenne	(1) Ajouter une donnée = 0	(2) Remplacer une donnée = 0	(3) Enlever une donnée = 0
	n(%)	n(%)	n(%)
Procédure de non-réussite la plus fréquente	Rien à signaler	Aucune réponse ou réponse sans explication Sec. 2: 5 (12,8) Sec. 3: 4 (10,3) Sec. 4: 4 (15,4) Total: 13 (12,5)	Conceptions inadéquates Sec. 2: 14 (35,9) Sec. 3: 9 (22,0) Sec. 4: 3 (12,0) Total: 26 (24,8)

Pour le problème 1, aucune procédure particulière n'attire l'attention, le problème ayant été bien réussi. On constate qu'au problème 2 (remplacement d'une donnée par 0), plusieurs élèves n'ont pas répondu à la question ou n'ont donné aucune explication de leur réponse (12,5%). Les autres procédures erronées pour ce problème sont variées. Pour les élèves de 4^{e} secondaire, les plus faibles (61,5% de réussite contre 79,5% et 71,8% respectivement en 3^{e} et 2^{e} secondaires, tableau 4) plusieurs n'ont rien répondu (aucune réponse) ou n'ont pas donné d'explication (15,4%), mais on trouve aussi, en pourcentage égal, une application de la formule de moyenne arithmétique avec erreur sur l'effectif (15,4% des élèves de 4^{e} secondaire).

Quant au problème 3 (retrait d'une donnée égale à 0), l'analyse des explications données par les élèves aux réponses fausses est éclairante. En effet, les réponses fausses les plus répandues (24,8% des élèves à qui a été soumis la question 3, soit 26/105), peuvent être associées clairement à une mauvaise interprétation de la moyenne que nous considérons dans le classement des réponses comme une conception inadéquate. En fait c'est 81,1% des réponses fausses du problème 3 (26 « conceptions inadéquates » sur 32 réponses fausses) qui sont dans cette catégorie. La première conception identifiée, le zéro neutre, est associée, dans un premier temps, à l'utilisation d'un schème de résolution inapproprié dans le contexte c'est-à-dire, considérer zéro comme un élément neutre, application erronée d'une propriété des groupes, qui se traduit ici par « on enlève une donnée nulle, ça ne change rien à la moyenne ». La deuxième conception, le partage égal, vient d'une généralisation abusive d'une représentation de la moyenne comme résultat d'un partage égal. Nous en donnons des exemples ci-dessous. Ces conceptions inadéquates se manifestent peu au problème 2 (2 élèves) et au problème 1 (3 élèves).

La première conception identifiée: Le Zéro neutre L'influence sur la moyenne d'une donnée égale à zéro a été étudiée dans différentes recherches et est documentée (Mevarech, 1983; Strauss & Bichler, 1988). Dans le même sens que ces auteurs, nous trouvons plusieurs élèves qui semblent penser qu'une donnée nulle n'a pas d'influence sur la moyenne. En effet, au problème 3 (retrait d'une donnée égale à zéro), les deux tiers (17/26) des élèves dont les réponses ont été classées conceptions inadéquates indiquent que la moyenne est inchangée et manifestent dans leur explication une telle conception. Les énoncés de ces élèves sont regroupés autour de l'explication suivante: « Pierre a zéro donc ça ne change rien ».

Voici des exemples d'explications d'élèves qui ont répondu que la moyenne était inchangée.

1. *C'est comme si Pierre n'existait pas.* (2^{e} secondaire)

2. *Parce que Pierre de toutes façons n'avait pas un sou.* (2^{e} secondaire)

3. *Car s'ils avaient 12$ en moyenne mais que Pierre n'avaient rien, alors ça reste 12$.* ($2^{e}$ secondaire)

4. *Puisqu'il n'avait pas un sou, il n'a pas influencé la moyenne donc lorsqu'il est parti, la moyenne n'a pas baissé ni augmenté car Pierre n'y a pas participé.* (3^{e} secondaire)

D'autres exemples, toujours pour la même classe d'énoncés, révèlent toutefois des aspects non mentionnés dans les études antérieures et pouvant interférer dans l'interprétation que les élèves font de la situation et sur leur décision. Ainsi, dans les énoncés 5 à 8, c'est comme si le fait que Pierre n'ait pas un sou l'excluait du calcul de la moyenne ou tout simplement du groupe d'amis dont on connaît la moyenne.

5. *Ici: 12$ pour chacun; parce que Pierre n'avait rien donné, donc la moyenne de 12$ chacun était pour 8 amis donc s'il s'en va ça ne change rien.* (2^{e} secondaire)

6. *Ici, Pierre il n'est pas là, il y a 8 personnes.*
   (Somme d'argent qui possède en groupe) 96 / 8 (Nombre de personnes).

   96 / 8
   12 (3^{e} secondaire)

Dans l'énoncé 5, l'élève considère que la moyenne est établie pour 8 amis puisque Pierre n'y participe pas. Dans l'énoncé 6, la somme totale est correctement établie mais elle est considérée pour 8 personnes et non pour 9. Dans l'énoncé 7 (ci-dessous), l'élève exclut Pierre du groupe de 9 amis et ne considère donc pas que Pierre participe à la moyenne donnée. C'est le cas aussi pour l'énoncé 8.

7. *Pierre n'avait pas d'argent donc ne fait pas partie des 9 amis du groupe car les 9 mettent 12$. Lui n'a pas un sou.* ($4^{e}$ secondaire)

8. *Comme Pierre n'avait pas d'argent, son montant n'a jamais été compté. Par contre sa part de la moyenne sera distribuée de façon égale pour les autres.* (2^{e} secondaire)

   moyenne générale

   L'élève pointe avec une flèche l'expression « moyenne générale ». On peut penser que l'élève a voulu signifier que le montant n'avait jamais été compté dans la moyenne générale.

Il pourrait y avoir une question de pertinence. En effet, pourquoi considérer zéro dans la moyenne? En fait, pourrait-on penser que la difficulté ne tiendrait pas tant à une généralisation d'une propriété de l'addition mais à un obstacle épistémologique de même nature que celui envisagé dans l'histoire dans le passage du « rien » à « zéro comme nombre » (Kaplan, 1999)? De plus, au-delà de la difficulté mathématique, il pourrait y avoir ici une influence du contexte social de la situation: on ne partage pas habituellement avec quelqu'un qui n'a pas un sou ou le partage s'effectue entre les personnes qui ont de l'argent: *Pierre n'a pas un sou, il n'existe pas* (6), *il ne fait pas partie du groupe* (7). La formulation du problème a pu aussi jouer dans l'exclusion de Pierre du groupe: « Pierre après avoir réfléchi dit qu'il doit aller travailler et quitte le groupe. De toutes façons, il n'avait pas un sou ». Était-il ou non membre du groupe des neuf? Les énoncés qui suivent suggèrent une autre explication à certaines réponses d'élèves.

9. *Parce que le texte dit qu'ils avaient 12$ chacun mais comme le texte dit que Pierre n'avait pas un sou cela ne change rien à la moyenne.* (4^{e} secondaire)

Dans l'énoncé 9, l'élève réinterprète la situation du problème en évacuant l'expression en moyenne; en réalité le texte dit que les amis « ont en moyenne 12$ chacun » et l'élève interprète qu' « ils avaient 12$ chacun ». Ce glissement peut expliquer divers autres énoncés classés sous l'interprétation « zéro ne change rien » comme 5, 6, 7 et 8 ci-dessus et le suivant:

10. *Pierre=0$*
9 amis = 12$* *Avec ou sans Pierre, la moyenne est de 12$/personne.
Exemple: 0, 12, 12, 12,12, 12, 12, 12, 12.

Notons que l'exemple que donne l'élève ne donne pas une moyenne de 12$ pour les 9 amis. Il en conclut tout de même que la moyenne est inchangée. Une moyenne de 12$ par ami est interprétée ici comme « chacun a 12$ » sauf Pierre évidemment qui lui n'a rien.

Nous y voyons ici l'interférence d'une représentation de la moyenne qui peut être utilisée pertinemment pour résoudre des problèmes de moyenne, celle du partage équitable. En effet, lorsque l'on dit une moyenne de 12$ par enfant, on peut comprendre que si la somme totale était répartie également entre les amis, chacun aurait 12$. Les élèves peuvent ne pas en saisir le sens conditionnel, oublier le *si*, ou simplifier à outrance la situation par incapacité à contrôler toutes les variations possibles. Watson et Moritz (2000) observent également une stratégie semblable chez des élèves capables de reconstituer le total mais incapable de construire une distribution respectant une moyenne alors que certaines données sont connues. Dans l'énoncé 9 ci-dessus, la représentation de la moyenne comme « partage équitable » entre en conflit avec le fait qu'une donnée est zéro. La gestion de ce conflit peut mener à exclure la donnée zéro.

La deuxième conception identifiée: Le partage égal Les énoncés précédents illustraient la conception « zéro ne change rien ». Dans les énoncés 9 et 10, nous en avons vu apparaître une autre « tous ont la même chose ». Dans ces cas, seul Pierre n'avait rien, il était exclut du groupe. Cependant, un certain nombre d'autres réponses fausses (8/26) sont aussi accompagnées d'explications de type « tous ont 12$ ». Les énoncés qui suivent sont caractéristiques de cette catégorie.

11. *Elle est inchangée.*

12 × 9 = 108   *108 - 12 = 96*   *96 ÷ 8 = 12*

96
- 8
16
-16
0

(3^{e} secondaire)

Le total est évalué à 9 x 12 = 108. Ensuite, le total est réduit de 12 pour devenir 96, puis est divisé en 8 puisque Pierre est parti. C'est comme si Pierre partait avec 12$.

12. *Elle diminue.*

Essai: 5 amis et chacun 12$ mais un a 0$
12 + 12 + 12 + 12 + 0 = 48 ÷ 5 ≅ 9,9
mais si 4 amis et tout le monde 12$
12 + 12 + 12 + 12 = 48 ÷ 4 ≅ 12 (2e secondaire)

Dans l'énoncé 12, l'élève considère le cas où il y a 5 amis. Il calcule d'abord la moyenne en donnant 12$ à chacune des personnes sauf à Pierre qui n'a rien, conformément à l'énoncé du problème. Ensuite, il calcule la moyenne en faisant partir Pierre. Il trouve une moyenne supérieure. L'élève aurait dû conclure que la moyenne augmentait. Toutefois, il a vraisemblablement confondu le 12$ trouvé lors du 2e calcul avec la moyenne du problème initial (qui était de 12$) et conclut qu'elle diminuait en passant à 9,9.

Les deux exemples qui précèdent illustrent un glissement. Comme nous l'avons dit précédemment, nous pouvons comprendre qu'une moyenne de 12$ correspond au montant qu'aurait chacun si le total était partagé également entre les amis. Toutefois, la distinction entre la part fictive qui résulte du partage égal (la moyenne) et une contribution réelle de chacun de 12$ est subtile! L'idée de faire partir Pierre avec 12$ est tentante. Le glissement est facile et est peut-être renforcé ici par le contexte du problème et sa formulation. Dans l'exemple suivant, dont la réponse est juste, l'élève partage un 12$ laissé par Pierre entre les 8 amis. Raisonnement juste ou erroné, la frontière est faible.

13. *Elle augmente.*

Comme il y a une personne qui s'en va, alors ils vont partager le 12$ entre les 8 autres personnes. (2e secondaire)

La formulation aussi peut induire des conceptions inadéquates. Bien que celle utilisée dans les problèmes « Ils ont en moyenne 12$ chacun» est très couramment employée dans la langue parlée et dans les manuels scolaires, une autre formulation comme « la moyenne des montants qu'ils ont mis en commun est de 12$» aurait peut-être évité certains glissements de sens. Cependant, dans l'énoncé 10, on constate que l'élève exprime bien le « ont en moyenne 12$ chacun » de l'énoncé par «la moyenne est de 12$/personne». Nonobstant cette formulation, une certaine abstraction du concept semble se développer avec l'âge. Ces interprétations inadéquates de la situation sont plus fréquentes en 2e secondaire et paraissent diminuer avec l'âge ou l'enseignement (Cf. tableau 5).

5. EVALUATION DE LA NOUVELLE MOYENNE (question b)

5.1. ANALYSE DES PROBLEMES A PRIORI

Après avoir demandé aux élèves d'anticiper si la moyenne augmentait, diminuait ou restait inchangée (question a), nous leur avons demandé quelle était la nouvelle moyenne. Ils pouvaient alors recourir à différentes stratégies, telles construire une liste de données ou reconstruire le total initial. Comme expliqué pour la partie (a), il était possible aussi de considérer que la moyenne est le résultat obtenu après partage égal et donc de reconstituer le total en multipliant le résultat du partage par l'effectif. Ce nouveau total pouvait se trouver également algébriquement en posant la somme initiale comme inconnue dans la formule de la moyenne. Une fois le total initial trouvé, il s'agit alors d'évaluer le nouveau total en ajoutant ou enlevant les données et en le redistribuant selon le nouvel effectif.

Si toutefois l'on se base sur la formule usuelle pour évaluer la moyenne, le calcul présente quelques difficultés, puisque dans ces problèmes, les données ne sont pas accessibles directement, seule la moyenne de ces données est connue. Ainsi, pour évaluer la somme des données, il faut savoir que la somme d'argent totale initiale égale le produit de la moyenne par l'effectif. Dans le premier problème, par exemple, si 7 amis ont en moyenne 12$ chacun, ils ont ensemble 7 x 12$. Pour trouver la nouvelle moyenne, il suffit alors d'ajouter ou soustraire la donnée supplémentaire ou la donnée à retirer et de diviser par le nouvel effectif. Ces problèmes peuvent également donner lieu à des résolutions algébriques en posant une équation dont l'inconnue est la somme initiale des données, avant modification de la situation.

5.2. RESULTATS

Les résultats à cette question sont présentés dans le tableau 6. Nous pouvons constater que 1) le problème 1 est encore le mieux réussi, 2) que le problème 2 est particulièrement faible (moins de 40% de réussite) et ce même en 4^e secondaire et 3) que le problème 3 est mieux réussi en 4^e secondaire. Dans tous les cas, un test de proportions donne des résultats significatifs.

Tableau 6. Réussite aux trois problèmes (question b)

Niveau	(1) Ajouter une donnée = 0	(2) Remplacer une donnée = 0	(3) Enlever une donnée = 0	Total
	n (%)	n (%)	n (%)	n (%)
sec. 2	34 (85,0)	18 (47,4)	19 (48,72)	71 (60,2)
sec. 3	29 (74,4)	12 (30,77)	17 (41,5)	58 (48,7)
sec. 4	20 (90,9)	10 (38,46)	18 (72,0)	48 (65,8)
TOTAL	83 (82,2)	40 (38,8)	54 (51,4)	177 (57,0)

5.3. CALCUL DE LA MOYENNE: EXAMEN DES STRATEGIES UTILISEES PAR LES ELEVES

Procédures de réussite Presque la totalité des procédures de réussite pour chacun de ces problèmes consiste en l'utilisation de ce que nous avons appelé la formule de la moyenne: somme des données divisée par l'effectif. Cette formule peut être appliquée pour reconstituer le total avant de poursuivre (total ÷ effectif = moyenne, donc, total = moyenne × effectif) ou après avoir construit une liste fictive de données. Toutefois, il est à remarquer que quelques élèves (7 élèves) utilisent une procédure très intéressante que nous avons appelée procédure de compensation: on ajuste la moyenne en lui ajoutant ou retranchant une partie, sans passer par le total. Ainsi, dans l'énoncé 14 qui apparaît pour le problème 2, où une donnée (4$) est remplacée par une autre nulle, l'élève enlève à la moyenne la fraction correspondant à la donnée qui part.

14. *Vu que Pierre a enlevé 4$ et qu'il y a 8 personnes dans le groupe, on diminue la moyenne de .50*

Alors 11.00
- .50
10.50 (2e secondaire)

Dans le problème 3, où une donnée nulle est enlevée, on voit apparaître une autre stratégie illustrée par les énoncés 15 et 16.

15. *Comme il y a une personne qui s'en va, alors ils vont partager le 12$ entre les 8 personnes.*

16. *12$ réparti sur 8 nouvelles personnes: 13,50$*
(4e secondaire)

Ils attribuent un montant équivalent à la moyenne à Pierre qui part et redistribue ce montant entre les personnes restantes. Cinq élèves font ce raisonnement.

On constate donc que des raisonnements différents apparaissent avec de nouveaux problèmes. De plus, l'analyse du tableau 6 montre que le taux de réussite varie selon les problèmes. La performance des élèves de 4e secondaire est intéressante sur ce plan. En effet, alors qu'ils utilisent beaucoup plus la formule que les élèves des deux autres niveaux au problème 3, ils semblent l'utiliser relativement peu dans le cas du problème 2! Les élèves de 4e secondaire réussissent mieux le problème 3 que les élèves de 2e secondaire (z=1,94; p= 0,061) et ceux de 3e secondaire (z=2,58; p=0,014). Toutefois, pour le problème 2, les élèves de 4e secondaire ne semble pas plus habiles pour évaluer la moyenne que les autres élèves. L'examen des procédures fausses utilisées peut éclairer encore ici.

Procédures dans le cas de non-réussite Dans le tableau 7, sont présentées les procédures les plus fréquentes dans le cas où les élèves n'ont pas réussi le problème. Les pourcentages expriment le rapport du nombre d'élèves ayant donné la réponse sur l'ensemble des élèves à qui a été soumis le problème à chacun des niveaux et au total.

Tableau 7. Procédures de non réussite (question b)

	(1) Ajouter une donnée = 0	(2) Remplacer une donnée = 0	(3) Enlever une donnée = 0
	n (%)	n (%)	n (%)
Procédure fausse la plus fréquente	Vides Sec. 2:3 (7,5) Sec. 3: 5 (12,8) Sec. 4: 0 (0) Total: 8 (7,9)	Fausses formules Sec. 2: 6 (15,4) Sec. 3: 8 (20,5) Sec. 4: 6 (23,1) Total: 20 (19,2)	Vides Sec. 2: 4 (10,3) Sec. 3: 10 (24,4) Sec. 4: 1 (4,0) Total: 15 (14,3)
2e procédure fausse la plus fréquente		Vides Sec. 2: 0 (0) Sec. 3: 13 (33,3) Sec. 4: 4 (15,4) Total: 17 (16,4)	Conceptions inadéquates + Total Sec. 2: 9 (23,1) Sec. 3: 5 (12,2) Sec. 4: 0 (0) Total: 8 (13,3)

Pour les trois problèmes, et surtout pour les problèmes 2 et 3, qui semblent plus difficiles, on constate beaucoup de réponses *vides* (l'élève n'a rien répondu ou dit qu'il ne sait pas).

Pour le problème 2 (remplacement d'une donnée par 0), le moins bien réussi, la procédure erronée la plus fréquente consiste en *fausses formules*. Nous avons appelé *fausse formule* tout faux calcul qui implique les données et l'effectif; la plupart du temps, il s'agit de la formule avec erreur sur l'effectif (19,2%). Ainsi au problème 2, on retrouve le nouveau total (88$ - 4$) divisé tantôt par 7, 9, 10 et 11. Même les élèves de 4e secondaire qui réussissent assez bien grâce à la formule, semblent avoir plus de difficultés avec le problème 2 (tableau 7). Eux aussi recourent en assez grand nombre à ce que nous avons appelé une fausse formule. C'est comme si, dans ce cas, ils n'arrivaient pas à tenir compte adéquatement des deux changements. Les réponses vides en témoignent également. Les deux procédures suivantes ont été classées fausses formules.

17. *7 × 11 = 77* *77 ÷ 7 = 11* *7 : 11$*
1 : 7$
1 : 0
———
9 : 84
84 ÷ 9 = 9,3$

(problème 2, 4e secondaire)

Pour les trois problèmes, et surtout pour les problèmes 2 et 3, qui semblent plus difficiles, on constate beaucoup de réponses *vides* (l'élève n'a rien répondu ou dit qu'il ne sait pas).

Dans l'énoncé 17, l'élève a considéré sept données égales à 11, une donnée égale à 7, vraisemblablement parce que le total devait être 4 de moins que l'ancien lorsqu'un ami part, puis une donnée égale à 0 pour le nouveau venu. En procédant ainsi, il se trouve à créer une donnée supplémentaire. En effet, huit personnes ont 84$ et non neuf. Douze élèves ont ainsi divisé le total de 84$ par 9. La procédure 17 pourrait expliquer cette erreur sur l'effectif. L'énoncé 18 est un autre exemple de fausse formule.

18. *Elle est de 12 car en moyenne ils avaient en tout (8 × 11) = 88$ maintenant 84 et 7 amis 84/7=12$.*

(problème 2, 4e secondaire)

C'est comme si l'élève avait oublié de considérer le nouvel arrivé, à moins que le fait qu'il n'ait rien, l'exclut du calcul, comme nous l'avons vu au problème 3 (a).

Pour le problème 3 (b) (retrait d'une donnée égale à 0), plusieurs élèves n'ont rien répondu ou ont répondu qu'ils ne savaient pas (« ? » ou « je ne sais pas »). Outre ces réponses, à la question b comme à la question a, nous retrouvons la manifestation d'une mauvaise interprétation de la moyenne (*conception inadéquate*): on considère un total de 96$ comme si le 12$ de moyenne était seulement pour les huit amis, Pierre ne changeant rien à la moyenne, puisqu'il avait zéro, ou comme si tous se trouvaient avec 12$ donc Pierre partant, il y a un 12$ de moins.

19. *Ils sont 8, Pierre étant parti.*
12$ × 8 amis = 96$ / 8 = 12$

(problème 3, 2e secondaire)

En plus des énoncés classés comme *conceptions inadéquates*, nous retrouvons un certain nombre d'énoncés classés *Total* qui révèlent aussi une mauvaise interprétation de la moyenne. Pour les problèmes 2 et 3, plusieurs élèves donnent comme moyenne la somme des données correspondant au total initial ou au total après changement. Ces réponses apparaissent surtout au problème 2 (10,6%) et, moindrement au problème 3 (5,8%). En voici un exemple.

20. *S'ils sont 8 amis et ont en moyenne 11$ chacun, donc tous ensemble auront un montant d'à peu près 88$. Si Pierre prend 4$, ils leur resteront en moyenne 84$.*
(problème 2, 4e secondaire)

Cette explication de l'élève pour sa réponse au problème 2 fait apparaître un nouvel élément qui permet d'interpréter autrement les réponses des élèves qui indiquent au problème 3 que la moyenne est inchangée; en fait c'est le total qui est inchangé. Ceci est plus fort en 2e secondaire (18,0% pour le problème 2 et 15,4% pour le problème 3). Il pourrait y avoir ici un glissement provenant de l'expérience des notes à l'école au Québec, où on appelle souvent moyenne la somme des notes accumulées. L'exemple 20 ci-dessus, comme le 8 dans la partie a, nous laisse penser que pour les élèves il y a la moyenne par personne et une moyenne générale pour le groupe exprimée par le total. De plus, dans environ la moitié des cas d'élèves qui répondent par un *total*, nous observons la manifestation des conceptions mentionnées plus haut « tous ont la moyenne » ou « zéro ne change rien », qui se superposent à la confusion total – moyenne. L'énoncé 20 en est un exemple.

On pourrait alors ajouter les cas classés *Total* à ceux classés *conceptions inadéquates* et ainsi, une interprétation inadéquate de la moyenne deviendrait la principale explication des erreurs pour le problème 3 comme pour la question a. Notons que ne sont pas considérés ici les élèves (3,4%) ayant reconduit leur réponse de a) (la moyenne ne change pas) sans ajouter d'explication supplémentaire.

6. DISCUSSION

Les problèmes 1 et 3 sont semblables, dans le premier cas, on ajoute une donnée ce qui fait diminuer la moyenne, dans le deuxième cas, on retranche une donnée ce qui fait augmenter la moyenne. On pouvait s'attendre à des comportements semblables chez les élèves. Le problème 2, impliquant deux modifications, le retrait et l'ajout d'une donnée, pouvait être plus difficile. Nous avons en effet observé des procédures de réussite semblables pour les problèmes 1 et 3 lorsque nous avons demandé d'anticiper l'effet sur la moyenne d'une modification de l'effectif. Les réponses des élèves mettent alors l'accent principalement sur l'effectif qui change. Pour le problème 2 (remplacement d'une donnée par 0), les réponses des élèves mettent l'accent sur le total qui change. Lors du calcul de la moyenne en (b) la procédure de réussite est dans tous les cas l'utilisation de la formule : somme des données sur l'effectif.

Cependant, lorsque nous faisons l'examen des procédures dans le cas de non-réussite, les problèmes font apparaître des procédures différentes et un degré de difficulté apparemment différent.

Le problème 3 (retrait d'une donnée égale à 0) s'est révélé plus difficile que les autres exposant des *conceptions inadéquates: «* zéro ne change rien », « tous ont 12$ ». Dans le problème 1 (ajout d'une donnée égale à 0), si on applique la conception « tous ont 12 », on a 7 amis à 12$ chacun, on obtient un total de 7 × 12$ = 84$, ce qui est correct. Parce que Pierre arrive par la suite, il semble que l'on soit moins tenté de lui attribuer un montant de 12$ que tous ont. On arrive donc à la bonne réponse. Dans le problème 3, le fait que le zéro soit une des données de départ semble rendre difficile le retrait. Si on applique la conception « tous ont 12 », on a neuf amis à 12$ chacun, on obtient un total de 9 × 12$=108$. Lorsque l'un des amis part, la question se pose: il part avec combien? (1) S'il part avec 0$, ce qu'il avait au départ, que fait-on de son 12$ étant donné que « tous ont 12$ » ? Il peut être distribué aux personnes restantes (12/8); peu d'élèves utilisent cette alternative. (2) S'il part avec 12$, la moyenne reste inchangée. On arrive donc à une mauvaise réponse. On peut donc considérer que la conception est inadéquate dans ce cas et c'est pourquoi elle se révèle plus souvent dans les réponses au problème 3. Considérer la moyenne comme le résultat d'un partage égal de la somme, donc que tous ont un montant équivalent à la moyenne, est une conception utile qui permet de résoudre facilement plusieurs problèmes de moyenne. Dans le problème 3, cette représentation de la moyenne provoque manifestement un glissement: le 12$ devenant une somme attachée à la personne et non une somme fictive en cas de partage égal.

Quant au zéro neutre, la conception « zéro ne change rien » pouvait conduire à dire que la nouvelle donnée ou la donnée retranchée étant nulle, elles ne changeaient rien à la moyenne. Dans le problème 1, seulement un élève a donné une telle réponse. La généralisation de la propriété du zéro neutre ou l'obstacle épistémologique ou l'influence sociale du contexte, dont nous avons parlé plus tôt, ne se sont donc pas manifestés, s'ils existent, dans le problème 1. Dans ce problème (1), la donnée nulle étant ajoutée après, il était clair qu'elle n'intervenait pas dans la moyenne initiale. Dans le problème 3, le fait que le zéro soit une des données de départ peut avoir conduit à se demander si Pierre avait été compté ou non dans le calcul de cette moyenne. En effet, certains énoncés révèlent que Pierre qui avait 0$ est exclu du calcul de la moyenne conduisant ainsi à une réponse fausse. De plus, dans le problème 2, la donnée « zéro » entre en conflit avec la conception « tous ont 12 », ce qui n'est pas le cas dans le problème 1 puisque la donnée « zéro » n'en est pas une de la distribution de départ. N'est-ce qu'une question de formulation ou de contexte? Il serait pertinent de vérifier cette différence entre les problèmes 1 (ajout d'une donnée) et 3 (retrait d'une donnée) dans d'autres contextes. Nous pensons que le comportement n'est pas seulement dû au fait que la donnée soit « zéro » mais aussi au fait que cette donnée différente de la moyenne entre en conflit avec des conceptions inadéquates de la moyenne.

Le problème 2 (remplacement d'une donnée par 0), par ailleurs, a fait apparaître une confusion entre le total et la moyenne. Comme dans ce problème, le total change, c'est sur ce total que l'accent est mis dans les procédures de réussite en (a). On peut alors comprendre que ce même accent se note dans certaines procédures erronées.

Ces résultats différents pour des problèmes semblables montrent l'importance de soumettre les élèves à des problèmes variés. L'influence des grandeurs en jeu peut favoriser certains raisonnements; ainsi, dans Mary et Gattuso (2003), nous avons fait l'hypothèse que le meilleur taux de réussite obtenu à un problème de poids, comparativement à un problème d'âge et à un problème de notes, pouvait s'expliquer par le sens que prend le total (la somme des poids dans un ascenseur a un sens, une pertinence, tandis que la somme des âges d'un groupe de personnes dans un autobus n'en a pas). Dans les problèmes discutés dans cet article, le contexte de partage était particulièrement propice à un raisonnement sur le total. Toutefois, nos résultats différents pour des problèmes de même contexte montrent bien l'importance de varier aussi la structure des problèmes de manière à confronter certaines conceptions qui pourraient apparaître dans une situation et pas dans l'autre et à prévenir des glissements de sens.

Sur un autre plan, l'influence du contexte de partage a pu être favorable à certaines stratégies adéquates d'égalisation des données, mais a pu aussi permettre l'apparition de stratégies inadéquates. Ainsi, s'il peut être intéressant d'appuyer une approche d'enseignement sur une représentation commune de la moyenne, par exemple celle de l'égalisation des données ou de partage égal *(la moyenne est la valeur qu'aurait chacune des données si on répartissait également le total des données),* les résultats obtenus ici *appellent* toutefois *à une certaine vigilance*.

Ces résultats laissent penser aussi que les *conceptions inadéquates* et glissements diminuent avec les années. C'est rassurant! Si les élèves de 2e secondaire recourent davantage à ces conceptions que les élèves de 4e secondaire, c'est sans doute qu'ils sont plus près du contexte et plus influencés par celui-ci que les élèves de 4e secondaire. Ces derniers utilisent beaucoup plus la *formule* dans le calcul de la moyenne (procédure gagnante) que les élèves des deux autres niveaux. Par contre, ils ne sont pas plus habiles pour évaluer une moyenne lorsque les informations se complexifient (problème 2). Ces résultats montrent les difficultés qui peuvent survenir lorsque des élèves collent *au sens contextuel* (en 2e secondaire) mais aussi les difficultés d'application correcte de la formule étant donné le contexte (4e secondaire). Il ne s'agit donc pas de privilégier l'utilisation d'algorithme sous prétexte d'éviter des glissements. Plusieurs auteurs ont montré les limites de cette représentation algorithmique de la moyenne (Cai, 1995; Garfield & Ahlgren, 1988; Lappan & Zawojewski, 1988). Il apparaît important de faire réfléchir les élèves sur les liens entre chacun des éléments impliqués (effectif, données, total, moyenne) et sur l'effet sur les autres de modifier l'un ou l'autre. Comme dit précédemment, cette relation peut être plus immédiate dans certaines situations que dans d'autres.

Dans les problèmes 2 et 3, la relation entre les données, l'effectif et la moyenne apparaît difficile à reconstituer pour les élèves. Les taux de réussite sont peu élevés. Ces problèmes pourraient être considérés comme de bons indicateurs pour évaluer le niveau de compréhension des élèves dans la perspective développementale de Watson et Moritz (1999). Cependant, nous ne retrouvons pas la progression des pourcentages de réussite selon l'âge qui pourraient être attendue pour le problème 2. Nos résultats mettent davantage en lumière les différences selon la structure des problèmes. Toutefois, certaines observations concernant le problème 3(b), telles la diminution du recours à des conceptions inadéquates avec l'âge et l'augmentation des procédures de réussite de type formule permettent de comprendre l'évolution des élèves sous l'influence des nouveaux savoirs enseignés et des nouvelles habiletés de raisonnement dont ils disposent. Ceci nous fait dire, qu'il est important de faire en sorte que les stratégies différentes soient confrontées de manière justement à faire le lien entre certaines plus primitives qui sont adéquates mais qui peuvent provoquer des glissements et d'autres plus sophistiquées mais dont le sens échappe parfois.

7. CONCLUSION

Pour cette étude, les données ont été recueillies à l'aide d'un questionnaire écrit. Bien que cette méthode ait l'avantage d'atteindre un plus grand nombre de sujets, elle comporte ses difficultés. La première et non la moindre, est celle de la formulation des questions. De plus, la réponse écrite est souvent moins révélatrice d'informations sur les raisonnements du sujet qu'une réponse qui peut s'expliquer en entrevue.

Malgré tout, les résultats sont intéressants et il faudra en tenir compte dans l'enseignement. À l'école, on met souvent l'accent sur la formule. Un traitement global des liens entre l'effectif, le total et la moyenne, par anticipation des effets d'une modification de l'un ou l'autre de ces éléments, nous apparaissait particulièrement intéressants pour développer chez les élèves le sens de la moyenne comme mesure représentative d'une distribution. Les élèves ne sont pas démunis devant ces anticipations. Comme nous l'avons vu, ils utilisent un certain nombre de stratégies pour anticiper l'effet d'une modification et réussissent dans une assez bonne proportion. Par exemple, une majorité d'élèves semble avoir combiné les effets de deux transformations sur le total pour évaluer l'effet sur la nouvelle moyenne. Il paraît pertinent d'exploiter ce type de raisonnement puisqu'il est utilisé par les élèves et qu'il semble productif. Prendre comme point de départ les anticipations et stratégies des élèves peut être l'occasion d'une discussion fructueuse pour une meilleure compréhension de la moyenne et pour affronter les conceptions inadéquates latentes dans l'esprit des élèves. Par ailleurs, comme nous l'avons développé dans cet article, les différences de comportement observées chez les élèves, selon les problèmes, demandent de varier les modifications effectuées justement pour que ces conceptions latentes se manifestent.

REFERENCES

Bell, A., Fischbein, R., & Greer, B. (1984). Choice of operation in verbal arithmetic problems: The effects of number size, problem structure and context. *Educational Studies in Mathematics, 15,* 129-147.

Cai, J. (1995). Beyond the computational algorithm: Students' understanding of the arithmetic average concept. In L. Meira & D. Carraher (Eds.), *Proceedings of the 19th International Conference for the Psychology of Mathematics Education* (vol. 3, pp. 144-151). Pernambuco, Brazil: Universidade Federal de Permanbuco.

Cerquetti-Aberkane, F. (1987). Des erreurs et des maîtres. Les erreurs des élèves trouvent souvent leur source chez les enseignants. *Prospectives,* octobre, 120.

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

Gattuso, L., & Mary, C. (1996). Development of concepts of the arithmetic average from high school to university. In L. Puig & A. Gutiérrez (Eds.), *Proceedings of the 20th International Conference for the Psychology of Mathematics Education* (pp. 401-408). Valenzia, Spain.

Gattuso, L., & Mary, C. (1998). Development of the concept of weighted average among high-school children. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee & W-K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 685-692). Voorburg, The Netherlands: International Statistical Institute.

Gattuso, L., & Mary, C. (2001). Pupils perception of the links between data and their arithmetic average. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th International Conference for the Psychology of Mathematics Education* (vol. 2, pp. 25-32). The Netherlands: Utrecht University.

Hardiman, P., Well, A., & Pollatsek, A. (1984). Usefulness of a balance model in understanding the mean. *Journal of Educational Psychology*, *76*(5), 792-801.

Kaplan, R. (1999). *The nothing that is a natural history of zero*. New York: Oxford University Press.

Lappan, G., & Zawojewski, J. (1988). Teaching statistics: Mean, median, and mode. *Arithmetic Teacher*, *March 88*, 25-26.

Lavoie, P., & Gattuso, L. (1998). An historical exploration of the concept of average. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee & W-K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 1051-1058). Voorburg, The Netherlands: International Statistical Institute.

Leon, M., & Zawojewski, J. (1990). Use of arithmetic mean: An investigation of four properties issues and preliminary results. In D. Vere-Jones (Ed.), *Proceedings of Third International Conference on Teaching Statistics Vol. 1: School and General Issues* (pp. 302-306). Voorburg, The Netherlands: International Statistical Institute.

Mary, C., & Gattuso, L. (2003). L'influence des grandeurs impliquées sur la résolution d'un problème de moyenne. *Communication dans le cadre du colloque EMF 2003*, Tozeur, Tunisie, décembre, 2003.

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

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

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

Semadeni, Z. (1984). A principle of concretization permanence for the formation of arithmetical concepts. *Educational Studies in Mathematics, 15*, 379-395.

Stern, E., & Lehmdorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 7*, 259-268.

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

Vergnaud, G. (1996). La théorie des camps conceptuels. In J. Brun (Ed.), *Didactique des mathématiques* (pp. 197-242). Lausanne: Delachaux et Niestlé.

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

Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1 & 2), 11-50.

CLAUDINE MARY
Département d'études sur l'adaptation scolaire et sociale
Université de Sherbrooke
Sherbrooke, J1K 2R1
Canada

PAST IASE CONFERENCES

1. SRTL-4 THE FOURTH INTERNATIONAL RESEARCH FORUM ON STATISTICAL REASONING, THINKING AND LITERACY
Auckland, New Zealand, July 2-7, 2005

The fourth research forum in a series of international research forums on Statistical Reasoning, Thinking and Literacy (SRTL) took place in winter at The University of Auckland in New Zealand. This particular gathering of researchers has played an important role in advancing our understanding of the richness and depth of reasoning about distribution, a key focus of statistics education.

The forum was sponsored by the Key College Press (USA), The American Statistical Association (ASA) Section on Statistical Education, the Department of Statistics, The University of Auckland, the Department of Mathematics, The University of Auckland, and the New Zealand Statistical Association (NZSA).

The focus of SRTL-4 on reasoning about distribution emerged from the previous three SRTL conferences. Distribution is a key concept in statistics, and yet statisticians and educators may not be aware of how difficult it is for students to develop a deep understanding of this concept. When students are given tasks involving comparing distributions or making inferences, they often fail to utilize relevant information contained in the underlying distributions. Curricular materials often focus on construction and identification of distributions, but not on what these distributions mean to students and how they interpret them.

Twenty researchers in statistics education from six countries shared their work and discussed important issues in a stimulating and enriching environment. Sessions were held in an informal style, with a high level of interaction. With emphasis on reasoning about distribution, a wide range of research projects were presented spanning learners of all ages, as well as teachers. These demonstrated an interesting diversity in research methods, theoretical approaches and points of view. As a result of the success of this gathering, plans are already underway for the next gathering (SRTL-5) in 2007.

The programme began with an overview talk by Chris Wild entitled: “A statistician’s view on the concept of distribution.” Eight presentations of SRTL-4 were thematically grouped into five clusters. A cluster included one or two ninety-minute research presentations to the entire group, small group discussions, and a whole group reflection on the cluster. All presenters showed a small subset of video segments of their research. Optional time was devoted to viewing and discussing the research video-tapes from methodological and interpretive perspectives. In addition, three doctoral students presented their current research findings in a poster session and a software developer discussed potential research questions to the entire group. The programme ended with three discussants’ reflecting on reasoning about distribution from research, curriculum, and technology viewpoints.

The research forum proved to be very productive in many ways. Several types of scientific publications will be produced including a CD-ROM of the proceedings edited by Katie Makar, papers in refereed journals, and a special issue of Statistics Education Research Journal (SERJ) on reasoning about distribution co-edited by Maxine Pfannkuch and Chris Reading. An additional product of the meeting was a new SRTL Website

Statistics Education Research Journal, 4(2), 103-111, http://www.stat.auckland.ac.nz/serj

hosted by the Department of Statistics, The University of Auckland that includes a variety of resources. These will all serve as a rich resource for statistics educators and researchers.

Maxine Pfannkuch, The University of Auckland, was the local SRTL-4 organiser ably assisted by Ross Parsonage, social programme, Chris Wild, finance, and Stephen Cope, computer technician and webmaster. Beyond the scientific programme, participants took part in a variety of social events and local excursions that helped to build a sense of community amongst the researchers and to enjoy the wild beauty of Auckland and its beaches, bush, and indigenous Maori and Pacifica cultures. A DVD of the scientific and social programme, edited by Pip Arnold, will be available for participants.

For further information please contact the SRTL Co-chairs Joan Garfield (jbg@umn.edu) and Dani Ben-Zvi (dbenzvi@univ.haifa.ac.il) or visit the SRTL-4 web page http://www.stat.auckland.ac.nz/srtl4/.

OTHER PAST CONFERENCES

1. UNITED STATES CONFERENCE ON TEACHING STATISTICS, USCOTS
Columbus, OH, USA, May 19-21, 2005

The first United States Conference on Teaching Statistics (USCOTS) was held on May 19-21, 2005 at the Ohio State University in Columbus, Ohio, hosted by CAUSE, the Consortium for the Advancement of Undergraduate Statistics Education. USCOTS was an active, hands-on working conference for teachers of Statistics at the undergraduate level, in any discipline or type of institution, including high school teachers of Advanced Placement Statistics. The theme of the 2005 USCOTS was "Building Connections for Undergraduate Statistics Teaching" and focused on ways that we can share teaching ideas, develop working relationships, and identify areas for future collaborations and projects at our own institutions. USCOTS focused on three major areas: curriculum, pedagogy, and research. Lots of good resources for each of these areas are provided on USCOTS web page http://www.causeweb.org/uscots/census/. For more information about USCOTS, please contact Deborah Rumsey, USCOTS program chair at rumsey@stat.ohio-state.edu. Website: www.causeweb.org/uscots/

2. PSYCHOLOGY OF MATHEMATICS EDUCATION, PME-29
Melbourne, Australia, July 10-15, 2005

The PME-29 conference was held on July 10-15, 2005 in Melbourne, Australia. More information from Helen Chick, h.chick@unimelb.edu.au.

Website: staff.edfac.unimelb.edu.au/~chick/PME29/

3. BEYOND THE FORMULA IX
Rochester, NY, USA, August 4-5, 2005

For the ninth straight year, teachers of introductory statistics from high schools, two- and four-year colleges gathered in Monroe Community College's Thomas R. Flynn Conference Center to learn about the latest developments in practicing their craft. The 125 participants from 18 states, Canada and Australia arrived for the August 4 and 5, 2005 conference. They heard talks, tried out new technology and discussed curriculum changes related to basic statistics, a course that continues to grow in popularity as both a requirement and elective at all levels. The conference was planned and hosted by members of MCC's Mathematics Department. Abstracts of papers given by seven invited speakers are available from their web site. Next BTF conference will be organized in 2007. Website: www.monroecc.edu/go/beyondtheformula/

4. JOINT STATISTICAL MEETINGS, JSM 2005
Minneapolis, MN, USA, August 7-11, 2005

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 included oral presentations, panel sessions, poster presentations, continuing education courses, placement service, society and section business meetings.

The ASA Section on Statistical Education organized three invited sessions and several contributed sessions for this conference. Topics included Measurement Issues in

Statistical Education, Interdisciplinary Approaches to Statistics Education, Assessments of Students Learning and Attitudes in Introductory Statistics, Publishing in Statistics Education Journals: Views from the Editors, Using Japanese Lesson Study to Develop Research Based Lessons in Statistics, Nuts and Bolts of Classroom Assessment, Implementing the GAISE Guidelines in College Statistics Courses, Ideas and Examples for Teaching Concepts in Statistics Classrooms, Career Advice in Statistics Education, CAUSE Updates on Research, CAUSEweb, and USCOTS, and Using Technology and the Web When Teaching Statistics.

Web site: www.amstat.org/meetings/jsm/2005/

FORTHCOMING IASE CONFERENCES

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

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

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

1.1. CALL FOR PAPERS

Statistics educators, statisticians, teachers and educators at large have been invited to contribute to the scientific programme. Types of contribution include *invited papers, contributed papers* and *posters*. No person may author more than one Invited Paper at the conference, although the same person can be co-author of more than one paper, provided each paper is presented by a different person.

Voluntary refereeing procedures have been implemented for ICOTS7. Details of how to prepare manuscripts, the refereeing process and final submission arrangements are available from the ICOTS7 website at www.maths.otago.ac.nz/icots7.

Invited Papers

Invited Paper Sessions have been organized within nine different Conference Topics 1 to 9. The list of Topic and Sessions themes, with email contact for Session Organisers is available at the ICOTS-7 website, under "Scientific Programme."

Contributed Papers

Contributed paper sessions are being arranged in a variety of areas. Those interested in submitting a contributed paper contact either Joachim Engel (Engel_Joachim@ph-ludwigsburg.de) or Alan McLean (alan.mclean@buseco.monash.edu.au).

Posters

Those interested in submitting a poster should contact Celi Lopes (celilopes@uol.com.br) before February 1, 2006.

Special Interest Group Meetings

These are meetings of Special Interest Groups of people who are interested in exchanging and discussing experiences and/or projects concerning a well-defined theme of common interest. Proposals to hold a SIG Meeting specifically oriented to reinforce Latin American statistics education cooperation in a particular theme are especially welcome. In this case the organisers may decide to hold the meeting in Portuguese and Spanish language. Individuals or groups may submit proposals to establish a Special Interest Group to Carmen Batanero at (batanero@ugr.es).

1.2. IMPORTANT DEADLINES

Deadlines for Submission of proposals

Contributed Papers: Refereed Nov. 1, 2005, Non-Refereed Jan. 1, 2006
Special Interest Groups, Special Sessions, Posters: Feb. 1, 2006

Closing date for Submission of papers (invited and contributed):

Final version: Nov. 1, 2005 (if to be refereed)
Final version: Jan. 1, 2006 (if not to be refereed)

1.3. PLENARY SESSIONS

Pedro Luis do Nascimento Silva (Brazil): *Statistical Education for Doing Statistics Professionally: Some Challenges and the Road Ahead*
Bryan Manly (New Zealand): *Cooperation and Conflict in Environmental Statistics*
Joan Garfield (USA): *Collaboration in Statistics Education Research: Stories, Reflections, and Lessons Learned*
Mike Shaughnessy (USA): *Students' Thinking about some Important Concepts in Statistics*
Chris Wild (New Zealand), Closing Speaker: *On collaboration, Competition and Making Connections*
Len Cook (UK), After Dinner Speaker: *Training Statisticians for Working in Public Affairs*

Panel Discussion: *The challenges for cooperation in statistics education*
Chair: Pedro Morettin (Brazil); Speakers include: Evelio Fabbroni (Panama), Jae C. Lee (Korea), Pilar Martín Guzmán (Spain) and Allan Rossman (USA).

1.4. TOPICS AND TOPIC CONVENORS

Topic 1. *Working cooperatively in statistics education*. Lisbeth Cordani, lisbeth@maua.br and Mike Shaughnessy, mike@mth.pdx.edu
Topic 2. *Statistics Education at the School Level*. Dani Ben-Zvi, benzvi@univ.haifa.ac.il and Lionel Pereira, lpereira@nie.edu.sg
Topic 3. *Statistics Education at the Post Secondary Level*. Martha Aliaga, martha@amstat.org and Elisabeth Svensson, elisabeth.svensson@esi.oru.se
Topic 4. *Statistics Education/Training and the Workplace*. Pedro Silva, pedrosilva@ibge.gov.br and Pilar Martín, pilar.guzman@uam.es
Topic 5. *Statistics Education and the Wider Society*. Brian Phillips, BPhillips@groupwise.swin.edu.au and Phillips Boland, Philip.J.Boland@ucd.ie
Topic 6. *Research in Statistics Education*. Chris Reading, creading@une.edu.au and Maxine Pfannkuch, pfannkuc@scitec.auckland.ac.nz
Topic 7. *Technology in Statistics Education.* Andrej Blejec, andrej.blejec@nib.si and Cliff Konold, konold@srri.umass.edu
Topic 8. *Other Determinants and Developments in Statistics Education.* Theodore Chadjipadelis, chadji@polsci.auth.gr and Beverley Carlson, bcarlson@eclac.cl
Topic 9. *An International Perspective on Statistics Education*. Delia North, delian@icon.co.za and Ana Silvia Haedo, haedo@qb.fcen.uba.ar
Topic 10. *Contributed Papers*. Joachim Engel, Engel_Joachim@ph-ludwigsburg.de and Alan McLean, alan.mclean@buseco.monash.edu.au

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

1.5. ORGANISERS

Local Organisers

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

IPC Executive

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

More information and the **Second announcement** is available from the ICOTS-7 web site at www.maths.otago.ac.nz/icots7 or from the ICOTS IPC Chair Carmen Batanero (batanero@ugr.es), the Programme Chair Susan Starkings (starkisa@lsbu.ac.uk) and the Scientific Secretary John Harraway (jharraway@maths.otago.ac.nz).

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

The 56th Session of the International Statistical Institute (ISI) will be held in Lisboa, Portugal. IASE is usually very active at ISI meetings (see SERJ 4(1) for the report of IASE activities at ISI-55 in Sydney) and will sponsor a list of Invited Paper Meetings (IPM).

2.1. IPMS SPONSORED BY THE IASE

IPM37 *Research on Reasoning about Distribution*
IPM38 *How Modern Technologies have Changed the Curriculum in Introductory Courses*
IPM39 *Preparing Teachers of Statistics*
IPM40 *Research on the Use of Simulation in Teaching Statistics and Probability*
IPM41 *Optimizing Internet-based Resources for Teaching Statistics* (co-sponsored by IASC)
IPM42 *Observational Studies, Confounding and Multivariate Thinking*
IPM43 *Teaching of Official Statistics* (co-sponsored by IAOS)
IPM44 *Teaching of Survey Statistics* (co-sponsored by IASS)
IPM45 *Studying Variability through Sports Phenomena* (co-sponsored by Sports Statistics)
IPM46 *Use of Symbolic Computing Systems in Teaching Statistics* (co-sponsored by IASC)

2.2. SPECIAL CONTRIBUTED PAPER MEETINGS (SCPMS)

The ISI-56 organizers would like to draw the attention of those who could not be included in the final IPM list, to the possibility of organizing Special Contributed Paper

Meetings (SCPMs). Proposals for SCPMs are welcome and the deadline is the 28th of February, 2006.

For these special sessions, the proponents should select a specific topic, preferably not yet contemplated in the IPMs list, together with five or six speakers. All participants will have to register and submit their papers, using the same procedures as for other Contributed Papers. The SCPM application form should be sent to the e-mail ivette.gomes@fc.ul. More information and the SCPM application form is available from the ISI-56 website.

2.3. IASE COMMITTEE

Allan J. Rossman (Chair, arossman@calpoly.edu), Gilberte Schuyten (Gilberte.Schuyten@UGent.be) and Chris Wild (c.wild@auckland.ac.nz).

More information can be found at ISI-56 web site: http://www.isi2007.com.pt/

OTHER FORTHCOMING CONFERENCES

1. INTERNATIONAL CONFERENCE OF THE MATHEMATICS EDUCATION INTO THE 21ST CENTURY PROJECT: "REFORM, REVOLUTION AND PARADIGM SHIFTS IN MATHEMATICS EDUCATION"

Johor Bharu, Malaysia, November 25 – December 1, 2005

The conference will be organized in Johor Bharu, Malaysia, close to Singapore and in the heart of Tropical South East Asia. Mercifully this part of Malaysia was unaffected by the tragic earthquake and tidal wave at the end of 2004. The time and place were chosen to encourage teachers and mathematics educators from around the world to communicate with each other about the challenges of Reform, Revolution and Paradigm Shifts in Mathematics Education. The Malaysia Conference is organised by the Mathematics Education into the 21st Century Project - an international educational initiative whose coordinators are Dr. Alan Rogerson (Poland) and Professor Fayez Mina (Egypt). Since its inception in 1986, the Mathematics Education into the 21st Century Project has received support and funding from many educational bodies and institutions throughout the world. In 1992 UNESCO published our Project Handbook "Moving Into the 21st Century" as Volume 8 in the UNESCO series Studies In Mathematics Education.

The Mathematics Education into the 21st Century Project is dedicated to the improvement of mathematics education world-wide through the publication and dissemination of innovative ideas. Many prominent mathematics educators have supported and contributed to the project, including the late Hans Freudental, Andrejs Dunkels and Hilary Shuard, as well as Bruce Meserve and Marilyn Suydam, Alan Osborne and Margaret Kasten, Mogens Niss, Tibor Nemetz, Ubi D'Ambrosio, Brian Wilson, Tatsuro Miwa, Henry Pollack, Werner Blum, Roberto Baldino, Waclaw Zawadowski, and many others throughout the world. For further conference details email to Alan Rogerson, arogerson@vsg.edu.au.

Website: http://math.unipa.it/~grim/21_malasya_2005.doc

2. ASIAN TECHNOLOGY CONFERENCE IN MATHEMATICS, ATCM2005

Cheong-Ju, South Korea, December 12-16, 2005

The 10th Annual conference of Asian Technology Council in Mathematics (ATCM) on the theme Enriching Technology in Enhancing Mathematics for All is hosted by Korea National University of Education in Cheong-Ju, South Korea. The aim of this conference is to provide a forum for educators, researchers, teachers and experts in exchanging information regarding enriching technology to enhance mathematics learning, teaching and research at all levels. English is the official language of the conference. The conference will cover a broad range of topics on the application and use of technology in Mathematics research and teaching. More information: Professor Hee-chan Lew, hclew@knue.ac.kr.

Website: http://www.atcminc.com/mConferences/ATCM05/

STATISTICS EDUCATION RESEARCH JOURNAL REFEREES DECEMBER 2004 – NOVEMBER 2005

Acknowledgement: We are grateful to the following IASE members and researchers who acted as referees for the Statistics Education Research Journal in the period from December 2004 to November 2005.

Carmen Batanero, Spain
Gail Burrill, USA
Beth Chance, USA
Jean-Hugues Chauchat, France
Megan Clark, New Zealand
Bob delMas, USA
Lyn English, Australia
Antonio Estepa, Spain
Linda Gattuso, Canada
Juan Godino, Spain
Sue Gordon, Australia
Kathy Green, USA
Peter Johnston-Wilder, UK
Ruben Klein, Brazil
Cynthia Langrall, USA
Katie Makar, Australia
Kevin McConway, UK
Peter Petocz, Australia
Jorges Romeau, USA
Rosalind Phang, Singapore
Mike Shaughnessy, USA
Thaddeus Tarpey, USA
James Tarr, USA
Joseph Wisenbaker, USA