Statistics Education Research Journal

Volume 6 Number 1 May 2007

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

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

Statistics Education Research Journal

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

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

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

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

Submissions

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

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

ISSN: 1570-1824

TABLE OF CONTENTS

Editorial 2

Call for Papers: Special Issue on Reasoning about Informal Statistical Inference 3

Luc Budé, Margaretha W. J. Van De Wiel, Tjaart Imbos, Math J. J. M. Candel, Nick J. Broers, Martijn P. F. Berger
Students' Achievements in a Statistics Course in Relation to Motivational Aspects and Study Behaviour 5

James E. Corter and Doris C. Zahner
Use of External Visual Representations in Probability Problem Solving 22

Mark A. Earley
Students' Expectations of Introductory Statistics Instructors 51

Past Conference 67

Forthcoming IASE Conferences 68

Other Forthcoming Conferences 74

EDITORIAL

I have recently returned from the second biennial United States Conference on Teaching Statistics (USCOTS 2), held in Columbus, Ohio. At the conference Joan Garfield, a member of the *SERJ* Editorial Board, received the CAUSE/USCOTS Lifetime Achievement award. In a special presentation at the conference, Joan reflected on the statistics education research collaborations in her past and present. It is clear that the field of statistics education research has developed because of Joan's curiosity, motivation, and perseverance in locating research spread among many and varied disciplines including mathematics, psychology, education and statistics. From 1987 through 1999 Joan published the *Newsletter of the International Study Group for Research on Learning Probability and Statistics*, which later evolved into the *Statistical Education Research Newsletter* (2000 through 2001) and finally into what *SERJ* is today. Thank you, Joan, and congratulations on your well-deserved recognition.

A healthy strand of statistics education research sessions ran through the program at USCOTS 2. Some presenter names familiar to *SERJ* readers included Beth Chance, Bob delMas, Marsha Lovett, and Mike Shaughnessy, in addition to Joan. We were also given a glimpse into the future of research in our area when Joan acknowledged her current Statistics Education Doctoral students at the University of Minnesota. I hope that the success of Joan's program will inspire others to sprout up around the world.

Iddo Gal and I have been working behind the scenes to disseminate awareness of *SERJ*, and I am pleased to report that *SERJ* is now listed in two indexing services. One is "Cabell's Directory of Publishing Opportunities" (www.cabells.com) and the second is PsycINFO, which is managed by the American Psychological Association (see www.apa.org/psycinfo/about/covlist.html). We are working out the details of also listing *SERJ* abstracts in EBSCOhost (see www.epnet.com), and will continue to seek other opportunities to abstract *SERJ* in indices so that researchers will be able to easily find the work we publish.

Iddo and I have also been monitoring the acceptance rate for manuscripts submitted to *SERJ*. We received 30 manuscripts in 2006, and four of them have so far been accepted for publication in *SERJ*. A few more of these are still in revision and may eventually be accepted and published. The acceptance rate indicates the selectivity and high standards we maintain for *SERJ*, but it also suggests that there is plenty of room for more high quality manuscripts from statistics education researchers.

As Iddo's term as *SERJ* co-editor comes to an end, we will be announcing a new co-editor later this year. Thanks to Iddo for his leadership, creativity, and attention to detail during his term.

Please enjoy the articles and announcements in this new issue of *SERJ*, and note the Call for Papers for a Special Issue on Reasoning about Informal Statistical Inference. Thank you for reading the journal, and please consider sending the results of your own research to us!

TOM SHORT, for IDDO GAL

Statistics Education Research Journal, 6(1), 2, http://www.stat.auckland.ac.nz/serj

CALL FOR PAPERS: REASONING ABOUT INFORMAL STATISTICAL INFERENCE

The *Statistics Education Research Journal* (*SERJ*), a journal of the International Association for Statistical Education (IASE), is planning a special issue for November 2008, focused on research on *Reasoning about Informal Statistical Inference*. **Submission deadlines: Letters of intent by Sept. 15, 2007; Full papers by Nov. 1, 2007**. Guest Editors will be Dave Pratt (University of London, UK, <d.pratt@ioe.ac.uk>), and Janet Ainley (University of Leicester, UK, <jma30@le.ac.uk>).

1. ABOUT INFORMAL STATISTICAL INFERENCE

The aim of the special issue is to advance the current state of research-based knowledge about the development, learning, and teaching of statistical inference, a foundational area in statistics education. For the special issue we seek articles focused on a critical *subset* of issues in this broad area, describing research related to the understanding, learning, or teaching of *informal* aspects of inferential statistical reasoning, and demonstrating a contribution to research-based educational practice in this area.

It is recognized at the outset that the definition of what counts as "informal statistical inference" may at times be slippery, that is, what is informal could depend on the nature of the inferential tasks being studied, on the complexity of the statistical or probabilistic concepts involved, on the educational stage, and on other factors. The editors will select papers for the special issue that focus on learners' informal ideas about statistical inference or on learners' intuitive ways of reasoning about statistical inference in diverse contexts (see possible research topics below), *not* on mastery of formal procedures or on the learning/teaching of formal methods of statistical estimation, significance tests, etc. The papers being sought will be based on empirical research of a quantitative and/or qualitative nature on individuals or groups involved in all stages of education, including all levels of schooling, teacher education, professional development, and workplace and adult education. Papers on informal inferential reasoning invoked when people face everyday statistical tasks may also be considered, provided that they discuss clear educational implications.

2. POSSIBLE RESEARCH TOPICS

Key examples of relevant topics for papers that may fit under the general heading of *Reasoning about Informal Statistical Inference* include:

a. How does reasoning about statistical inference develop from simple forms towards more complex ones? What stages exist in the acquisition of informal knowledge about statistical inference, or in learning to communicate information or interpret displays about statistical inference, and how do students develop and understand the concepts and language that can be used in this regard (e.g., sampling, significance, confidence)?

b. What technological tools can be used to promote the understanding of statistical inference? How are such tools utilized by learners to help in understanding the building blocks or intermediate steps in statistical inference?

Statistics Education Research Journal, 6(1), 3-4, http://www.stat.auckland.ac.nz/serj

c. What tasks or sequences of instructional activities can help learners develop a conceptual understanding of some aspect of statistical inference? How does the design of tools and/or tasks shape students' informal inferential reasoning?
d. What types of barriers to students' informal reasoning about statistical inference are found, and how can they be avoided or overcome?
e. What types of foundational knowledge (statistical, general) or thinking processes are needed for or used by learners to informally understand and reason about statistical inference? How does an informal understanding of statistical inference connect with or depend on understanding of other statistical concepts?
f. What assessment approaches and research methodologies can be used to effectively assess understanding, reasoning or learning of informal statistical inference?

3. OTHER TOPICS RELATED TO FORMAL STATISTICAL INFERENCE

SERJ is also inviting research-based papers on learning, reasoning or understanding of formal aspects of statistical inference, that is, papers that fall outside the scope of the notion of "informal statistical inference" as described above, but that otherwise fit the general aims of the *Journal*. Such papers would be processed by *SERJ* as regular papers and if accepted will be published in a regular issue. Should enough such papers be accepted for publication, they will be grouped together in a special section and prefaced with an introductory paper by a member of the *SERJ* Editorial Board.

4. SUBMISSION GUIDELINES

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

5. DEADLINES AND CONTACT INFORMATION

Interested authors should send *a letter of intent by Sept. 15, 2007*, but preferably earlier, with a 150-250 word abstract describing key aspects of the research. This letter should be sent by e-mail to *SERJ* co-editor Iddo Gal: <iddo@research.haifa.ac.il>, and authors can expect to get a quick response within 10 days. Authors wishing to send informal queries regarding the suitability of a planned paper can also contact Iddo Gal.

Full manuscripts must be submitted by Nov. 1st, 2007 at the latest to Iddo Gal at the address above, in accordance with the submission guidelines listed earlier**.**

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

STUDENTS' ACHIEVEMENTS IN A STATISTICS COURSE IN RELATION TO MOTIVATIONAL ASPECTS AND STUDY BEHAVIOUR

LUC BUDÉ
Maastricht University, The Netherlands
Luc.Bude@stat.unimaas.nl

MARGARETHA W. J. VAN DE WIEL
Maastricht University, The Netherlands
M.vandeWiel@psychology.unimaas.nl

TJAART IMBOS
Maastricht University, The Netherlands
Tjaart.Imbos@stat.unimaas.nl

MATH J. J. M. CANDEL
Maastricht University, The Netherlands
Math.Candel@stat.unimaas.nl

NICK J. BROERS
Maastricht University, The Netherlands
Nick.Broers@stat.unimaas.nl

MARTIJN P. F. BERGER
Maastricht University, The Netherlands
Martijn.Berger@stat.unimaas.nl

ABSTRACT

The present study focuses on motivational constructs and their effect on students' academic achievement within an existing statistics course. First-year Health Sciences students completed a questionnaire that measures several motivational constructs: dimensions of causal attributions, outcome expectancy, affect, and study behaviour, all with respect to statistics. The results showed that when the cause of negative events was perceived as uncontrollable, outcome expectancy was negative. When the cause of negative events was perceived as stable, affect toward statistics was negative. Furthermore, negative affect toward statistics and limited study behaviour led to unsatisfactory achievements. Path analysis (Lisrel) largely confirmed the causal relations in a model that was based on attributional and learned helplessness theories. The consequences of these findings for statistics education are discussed.

Keywords: *Statistics education research; Motivation; Conceptual understanding; Study behaviour*

Statistics Education Research Journal, 6(1), 5-21, http://www.stat.auckland.ac.nz/serj

1. INTRODUCTION

Motivation influences the scope and the quality of study behaviour of students (see e.g., Bruning, Schraw, & Ronning, 1999; Deci & Ryan, 1985; Graham & Weiner, 1987; Pintrich, 2000). High-quality study behaviour involves active knowledge construction. Active knowledge construction is known to enhance understanding of the material in many courses (see e.g., Chi, de Leeuw, Chiu, & LaVancher, 1994; Phye, 1997; Steffe & Gale, 1995), including statistics courses (see e.g., Garfield, 1993; Giraud, 1997; Keeler & Steinhorst, 1995; Magel, 1998). Therefore, in attempts to improve statistics education, it is fundamental to stimulate motivation.

Research on motivation is quite extensive and covers heterogeneous constructs (see e.g., Ames, 1992; Boekaerts, 1997; Volet, 1997; Weiner, 1992). Some of these constructs involve phenomena that are difficult to change, because they are to a large extent determined by traits of the individual that is involved, such as goal orientation, self-determination, and competence. Our aim is not to focus on such phenomena, but rather to focus on constructs that have practical implications for statistics education, that is, constructs that can be manipulated and acted upon while trying to improve statistics education.

For that reason we have focused on two motivational theories that offer opportunities to intervene in motivational processes. Both theories take the starting-point of the explanations people perceive for events they experience. These so called causal explanations have cognitive, affective, and behavioural consequences. Examples of cognitive consequences in a statistics educational context are expected outcomes of attending lectures or studying a course book; examples of affective consequences are enjoyment, pleasure, and interest; and examples of behavioural consequences are effort and persistence. The influence of causal explanations on cognition, affect, and behaviour might be manipulated and driven toward outcomes that are more positive, in terms of motivation. As a consequence, these causal explanations have practical implications for statistics education, because the obtained improvement of motivation might result in study behaviour that enhances understanding. The goal of the study was to investigate these phenomena in the context of statistics education.

2. MOTIVATIONAL MODEL

In statistics education one can sometimes encounter students who think that there is a stable cause for failing an exam (e.g., statistics is a difficult subject). These students may no longer expect to benefit from studying statistics; they may start to dislike it and will not spend much study time on this subject. Other students may think that they have no control over the outcomes of their actions. For example, “no matter how hard I study, I will not be able to understand it.” These students may in advance expect to fail on the exam, will also start to dislike statistics, and will not spend much time studying the material. These examples show the influence of causal attributions (stability of causes, non-controllability of causes) on cognitions such as outcome expectancies (no benefit from studying statistics, expectancy to fail on the exam) and consequently on emotions (affective reactions of starting to dislike statistics) and behaviour (disregarding statistics), which will finally have an effect on achievement. This chain effect, which is consequential for statistics education, is reflected in a model that was developed and tested in this study.

The model as a whole stands for motivation (see Figure 1). Motivation is not a separate entity in our model for two reasons. Firstly, it is difficult to insert it separately into a model, because it is an abstract, complex (Weiner, 1992), and ill-defined (Murphy & Alexander, 2000) construct, which is frequently used in colloquial language and consequently has several connotations. Moreover, motivation is studied in different domains and from different perspectives, which has led to distinct and changing conceptualisations and approaches. Various motivational constructs are studied, such as self-efficacy, goal orientation, metacognitive strategies, value, strategy use, causal perceptions, autonomy, social relatedness, as so forth. (See e.g., Ames, 1992; Boekaerts, 1997; Dweck, 2000; Pintrich & Schunk, 1996; Volet, 1997; Weiner, 1986.) In these studies it is often left implicit whether these constructs are part of motivation or are merely related to motivation (Murphy & Alexander, 2000). Our model as a whole reflects our perspective on motivation.

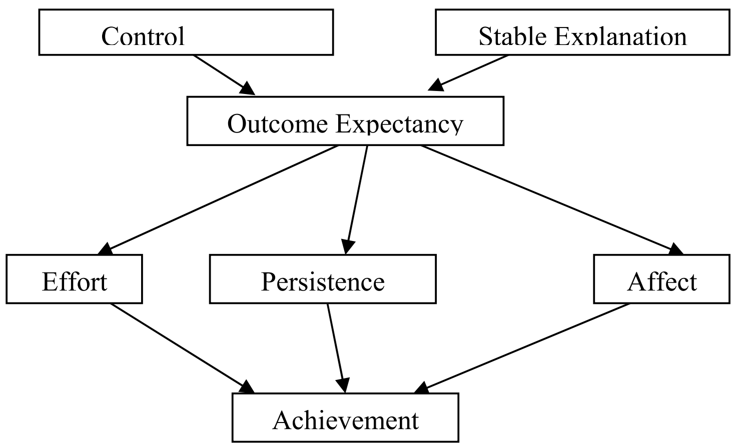

Figure 1. Statistics motivational model based on the attributional and the learned helplessness theory

Secondly, it is in our view not necessary to integrate motivation as a separate construct in the model. Traditionally, motivation was seen as an isolated latent construct that drives behaviour, cognition, and affect. We think that motivation merely is the sum of behaviour, cognition, and affect. Our opinion is in accordance with the remark of Weiner (1992), referring to Kelly (1958), that motivation as a model construct might be redundant; it is sufficient to represent only those variables that make up motivation. This view is also compatible with the fact that most motivational models do not explicitly contain motivation as a construct (see e.g., Bruning, Schraw & Ronning, 1999; Deci & Ryan, 1985; Pintrich, 2000; Pintrich & Schunk, 1996; Weiner, 1992). Therefore, the model that we developed contains only manifest variables that together stand for motivation, and does not contain motivation as a separate latent entity.

Two specific motivational theories were used for our model; namely the attributional and the learned helplessness theory, because they both use the starting-point of perceived causes for aversive events. The attribution-based theory of motivation (Graham & Weiner, 1987; Pintrich & Schunk, 1996; Weiner, 1986, 1992) commences with perceived causes for failure, unexpected outcomes, unusual events, and important situations. Perceived causes are the way people explain to themselves such outcomes, events, and situations. The connotations of the explanations are determined by underlying properties. In attribution-based theory these underlying properties of such explanations are divided into three dimensions: stability, control, and locus. Pintrich and Schunk (1996) propose, however, that the stability dimension is most closely linked to beliefs regarding future success (outcome expectancy) and subsequently to affect and actual achievement behaviour. Therefore, we integrated *stable explanation* in our model in Figure 1. It can be defined as the invariability over time of such perceived causes, namely causal explanations.

Peterson, Maier, and Seligman (1993) present a motivational theory, which originally emanates from the learned helplessness paradigm. In this paradigm, individuals are thought to become passive and to develop affective deficits if they cannot control and avoid the causes of aversive stimuli. They claim therefore, in contrast to Pintrich and Schunk (1996), that controllability is the major factor contributing to a negative outcome expectancy. Uncontrollable events will, according to Peterson et al., lead to a perceived non-contingency between people's actions and the outcomes of their actions. This negative outcome expectancy will lead to pessimistic thoughts, negative emotions (affect), and passivity (behaviour). This is what is called learned helplessness. We integrated *control* influencing outcome expectation as a separate construct in our model. Control is defined as the ability to avoid the causes of aversive stimuli.

Although the two presented theories slightly differ in the emphasis of the causal dimensions control and stability, they both reflect the way these properties of negative causal explanations contribute to a negative *outcome expectancy*, and how this will act upon *affect* and on behaviour, such as *effort* and *persistence,* which will finally result in an effect on *achievement*. The causal relations among these constructs are symbolised by arrows in our model that is presented in Figure 1.

This model was examined within the domain of statistics education. This means that all the constructs were measured with respect to statistical events and phenomena. It is known that perceived causal explanations via expectancy, affect, and behaviour determine future achievements in mathematics (see e.g., Seegers & Boekaerts, 1993; Vålas & Søvik, 1994). Our question was whether this is also true for statistics education and if the results would provide useful information for the reformation of statistics education.

The following research questions were addressed:

1. How do students causally explain statistics related events? Do they think that they have control over, for example, the mastery of the material, the amount of time they can spend on studying statistics, and the result on the tests? We also wanted to know whether or not the causes that the students reported for these events were stable.
2. We further measured the outcome expectancies, that is, whether students experience a contingency between studying statistics and their understanding of the topics and the grades they receive on statistics tests. We also investigated the influence of outcome expectancy on effort, persistence and affect.
3. Finally, we investigated the relations between these motivational constructs and achievement. The potential causal relations among these constructs were tested with structural equation modelling via Lisrel (Jöreskog & Sörbom, 1989).

3. METHOD

3.1. PARTICIPANTS

Two hundred (n = 200) first-year students of the faculty of Health Sciences participated in a pilot study to establish the reliability of a questionnaire that was developed to measure the motivational constructs. In the subsequent year n = 94 first-year students of the faculty of Health Sciences participated in the main study; 79 of these participants were female, 15 were male. The ages ranged from 19 to 26 years. Approximately 75 percent of the first-year Health Sciences student body is female. The participants were recruited during educational activities before the start of the introductory statistics course in which this study was executed. During recruitment they were told that they had to answer questions about statistics education and that they would be paid 10 euro. This payment was given to avoid attracting only motivated students who were particularly interested in statistics. All participants took the introductory statistics course.

3.2. MEASUREMENT INSTRUMENTS AND PROCEDURE

A questionnaire to measure the motivational constructs that are relevant for our model was developed. This Motivation toward Statistics Questionnaire (MSQ) consisted of 38 items, divided into six subscales. The items were phrased as statements and participants responded on a 7-point Likert scale. The questionnaire is partly a Dutch translation of the Survey of Attitudes Toward Statistics (SATS) (Gal, Ginsburg, & Schau, 1997). Additional items with regard to causal explanations were formulated using the same principles as the Attributional Style Questionnaire (ASQ) (Peterson et al., 1993), in particular for two attributional dimensions: stability and control. Finally, items were added to measure the two aspects of study behaviour: effort and persistence. All MSQ items concentrated on statistics related events. Because the MSQ was for the greater part based on existing surveys that have been proven to be valid (Peterson et al., 1993; Schau, Stevens, Dauphinee, & Del Vecchio, 1995), it can be considered an adequate measurement instrument regarding the relevant motivational constructs. Example questions are presented in Table 1. Based on content the items were divided into six subscales. To establish the reliability of the MSQ, it was administered to 200 first-year Health Sciences students and Cronbach’s alpha was computed for each subscale. Six questions that did not fit in the subscale were identified. Four questions were removed; two were rephrased. The MSQ was used the subsequent year for collecting data for the main study. It was administered to the students at the beginning of the introductory statistics course. Students received written instructions before they completed the MSQ. The whole procedure took approximately half an hour.

A second instrument was used to assess participants on effort and persistence, because it is well known that self reports and students’ responses to questionnaires may not always adequately reveal mental processes and behaviour (Biggs, 1993; Nisbett & Wilson, 1977; Schwartz, 1999; Watkins, 1996). The goal was to obtain more reliable data on study behaviour. The instrument consisted of two rating scales ranging from zero to ten. It was distributed to the tutors of tutorial group meetings. These are weekly two hour sessions supervised by a tutor, in which the students discuss the subject matter. The sessions are an essential part of the course. The tutors were given instructions on how to infer students’ effort and persistence. They were told what was meant by effort and persistence, examples were given, and they were told how to use the rating scale (grades ranging from zero to ten are customary in our education). This came down to instructing them to ask and register whether students attended the lectures, whether students were prepared for the tutorial group meetings, and whether students were actively involved in the discussion during the obligatory meetings. The tutors had to convert their impression concerning these aspects into a grade called effort. Persistence was analogously a grade based on the tutors’ judgement concerning whether students continued asking questions during the meetings until they really understood the subject matter, whether students at home persisted in trying to solve their assignments by using lecture notes and/or their books, or whether they consulted their teacher when they were not able to solve an assignment. The participants were evaluated by their tutors in the week before the end of the course. Finally, the scores on the exam at the end of the course were used as an indicator for participants' achievements. The exam consisted of 30 multiple choice questions and grades could range from zero to ten. Example questions of the exam are presented in the appendix.

3.3. ANALYSIS

Sum scores of the responses to the questionnaire were computed for each subscale by summing the scores of individual items. Some items were positively phrased, others negatively. Responses on the negatively phrased items were mirrored so that all answers were in the same direction. The sum scores were called: *Stable Explanation, Control, Outcome Expectancy, Affect, Effort*, and *Persistence.* To reflect the facts that people seek causes especially for failure (Graham & Weiner, 1987) and that motivation to study statistics is usually modest, the coding on the variables *Stable Explanation* and *Control* was done in such a way that high scores corresponded with respectively a stable negative explanation and lack of control. Cronbach's α was computed for each subscale. The exam grades (*Achievement*) and the tutor ratings *Effort(T)* and *Persistence(T)* consisted of grades ranging from zero to ten. They were included into the analyses as raw data.

Four analyses were done. First, several t tests were done to test for possible selection biases. A comparison was done between the male and female participants on *Achievement, Stable Explanation, Control, Outcome Expectancy, Affect, Effort(T),* and *Persistence(T)*. Moreover, achievement was compared between the participants in our study and the rest of the cohort that took the introductory course. Second, bivariate correlations between all variables were calculated to inspect the correlation patterns. The covariance structure modelling was, because of the rather small sample size, done in two separate steps (Scott Long, 1983), resulting in the third and fourth analysis. The third analysis was a robust maximum likelihood confirmative factor analysis (the simultaneous analysis of the covariance and the asymptotic covariance matrix; Jöreskog & Sörbom, 1989), which was done to confirm the measurement structure. Fourth, a path analysis (a robust maximum likelihood structural equation modelling) was done with Lisrel. Due to the sample size it was necessary to disregard the measurement structure in this analysis. Hence, the analysis was done without latent variables and the sum scores of the separate items of the MSQ served as manifest variables. With this path analysis the model presented in Figure 1 was tested.

4. RESULTS

From the pilot study, Cronbach's α for each subscale (after the removal of the four items) and some example questions are presented in Table 1.

A robust maximum likelihood confirmatory factor analysis was executed on those data of the MSQ that were also used in the path analysis of the main study (n = 94). The content based classification of the items on the subscales *Control, Stable Explanation*, *Outcome Expectancy,* and *Affect* was supported by the results of this confirmatory factor analysis; indices showed a proper fit. The Satorra-Bentler chi-square was used. It is considered to be more robust against a small sample size and violations of distributional assumptions (Hu, Bentler, & Kano, 1992; Satorra & Bentler, 1994).

Table 1. Subscales of the MSQ (n = 200)

Subscales and example questions	*Number of items*	*Cronbachs's α*
Stable explanation: Statistics is just a difficult subject. I have always had difficulties with statistics.	4 items	.8427
Control: The result on the statistics exam is determined by my own endeavour. Whenever I don't understand a statistical topic, I know what to do.	5 items	.7797
Outcome Expectancy: It pays off to study statistics. The time I spend on statistics is wasted.	6 items	.6048
Affect: To study statistics is enjoyable. I think statistics is interesting.	8 items	.7813
Effort: I spend a lot of time on statistics. I never prepare myself for the statistics tutorial group meeting.	8 items	.8058
Persistence: Whenever I don't understand something from statistics, I quit. When I cannot complete a statistics assignment, I go through the book once again.	7 items	.7405

The Lisrel program provides several additional indices for how well the model fits the data (Jöreskog & Sörbom, 1988). A goodness of fit index (GFI) is given for the whole model. It compares the tested model with a so called null-model, that is, all parameters are fixed on zero. A second index is the normed fit index (NFI), which compares the tested model with an independence model (variances are set free, covariances are fixed on zero). This index, however, continues to improve when paths are added and therefore does not appraise parsimonious models adequately. The most meaningful index is the non-normed fit index (NNFI). In this index the degrees of freedom are taken into account and consequently it appraises not only the best fitting, but also the most parsimonious model. All three fit indices should be close to one. Finally the root mean square residual (RMR) is given. This index, as the residuals, is ideally close to zero. The indices presented in Table 2 show a proper fit for this model; that is, the items adequately fit into their subscales.

Table 2. Fit indices for the confirmatory factor analysis on Control, Stable Explanation, Outcome Expectancy, and Affect

Satorra-Bentler chi-square ($df = 224$, $n = 94$)	277.18; $p = .04$*	GFI	.86	Standardised RMR	.22
		NFI	.89		
		NNFI	.93		
		CFI	.94		

* $p < 0.05$

In Table 3 descriptive statistics of all the variables as measured by the MSQ, as well as the tutor ratings and the exam grades are given.

Table 3. Descriptives of the motivational variables and achievement

	Mean	*SD*	*Items*	*Scale min*	*Scale max*	*Min score*	*Max score*	*Skewness*	*Kurtosis*
Stable Explanation	16.39	5.61	4	4.00	28.00	4.00	28.00	.100	-.459
Control	16.22	4.84	5	5.00	35.00	5.00	31.00	.697	.637
Outcome Expect	29.13	5.13	6	6.00	42.00	14.00	40.00	-.562	.826
Affect	26.66	7.60	8	8.00	56.00	12.00	51.00	.196	.174
Effort	37.88	7.62	8	8.00	56.00	16.00	54.00	-.477	.347
Persistence	31.93	6.48	7	7.00	49.00	11.00	46.00	-.207	.293
Effort(T)	7.11	1.55	4	0.00	10.00	2.00	10.00	-.803	1.212
Persistence(T)	6.64	1.73	4	0.00	10.00	1.00	10.00	-1.028	1.862
Achievement	7.05	1.90	30	0.00	10.00	1.60	9.40	-.780	-.133

The results of the *t* tests showed no significant differences between female and male participants. This might partly be because of the restricted power of the tests, so additionally the effect sizes (Cohen's *d*) were computed. The results are respectively for *Achievement* (d = .13; p = .65), *Control* (d = .17; p = .51), *Stable Explanation* (d = .53; p = .08), *Outcome Expectancy* (d = .22; p = .49), *Affect* (d = .008; p = .97), *Effort(T)* (d = .02; p = .94), and *Persistence(T)* (d = .20; p = .52). Combined, these results indicate no substantial differences between male and female participants. An additional *t* test was done to test for another possible selection bias. In this *t* test the achievement of the students who participated in our study was compared to the rest of the cohort (n = 122). No significant difference was found, nor a consequential effect size (p = .82; d = .06).

A correlation matrix of all variables was computed and is presented in Table 4. The significance level was adjusted with a Bonferroni correction. Both dimensions of attribution (*Stable Explanation* and *Control*) were significantly correlated to *Outcome Expectation*. The notion of having no control was most strongly correlated to *Outcome Expectation*. *Outcome Expectation* was significantly correlated with *Affect* toward statistics.

Affect was significantly correlated to *Achievement*, but as expected not to the self-reported behavioural constructs (*Effort* and *Persistence*), which were also not correlated to *Achievement*. The tutor ratings *Effort(T)* and *Persistence(T)* on the other hand were much better predictors for *Achievement* and were more highly correlated to *Affect*. This is consistent with research that established the inaccuracy of self-reports and research that showed that students' responses to questionnaires may not always adequately reveal their own learning (Biggs, 1993; Glenberg, Sanocki, Epstein, & Morris, 1987; Nisbett & Wilson, 1977; Schwartz, 1999; Watkins, 1996).

A path analysis with Lisrel was conducted, because of this above-mentioned inaccuracy of self-reports, on a model where the tutor ratings *Effort(T)* and *Persistence(T)* were inserted instead of the self-reported study behaviour (*Effort* and *Persistence*). We started with our model that was presented in Figure 1. The relation between *Stable Explanation* and *Outcome Expectancy* based on attributional theories was not significant (Standardised Path coefficient β = .06; p = .31). We did find a strong negative relation between the notion of having no control (*Control*) and *Outcome Expectancy* (β = -.68; p < .001). Apparently, if a student thinks that there is no contingency between, for example, his study activities and the result on an exam, he will not expect a positive outcome of his actions.

Table 4. Correlations between the motivational variables and achievement.

	Stable Explanation	*Control*	*Outcome Expectancy*	*Affect*	*Effort*	*Persistence*	*Effort(T)*	*Persistence(T)*	*Achievement*
Stable Explanation	1	.584* $p < .001$	-.336* $p = .001$	-.550* $p < .001$	.156 $p = .067$	-.052 $p = .310$	-.116 $p = .132$	-.138 $p = .093$	-.392* $p < .001$
Control		1	-.647* $p < .001$	-.306 $p = .003$	.152 $p = .072$	-.099 $p = .172$	-.020 $p = .423$	.016 $p = .439$	-.121 $p = .123$
Outcome Expectancy			1	.312* $p = .001$	.020 $p = .424$	.157 $p = .065$	.127 $p = .112$	-.006 $p = .479$	.226 $p = .015$
Affect				1	.125 $p = .115$	.239 $p = .010$	.266 $p = .005$	.216 $p = .018$	.429* $p < .001$
Effort					1	.746* $p < .001$	.273 $p = .004$	.263 $p = .005$	.261 $p = .006$
Persistence						1	.368* $p < .001$	.337* $p = .001$	.294 $p = .002$
Effort(T)							1	.843* $p < .001$	.455* $p < .001$
Persistence(T)								1	.478* $p < .001$
Achievement									1

* $p \leq 0.001$ (Bonferroni corrected)

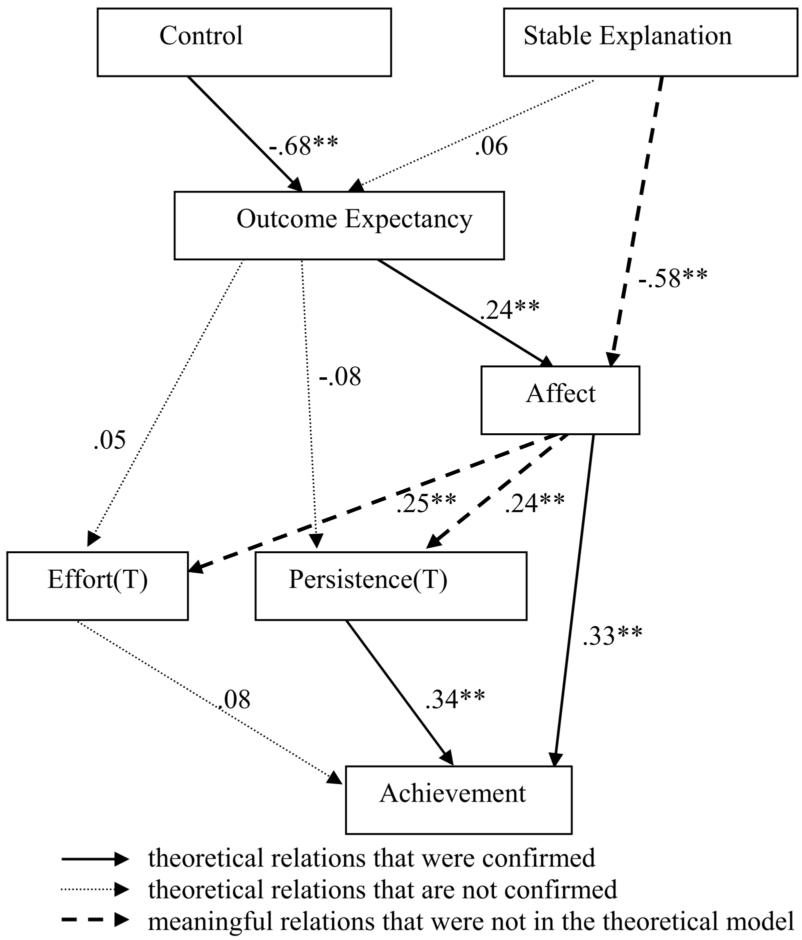

Figure 2. Statistics motivational model, as confirmed by path analysis (Lisrel)
Notes: Coefficients are standardised; **p* < 0.05; ***p* < 0.01.

The relations among the motivational constructs as well as the coefficients are displayed in Figure 2. The solid arrows in Figure 2 stand for the theoretical relations that were confirmed, the dotted arrows stand for the theoretical relations that were not confirmed, and the dashed arrows indicate meaningful relations that were not in the hypothesised theoretical model, as shown in Figure 1.

Figure 2 shows a strong direct relation between *Stable Explanation* and *Affect*, that is, if students think that there are stable causes for negative statistics related events, failing their exams for example, they will develop negative feelings toward statistics. In the model, as displayed in Figure 1, this relation was mediated by *Outcome Expectancy*.

A negative *Outcome Expectancy* also had an adverse effect on *Affect*. *Affect* is related to all other constructs except to the notion of no control (*Control*) ($\beta = .19$; $p = .08$). To emphasise the importance of *Affect*, it has been placed in a more central position in Figure 2. It is strongly related to *Achievement* directly, as well as via *Persistence(T)*. Also important is that *Achievement* is determined by *Persistence(T)* ($\beta = .34$; $p < .001$) but not by *Effort(T)* ($\beta = .08$; $p = .36$).

To enhance the fit of the model, the residuals of the behavioural constructs *Effort(T)* and *Persistence(T)* had been set free to correlate (error covariance = 2.12; $t = 5.34$) in

Lisrel. With this relaxation of the model (as presented in Figure 2), all fit indices showed a good fit. The values of these indices for our model are provided in Table 5. Again the Satorra-Bentler chi-square is presented because of its robustness against a small sample size and violations of distributional assumptions (Hu et al., 1992; Satorra & Bentler, 1994).

Table 5. Fit indices for the model in Figure 2

Satorra-Bentler chi-square (df = 7, n = 94)	13.40; p = .063	GFI	.96	Standardised RMR	.042
		NFI	.95		
		NNFI	.93		
		CFI	.98		

5. DISCUSSION

This study was done in an introductory statistics course. It focussed on causal explanations of statistics related events, perceived outcome expectancy of students' activities within this statistics course, affect and study behaviour toward statistics, and the relation of these constructs to the results on the exam at the end of the course. These constructs were chosen because of their practical implications for the teaching of statistics.

Our first findings concern causal explanations. In the two presented motivational theories, perceived causes for events have underlying properties that have affective, behavioural, and cognitive consequences (Peterson et al., 1993; Pintrich & Schunk, 1996). In our study we focused on the dimensions of control and stability of causal explanations.

The first result concerns control. The model in Figure 2 indicates that the perception of having no control over causes of statistics related events may lead to decreased outcome expectancy. For example, a student who thinks that there is nothing he can do about the causes for failing the statistics exams, or thinks that he is not able to understand statistics anyway, may not expect a positive outcome from attending the lectures or studying the material. This mechanism is intuitively appealing.

The second result indicates that the stability of causal explanations may be more directly related to affect. As is seen in Figure 2 we found a significant path from *Stable Explanation* of such causes to *Affect*. The path that we found may be interpreted as follows. The perception of stable causes for aversive events related to statistics may lead to displeasure and frustration. If students perceive that failing statistics exams is not easily changeable, students may start to dislike statistics. This was reflected in responses like: *I dislike statistics; I do not have a positive perception of statistics*; and so forth.

In sum, these two findings indicate that students who think that they lack control may not expect to profit from studying statistics, and students who do invest time but think that there are stable causes for failing in spite of that, may start to dislike statistics.

The last path from *Stable Explanation* to *Affect*, though intuitively appealing, was not anticipated. The model in Figure 1 contained a relation between *Stable Explanations* and *Outcome Expectancy*. This relation was based on the general attributional position that the stability of a cause has the most influence on shifts in expectancy (Pintrich, 2000; Pintrich & Schunk, 1996; Weiner, 1986, 1992). Our findings are more consistent with the basic assumption from Peterson et al. (1993) that controllability is the major factor influencing outcome expectancy. Yet, the direct influence of *Stable Explanation* on *Affect* may also have important practical implications for statistics education.

The implication for education from our findings may be that when students discover the material is comprehensible to them and they experience success, they will be stimulated to study the material. This means that in constructing a learning environment, there should be tasks built in that are feasible for students. In that way the sequence of events that may lead to diminished motivation (Weiner, 1986) may be interrupted. Students will gradually sense that they can master the topics, they will discover they can control their learning outcomes, they will experience success, and they will abandon the idea that there are stable causes for failure. Control over learning outcomes may foster the positive expectation of future study activities. This positive expectation, together with the reduction of the perception of stable negative causes for failure, may even promote students to enjoy studying statistics. Only then should more difficult tasks be administered.

A second finding of interest in our study seems to be the central position of *Affect* in our model in Figure 2. Students who appreciate the value and relevance of statistics, who think it is interesting, challenging, and who like statistics, appear to study statistics more and qualitatively better, and perform better on the exams. In attributional theories (Pintrich, 2000; Pintrich & Schunk, 1996; Weiner, 1986, 1992) as well as in the learned helplessness theory of Peterson et al. (1993), affect is on the same level as behaviour and cognition. In the model in Figure 1 *Affect* was therefore put on par with behavioural consequences of *Outcome Expectancy*. However, affect seems to have a more prominent role in motivational processes in the present statistics education context. In our study we found that *Affect* directly and positively influenced *Achievement*. It also influenced study behaviour, namely *Effort(T)* and *Persistence(T)*. *Persistence(T)* in turn also influenced *Achievement*. Thus, *Affect* seems to determine achievement directly, as well as indirectly. Moreover, we found that *Affect* functioned as a mediator between *Control*, *Stable Explanations,* and *Outcome Expectancy* on the one hand, and the rest of the motivational constructs on the other. For this reason *Affect* holds a more central position in our model in Figure 2 than in the model presented in Figure 1.

The central role of *Affect* suggests that the students' feelings toward statistics appear to be an important theme for innovating and improving statistics education. Our results with respect to *Affect* are in line with Malone and Lepper (1987), who state that implementing features that make learning more appealing, enjoyable, and challenging makes learning more intrinsically motivating. Our finding that the feelings toward statistics are crucial in reaching satisfactory achievements corroborates the results of Isen, Daubman, and Gorgoglione (1987). In their study they found that positive affect may foster student's tendencies to see relations among stimuli, because positive affect leads to different ways of information processing, for example using different strategies. More relations between concepts are characteristic for richer knowledge networks, which indicate better integrated knowledge and deeper understanding (Kintsch, 1988, 1998).

It seems to be of relevance in the improvement of statistics education to make statistics courses more attractive, interesting, and enjoyable. One of the ways this might be achieved is by making the courses less theoretical. We think that a small experiment may engage students in a more active way, it may be fun to analyse data that are collected by the students themselves, and it may foster the notion of relevance of statistics.

A final result in our study was that *Effort(T)* had no significant relation with *Achievement*. Both *Effort(T)* and *Persistence(T)* were determined by the tutors. *Effort(T)* reflected the amount of time students studied, and whether students prepared themselves, attended lectures, or were actively involved in the discussion during group meetings.

Effort per se seemed to have a minor effect on achievement. What counts seems to be the way students study. In our study, *Persistence(T)* contributes significantly to exam performance. Students who did not quit that easily, who persisted, who turned to their lecture notes or their books, or consulted a teacher when they were not able to solve a statistical problem, those students did better on the exam. This result suggests that persisting is the best way to study statistics. It is in line with research in other subjects that established the importance of learning strategies and mastery goals for achievement in educational settings. (see e.g., Ames, 1992; Boekaerts; 1997; Pintrich, 2000; Dweck, 2000).

This finding may also be important for educational purposes. In the teaching of statistics, students should be stimulated to try to solve their problems. They should try to persist instead of quitting all too easily. This can be done by guiding them through the topics and by pointing them in the correct direction, instead of giving the solution to a problem promptly. Persisting and learning from mastering their own difficulties may be the most valuable way of learning.

The student population from which we recruited our participants consists largely of female students. Consequently, most of our participants were female (79 female versus 15 male). This could have affected our results. However, *t* tests on all the core variables (*Control, Stable Explanation, Outcome Expectancy*, *Affect, Effort(T)*, *Persistence(T),* and *Achievement*) in our models showed no significant differences between the female and male students. Therefore, the fact that the majority of our participants was female seems not to affect the motivational processes that were studied.

The tutor ratings that we used to measure effort and persistence are another limitation of our study. We instructed the tutors in great detail and asked them to record students' activities that we hold indicative for effort and persistence. We are confident that the ratings of the tutors are a quite valid and reliable measurement of the relevant behaviour. Still these ratings only reflect observable, external behaviour. Consequently we cannot discuss internal processes of reflection and mental activity. Our results only pertain to self-reported cognitions, affect, and observed behaviour.

In the present study only first-year students were studied. In future research second- and third-year students could be studied. Secondly, our results could be corroborated in studies with a larger sample. In our study a rather small sample was used (n = 94). It could also be investigated how in a practical educational context we can determine whether students persist during studying statistics. How can students optimally be guided to the correct solution of the problems? Will this reduce the perception of stable negative causes for failure and enhance the notion of control? Will such a reduction lead to a positive expectation of future study activities and to more enjoyment? Will all this eventually lead to more persistence and better results on the exam? Finally, further research is needed to investigate additional ways statistics education can be made more enjoyable. In the past, our department spent most attention on how to make lectures more informative, to select the best instruction books, and to develop assignments that are mainly educational. Now our attention has somewhat shifted toward making the courses more attractive, interesting, and enjoyable. We have tried to make the courses less theoretical by introducing a small experiment. Even so, future research may include investigating the most effective ways of making statistics education more enjoyable.

REFERENCES

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261-271.

Biggs, J. B. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*(1), 3-19.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161-186.

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction.* Upper Saddle River, NY: Merrill, Prentice Hall.

Chi, M. T. H., de Leeuw, N., Chiu, M-H., & LaVancher, C. (1994). Eliciting self-explanations improves learning. *Cognitive Science, 18*(3), 439-477.

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

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* Philadelphia: Psychology Press.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal, & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp 37-51). Amsterdam: IOS Press and the International Statistical Institute.

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

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

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology, 116*(2), 119-136.

Graham, S., & Weiner, B. (1987). Some educational implications of sympathy and anger from an attributional perspective. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction. Volume 3: Conative and affective process analyses* (pp. 199-221). Hillsdale: Lawrence Erlbaum Associates.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*(2), 351-362.

Isen, A. M., Daubman, K. A., & Gorgoglione, J. M. (1987). The influence of positive affect on cognitive organisation: Implications for education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction. Volume 3: Conative and affective process analyses* (pp. 143-164). Hillsdale: Lawrence Erlbaum Associates.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications.* Chicago: Scientific Software, Inc.

Keeler C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, *3*(2).
[Online: www.amstat.org/publications/jse/v3n2/keeler.html]

Kelly, G. A. (1958). Man's construction of his alternatives. In G. Lindzey (Ed.), *Assessment of human motives* (pp. 33-64). New York: Grove.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition.* Cambridge: Cambridge University Press.

Magel, R. C. (1998). Cooperative learning and statistics instruction. *Journal of Statistics Education*, *6*(3).

[Online: www.amstat.org/publications/jse/v6n3/magel.html]
Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction. Volume 3: Conative and affective process analyses* (pp. 223-253). Hillsdale: Lawrence Erlbaum Associates.
Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*(1), 3-53.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.
Peterson, C., Maier, S. F., & Seligman, M. E. P. (1993). *Learned helplessness. A theory for the age of personal control.* New York: Oxford University Press.
Phye, G. D. (1997). *Handbook of academic learning: Construction of knowledge*. San Diego: Academic Press.
Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P .R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego: Academic Press.
Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research and applications.* Upper Saddle River, NY: Merrill, Prentice Hall.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis: Application for developmental research* (pp. 399-419). Thousand Oaks: Sage.
Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement,55*(5), 868-875.
Schwartz, N. (1999). Self-reports. How the questions shape the answers. *American Psychologist, 54*(2), 93-105.
Scott Long, J. (1983). *Covariance structure models.* Newbury Park: Sage Publications.
Seegers, G., & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction, 3*(2), 133-150.
Steffe, L. P., & Gale, J. (1995). *Constructivism in education.* Hillsdale: Erlbaum.
Valås, H., & Søvik, N. (1994). Variables affecting students' intrinsic motivation for school mathematics: two empirical studies based on Deci and Ryan's theory on motivation. *Learning and Instruction, 3*(4), 281-298.
Volet, S. E. (1997). Cognitive and affective variables in academic learning: The significance of direction and effort in students' goals. *Learning and Instruction, 7*(3), 235-254.
Watkins, D. (1996). The influence of social desirability on learning process questionnaires: A neglected possibility? *Contemporary Educational Psychology, 21*(1), 80-82.
Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer Verlag.
Weiner, B. (1992). *Human motivation. Metaphors, theories, and research.* Newbury Park: Sage Publications Inc.

LUC BUDÉ
Department of Methodology and Statistics
Maastricht University
PO Box 616
6200 MD Maastricht

APPENDIX

Example questions from the exam at the end of course (used for the measurement of achievement).

1. In a sample of 101 newborn babies, the mean birth weight is 3.8 kg and the standard deviation is 0.85. The null hypothesis is H_0: $\mu = 4$ kg.
   If this null hypothesis holds, then:

a. The probability that we will find a sample mean smaller than or equal to 3.8 kg is 50%
b. The probability that we will find a sample mean smaller than or equal to 3.8 kg is 80%
c. The probability that we will find a sample mean smaller than or equal to 3.8 kg is less than 50%
d. The probability that we will find a sample mean smaller than or equal to 3.8 kg is greater than 50%

2. Given the same sample as in question 1, we are testing H_0: $\mu = 4$ kg against H_1: $\mu \neq 4$ kg. The *p*-value of the sample mean of 3.8 kg is:

a. $p \leq .01$
b. $.01 < p \leq .02$
c. $.02 < p \leq .05$
d. $p > .05$

3. Given the same sample as in question 1, we are again testing H_0: $\mu = 4$ kg against H_1: $\mu \neq 4$ kg. Suppose the null hypothesis is rejected at $\alpha = .10$. What is the implication of this $\alpha = .10$?

a. In 10 % we will wrongfully conclude that H_0: $\mu = 4$ kg holds.
b. In 10 % we will wrongfully conclude that H_1: $\mu \neq 4$ kg holds.
c. In 5 % we will wrongfully conclude that H_0: $\mu = 4$ kg holds.
d. In 5 % we will wrongfully conclude that H_1: $\mu \neq 4$ kg holds.

4. The effects of 3 instructional methods on comprehensibility of the information (SCORE) were investigated. The 3 methods were: a standard method and 2 experimental methods (experimental method 1 and experimental method 2). The coding of the dummy variables was as follows:

	D_EXP1	D_EXP2
Standard method	0	0
Experimental method 1	1	0
Experimental method 2	0	1

It is tested whether the comprehensibility of the information (SCORE) for all methods is equal (H_0), or if at least one of the three methods is different (H_1).
Part of the output of the SPSS analysis is presented below:

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3341.722	2	1670.861	6.317	.005
	Residual	8727.917	33	264.482		
	Total	12069.639	35			

Predictors: (Constant), D_EXP2, D_EXP1
Dependent Variable: SCORE

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	37.750	4.695		8.041	.000
	D_EXP1	14.250	6.639	.367	2.146	.039
	D_EXP2	23.417	6.639	.603	3.527	.001

Dependent Variable: SCORE

Question: What conclusion can be drawn? Assume $\alpha = 0.05$.

a. There is a difference between the instructional methods because the *p*-value of *F* is smaller than 0.05.
b. There is a difference between the instructional methods because the *p*-value of *F* is smaller than 0.05/2 =0.025.
c. There is no difference between the instructional methods because the *p*-value of *F* is smaller than 0.05/2 =0.025.
d. There is no difference between the instructional methods because the *p*-value of *F* is smaller than 0.05.

Given the same research and the same results as in question 4, suppose that the F-test indicates a difference between the three methods. Which groups differ significantly? Assume $\alpha = 0.01$.

a. Each method differs significantly from the others.
b. The standard method differs significantly from experimental method 1.
c. The standard method differs significantly from experimental method 2.
d. The experimental method 1 differs significantly from experimental method 2.

5. Given the same research and the same results as in question 4, what is the proportion of explained variance in the SCORE variable?

a. 0.28
b. 0.72
c. 0.38
d. 0.86

USE OF EXTERNAL VISUAL REPRESENTATIONS IN PROBABILITY PROBLEM SOLVING

JAMES E. CORTER
Teachers College, Columbia University
corter@exchange.tc.columbia.edu

DORIS C. ZAHNER
Teachers College, Columbia University
dwc14@columbia.edu

ABSTRACT

We investigate the use of external visual representations in probability problem solving. Twenty-six students enrolled in an introductory statistics course for social sciences graduate students (post-baccalaureate) solved eight probability problems in a structured interview format. Results show that students spontaneously use self-generated external visual representations while solving probability problems. The types of visual representations used include: reorganization of the given information, pictures, novel schematic representations, trees, outcome listings, contingency tables, and Venn diagrams. The frequency of use of each of these different external visual representations depended on the type of probability problem being solved. We interpret these findings as showing that problem solvers attempt to select representations appropriate to the problem structure, and that the appropriateness of the representation is determined by the problem's underlying schema.

Keywords: *Statistics education research; Probability problem solving; Visual representations; Trees; Outcome listings; Venn diagrams*

1. INTRODUCTION

Consider the following probability problem:

An apartment building has four parking spaces in front (call them A, B, C, and D). There are four apartments in the building (#1, #2, #3, and #4), and each apartment has a single occupant with a single car. Every evening, all four occupants come home and park in a randomly chosen space. What is the probability that this evening they park so that the occupant of Apt #1 is in space A, the occupant of #2 is in space B, the occupant of #3 in space C, and the occupant of #4 in space D?

How would you go about solving this problem? Many people report visualizing the cars and parking spaces. After that, strategies for solving the problem tend to diverge (as do success rates). One of our points in presenting this problem (used in the present study) is that probability word problems are often simple to pose, yet difficult for many students to solve. Another point is that visualization and visual solution methods, such as self-generated external pictures and diagrams, can be very helpful in solving some probability problems.

Statistics Education Research Journal, 6(1), 22-50, http://www.stat.auckland.ac.nz/serj

Probability problem solving (PPS) can be quite difficult for students (Garfield & Ahlgren, 1988; Konold, 1989; O'Connell, 1993; Pollatsek, Well, Konold, Hardiman & Cobb, 1987), even when the mathematics involved is simple. Of course, other types of mathematics word problems are also difficult for many students, perhaps because solving them requires the problem solver to think abstractly about situations, and then model these situations using mathematical concepts. However, some researchers (Garfield & Ahlgren, 1988; Konold, 1989) have suggested that probability problem solving may be especially difficult because people have natural misconceptions about probabilistic concepts (e.g., Kahneman, Slovic, & Tversky, 1982).

Recommendations have been made for how to teach concepts in probability (e.g., Bantanero, Godino, & Roa, 2004; Gelman & Nolan, 2002; Gigerenzer, 1994; Keeler & Steinhorst, 2001; Konold, 1995, 1996; Sedlmeier & Gigerenzer, 2001). However, as pointed out by Garfield and Ahlgren (1988), only a few articles have tried to gather empirical evidence on the processes by which students solve probability problems. In one such study, O'Connell (1999; O'Connell & Corter, 1993) described a pedagogical model of recommended process steps by which students should solve probability problems. O'Connell (1993, 1999) classified student errors in probability problem solving, showing that they could be grouped into several categories: text comprehension errors, conceptual errors, procedural errors, and computational errors. Konold, Pollatsek, Well, and Lohmeier, and Lipson (1993) documented inconsistencies in probabilistic reasoning and discussed implications for probability problem solving. Due to this paucity of research on PPS, Chance and Garfield (2002) call for more research on the cognitive processes of probability problem solvers using innovative methods such as videotaped clinical interviews. The present study is intended as a step in that direction.

1.1. IS THERE A SPECIAL ROLE FOR VISUALIZATION IN PROBABILITY PROBLEM SOLVING?

In studying the cognitive processes of probability problem solvers, one issue that deserves special attention is the role of visualization. After all, anecdotal evidence suggests that visualization plays an important role in how experts solve probability problems (and mathematics problems generally). Also, informal observations of how students in statistics courses solve probability problems provide ample evidence that they sometimes spontaneously use visual devices (e.g., outcome trees) in their written work. Finally, Sedlmeier (2000) has suggested that common cognitive "fallacies" in reasoning about conditional probabilities may be ameliorated by graphical representations. Visualization may be especially important for probabilistic reasoning and probability problem solving because of the inherently abstract nature of the concepts introduced in probability.

To better understand the literature on visualization in mathematics problem solving, it is important to distinguish between *internal* visual representations (i.e., "mental imagery") and *external* visual representations (e.g., graphs, charts, pictures, etc.). Another distinction about the way external representations may be used in problem solving concerns whether the external representations are provided to the student by an instructor or experimenter, or are spontaneously generated by the student in the course of solving the problem. Although there is an extensive literature on how instructor-provided graphics can aid in scientific problem solving (summarized below), there has been little or no research on students' spontaneous creation and use of pictures, graphics and other visual devices in the course of mathematics problem-solving activities. In the present study, we use written and think-aloud protocols to study when and why probability problem solvers spontaneously produce external visual representations in their written work (when not required to do so), and what types of visual representations they employ.

1.2. PREVIOUS RESEARCH ON EXTERNAL VISUAL REPRESENTATIONS AND PROBLEM SOLVING

Results from previous research on scientific problem solving by schoolchildren (e.g., Lehrer & Schauble, 1998; Penner, Giles, Lehrer, & Schauble, 1996) and by high school and college students (e.g., Hall, Bailey, & Tillman, 1997; Hegarty & Just, 1993; Kaufmann, 1990; Mayer, 1989; Mayer & Anderson, 1991, 1992; Mayer & Gallini, 1990; Mayer, Mautone, & Prothero, 2002; Molitor, Ballstaedt, & Mandl, 1989; Santos-Trigo, 1996; B. Tversky, 2001) suggest that experimenter-provided external visual representations can aid scientific problem solving. The visual representations investigated in these studies ranged from diagrams that accompanied text (Hall, Bailey, & Tillman, 1997; Mayer, 1989; Mayer & Anderson, 1991, 1992; Mayer & Gallini, 1990; Mayer, Mautone, & Prothero, 2002) to actual physical models of scientific systems (Lehrer & Schauble, 1998; Penner, Giles, Lehrer, & Schauble, 1996). In spite of the wide range of external visual representations used in these studies, a common finding was that experimenter-provided external visuals often facilitate problem-solving success. Many of the studies also conclude that such external visual representations can aid in the development of student understanding of physical systems and mechanisms.

Incidentally, it is likely that individuals vary in the extent to which they use and benefit from visual representations. Some researchers in this area (e.g., Hegarty & Kozhevnikov, 1999; Kozhevnikov, Hegarty, & Mayer, 2002) have taken an individual differences perspective, grouping problem solvers into one of several types: those who tend to use verbal representations, and those who primarily use visual/spatial representations. Kozhevnikov et al. (2002) suggest that the visualizer group can be further split into object visualizers and spatial visualizers, with spatial visualizers showing some advantages in scientific and mathematical tasks.

Research conducted specifically in the domain of mathematics has also shown that experimenter-provided external visual representations can be useful in mathematics problem solving (e.g., Sedlmeier & Gigerenzer, 2001; Koedinger & Anderson, 1997; Nemirovsky, 1994). In particular, a number of studies (e.g., Hollebrands, 2003; Hannafin, Burruss, & Little, 2001; Hannafin & Scott, 1998) have found that the use of Geometer's Sketchpad®, a geometry graphing computer program, can be helpful in developing students' concepts and problem solving in geometry. Schwartz and Martin (2004) investigated the use of graphical tools in statistics instruction and found that experimenter-prompted graphical "invention activities" by students led to significant gains in understanding of statistical concepts.

Previous work (e.g., Russell, 2000; Zahner & Corter, 2002) in our own lab has shown that most probability problem solvers choose to use external visual representations while solving problems (after being taught the use of such visuals in an introductory statistics course), and that a wide variety of such external visual devices are used. External visual representations used by probability problem solvers include at least these types: graphs, tree diagrams, contingency tables, Venn diagrams, and pictures. Arguably, formulas and mathematical symbols could be included in this list, because they incorporate visuospatial relationships (cf. Presmeg, 1986). However, their usefulness in solving probability problems is not in question.

1.3. WHY ARE EXTERNAL VISUAL REPRESENTATIONS USEFUL IN PROBLEM SOLVING?

In order to use research results on visualization and problem solving to improve mathematics teaching and learning, it is important to ask *why* external visual representations are useful in mathematics problem solving. One possible answer to this question is that external visual representations help to augment cognitive capabilities in certain ways (e.g., Lowrie & Kay, 2001; Novick, 2001; Qin & Simon, 1995), for example by aiding memory. Tversky (2001) lists a number of possible functions of external diagrams and visual devices, including attracting attention, recording information and supporting memory, communication, providing models, and facilitating inference and discovery. Another possibility is that using multiple representations of a problem (including visual ones) leads to a fuller understanding of the problem and an increased "depth of processing" (Logie & Baddeley, 1990; Mayer, 1989, 2001; Mayer & Gallini, 1990). Other potential explanations for the use of external visual representations include the possibility that such representations can help problem solvers build a mental model of the described problem situation (Schwartz & Black, 1996). Finally, for certain problems the graphical devices may be used as a solution tool in a more specific way: for example, reading a value from a graph, or counting outcomes in an outcome tree. Alternatively, it might be that there is no benefit in using external visual representations, rather their use is just an epiphenomenon, a reaction to training from classroom instruction.

Of course, these accounts of why visualizations might be useful are not all mutually exclusive or contradictory. But only fragmentary data exist that might support or discredit any of these explanations. Some hints might come from studies examining *when* problem solvers choose to use external visual representations. For example, there is some evidence that both internal (Hampson & Morris, 1990) and external (Lowrie & Kay, 2001; Zahner & Corter, 2002) visual representations tend to be used more for unfamiliar or more difficult problems. This observation seems to support certain explanations (e.g., visuals as supporting memory, or facilitating inference and discovery) more than others.

1.4. THE PRESENT STUDY

This study focuses on the use of external visual representations in probability problem solving (PPS). We are interested in what types of problems tend to elicit use of visual representations, how and when external visual representations are used in PPS, and finally, if external visual representations facilitate correct solution of the problems. We used a variety of problem types, in order to investigate if the usefulness of visuals and the type of visual device chosen by the problem solver depends on specific aspects of the problem being solved. Specifically, we ask: Are particular types of representations used with particular problem topics (for example, problems dealing with permutations)? Also, we investigate if external visual representations are used more often with unfamiliar types of problems, because the student may have a higher cognitive load in these cases, or because the elicitation of a familiar problem-solving schema may be less likely.

As background to the present work, we assume that the process of solving a probability word problem can be broken down into roughly sequential stages (cf. Kintsch & Greeno, 1985; Mayer, 1992; O'Connell, 1993, 1999; O'Connell & Corter, 1993; Reusser, 1996). These stages are assumed to be:

i) initial problem understanding (text comprehension),
ii) formulating the mathematical problem,
iii) finding a solution method or schema,

iv) computing the answer.

Novick and Hmelo (1994) make a more gross distinction between *problem representation* and *solution procedure* phases of problem solving. Consistent with this simpler classification, our coding scheme for written protocols of students did not attempt to code use of visuals separately for stages i-iii, because we do not believe this can be done reliably with the present data. Rather, our scheme coded two types of uses of graphs or other external visual representations: a) for problem understanding, mathematical formulation, or for selection of a solution schema (i.e., any such use in the first three stages above), and b) for any use in the final stage, that of actually computing a numerical answer. We refer to the latter type of use of visual devices under the term "computational method." As an example of the first type of use, consider the use of a picture of a spinner or a Venn diagram to depict aspects of the probability word problem. An example of the second type of use of external visualizations, using them to compute an answer, would be counting the number of outcomes (leaves) in an outcome tree to find the denominator for a probability calculation.

Finally, we are interested in knowing if the use of external visual representations is associated with solution success for these probability problems. If external visual devices are used because they are helpful, then we ought to be able to find evidence of that. However, there are several factors that complicate this relationship, including the student's prior knowledge of the visual devices used, student spatial and mathematical ability, student cognitive style and the difficulty of the problem. Alternatively, it might be that the use of external visual representations is associated with solution failure, because participants might be more likely to use visual representations when they find a problem confusing or difficult (cf. Hegarty & Kozhevnikov, 1999; Lowrie & Kay, 2001).

2. METHOD

2.1. PARTICIPANTS

Twenty-six students were recruited from introductory probability and statistical inference classes during the Fall semester of 2002 from an urban college of education and psychology in the U.S.A. All participants were graduate students (post-baccalaureate) in education and social sciences, with widely varying math backgrounds. Participants were each paid ten dollars. They were informed that they were going to participate in a study of probability problem solving, and that the primary focus of the study was on the methods by which students solve problems. Because all the participants were enrolled in the same introductory statistics class at the college, their recent curricular background in probability problem solving was well-controlled and known, though the degree to which each participant mastered the material in that course was not measured. This course included approximately six lectures in probability. Topics included: events and outcome spaces, definition of probability for equally-likely and unequally-likely events, combinatorics, compound events, conditional probability, independence of events, and Bayes' Rule.

2.2. MATERIALS

Each respondent was asked to solve eight probability problems. This set of eight problems (see Appendix) was designed to include four different probability topics ("problem types") each represented at two different levels of typicality for that topic. The four different problem types were labeled: "*Combinations,*" "*Sequential,*" "*Permutations*," and "*Conditional probability*." The problems representing each of these four topics were thought to have distinct "deep structures" corresponding to four distinct problem schemas, tapping somewhat different sets of knowledge and solution skills. This factor will be referred to as "problem type." As an example, consider Problem P1, an example of a *Combinations* problem:

P1. There are 10 books on Mary's bookcase. She randomly grabs 2 books to read on the bus. What is the probability that the 2 books are "Little Women" and "War & Peace"? (Both these books are on her bookshelf.)

In the curriculum to which these participants had recently been exposed, this problem typically would have been solved using the formula for the number of *combinations of n objects selected k at a time.* That formula gives the number of possible outcomes in the sample space:

$$_nC_k = \frac{n!}{k!(n-k)!} = \frac{10!}{2!(8!)} = 45\,.$$

Then the probability of Mary selecting one particular combination (two specific books) can easily be calculated to be 1/45. Any problem requiring use of this formula (or a variant of it) is therefore assumed to share the same problem schema, and is said to be of the same basic Problem Type (*Combinations* in this case).

For each problem type, there was one typical variant and one atypical variant. The typical version was a problem that could be solved using a straightforward application of a standard probability formula known to have been taught in the participants' introductory statistics class. The atypical version was a problem that was very unlikely to be isomorphic to any problem encountered in their course, and that could *not* be solved using a single application of a standard probability formula. This manipulation of typicality may be clarified by Table 1, which presents summaries of what we judge to be appropriate *formula-based* solutions for the typical and atypical variants of each problem type. These solutions are presented to show the basic underlying structure of each problem type, without reference to surface content, and to illustrate how each problem might be solved by application of one or more standard probability formulas.

Table 1 also makes clear the types of specific problem manipulations that were used to create the atypical variant of each problem type. For the *Combinations* problem, the predicted solution for the typical version requires the problem solver to use the standard formula for the number of combinations of *n* things selected *k* at a time to calculate the number of possible outcomes. Our predicted solution for the atypical *Combinations* problem requires using this formula twice, once in the numerator and once in the denominator. For the *Sequential* problems, problem solvers must use the formula for calculating the probability of three independent events. In the typical variant the three events are identical, whereas in the atypical variant they are different events with differing probabilities. The typical variant of the *Permutations* problem asks how many different ways four items can be matched up with four "slots." The atypical version of this problem asks the same question, but orders the objects only with respect to the first two slots. We consider this an atypical problem because the computational method does not correspond to straightforward application of the formula for number of permutations of *n* objects, which is known to have been taught in the participants' introductory statistics course. Note that this atypical variant is actually simpler computationally than its typical version. For the *Conditional Probability* problems, the typical version closely resembles examples used in the students' introductory statistics course, and requires the problem solver to use the formula for conditional probability (twice). The atypical variant adds a final step, in which the formula must be used a third time.

Table 1. Example formula-based solutions for typical and atypical variants of each problem type

Problem Type	Typical	Atypical
Combinations	$_nC_k = \dfrac{n!}{k!(n-k)!}$ where n = total number of books k = number selected	$P(A) = \dfrac{_mC_k}{_nC_k} = \dfrac{\frac{m!}{k!(m-k)!}}{\frac{n!}{k!(n-k)!}}$ where n = total number of books m = number of novels k = number selected
Sequential	$P(A_1 \cap A_2 \cap A_3) = P(A)P(A)P(A)$	$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$, A_1, A_2, A_3 not necessarily equal
Permutations	# outcomes = $n!$ = $(n)(n\text{-}1)\ldots(2)(1)$ where n = number of objects	# outcomes = $(n)(n\text{-}1)$ where n = number of objects
Conditional Probability	$P(B) = P(A \cap B) + P(A^c \cap B)$ $P(A \cap B) = P(B \mid A) \times P(A)$ $P(A^c \cap B) = P(B \mid A^c) \times P(A^c)$	$P(B) = P(A \cap B) + P(A^c \cap B)$ $P(A \cap B) = P(B \mid A) \times P(A)$ $P(A^c \cap B) = P(B \mid A^c) \times P(A^c)$ $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Across two different forms (A and B) of the test booklet, each of the eight problems was formulated with two different cover stories, or *surface content.* Surface content was counterbalanced with problem typicality across test forms. For example, the typical *Combinations* problem P1 given above involved books on a bookshelf, so a participant who saw that problem would see an atypical *Combinations* problem involving cookies in a cookie jar. For another participant who saw the second test form, the atypical *Combinations* problem would involve books on a bookshelf, and the typical version would involve cookies in a cookie jar. The Appendix shows only test form A.

2.3. PROCEDURE

A structured interviewing protocol (cf. Ginsburg, 1997) was developed for use in the interviews, and was designed mainly to elicit a reasonable level of detail in the participant protocols. The same interviewer worked with all of the participants, interviewing only one participant at a time. Participants were asked to think aloud while solving the problems, and also to show their written work with provided pen and paper. The task was not timed. However, most participants finished in less than an hour. A probability formula sheet was available to them at all times (but left face down), though no participant was observed to use it. A videotape recorder was used to capture the participants' work and student/interviewer comments. The present analyses mainly focus on the participants' written work, though the verbal transcripts were analyzed as well.

The interviewer stepped in with verbal prompts in any of four circumstances, to elicit continuation of the work or more detail about the participant's solution process. The first circumstance was if the participant could not see any way to begin solving the problem. In this situation the script called for the interviewer to ask, "In general, what would be a good first step in solving this problem?" with other follow-up questions ("How would you apply it in this case?") if the first prompt did not elicit useful work. The second type of circumstance in which the interviewer stepped in was when the participant paused for a long time (more than 30 seconds) without thinking aloud or writing. This could indicate either that the participant was thinking silently or was at an impasse, and was responded to with "What are you thinking?" and other follow-up prompts ("Let's back up and look at this again. How else could you solve this?"). The third situation in which prompting occurred was when the participant's verbal or written process explanations lacked sufficient detail, for example, consisting of only a few calculations with no explanation ("Can you explain how you arrived at this?"). The fourth situation eliciting interviewer intervention was when the participant indicated that he or she was finished with the problem. In most cases, this occurred when the participant had arrived at what he or she believed was the correct answer. In other cases, this was because the participant gave up on solving the problem. In either case, the interviewer then asked the participant to explain in detail all of the steps used in the solution attempt.

Coding of the written protocols The focus of the present study is on use of external visual representations in problem solving and on the methods used to solve problems. Thus, the analyses reported in the present study focus mainly on coding of the participants' written work. Three particular aspects of the written problem solutions were coded, based upon a scheme developed in previous research (Russell, 2000). The first coded aspect was whether or not the participant gave the correct answer to the problem. The second aspect coded, described in more detail below, was the type of external visual representation used (if any) by the participant. The third aspect coded, also described in more detail in the next section, was the type of general computational method used to solve the problem. Here the identified types were: formula, graphical, or procedural.

External visual representations Written protocols for each problem solution were coded for use of different types of external visual representations. The coded types of external visual representation included *pictures, outcome listings, trees, contingency tables, Venn diagrams, novel schematic representations*, and *spatial reorganization of the given information*.

An external visual representation was coded as a *picture* if it attempted to represent the real-world situation conveyed in the problem in a non-symbolic, pictorial way. For example, in a problem about use of a spinner with separate areas marked "red," "blue," and so on, any picture of a spinner type device would count as a *picture* (see Figure 1 for an example). A visual device was coded as an *outcome listing* if it gave a list of outcomes in some relevant outcome space, for example: {HH, HT, TH, TT} as the outcomes space for the experiment of flipping a coin twice. A visual representation was coded as a *tree* diagram if the participant attempted to organize the information from the problem in either a complete or a partial outcome tree. An example of the use of a tree diagram is shown in Figure 2. A visual representation was coded as a *contingency table* if the participant presented the information from the problem as probabilities or frequencies in a two-way table. A visual representation was coded as a *Venn diagram* if the participant used a Venn diagram to represent set relationships.

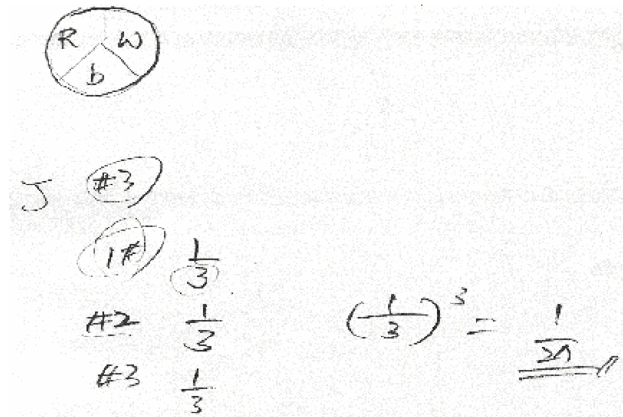

Figure 1. A participant's written work for the typical version of the Sequential problem, illustrating use of a picture

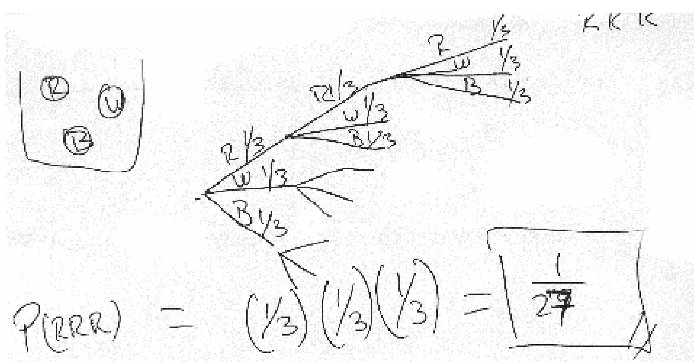

Figure 2. A participant's written work for the typical version of the Sequential problem, illustrating use of an outcome tree (and a picture)

Besides these standard pictorial and schematic representations used in previous studies (Russell, 2000; Zahner & Corter, 2002), we created two additional coding categories to cover cases not handled by the above classes. The first is a code indicating any attempt to invent and use what we termed a "*novel schematic representation*." Use of the term "novel" is meant to denote a schematic visual device that was not taught in the introductory class the participants were taking or had taken, nor used in standard probability texts. It is not meant to imply that the student invented and used a previously unknown type of visual device. An example of this category is a graphic used by several subjects for the *Permutations* problems: a list of four names whose elements are connected by lines or arrows to elements in a list of four numbers (see Figure 3). This type of representation (that we would classify as a directed graph) is apparently an attempt to develop or discover the correct outcome space for the problem. This type of representation uses spatial information and schematic elements (lines or arcs) to represent relational aspects of the problem, and is thus different from a simple outcome listing. The second additional coding category was defined to include any *spatial reorganization of the given information*. Use of a spatial organization scheme for information is not a formal graphical representation nor a purely pictorial representation. However, we have included this coding category because we have observed frequent use of spatially-organized rewriting of information to aid in problem solving. In the present study, many participants were observed to line up corresponding given probabilities or conditional probabilities (see Figure 4). This practice may make it easier for novice problem solvers to check for needed or missing information, to break down problem solution into subparts, or to make visual associations to relevant formulas.

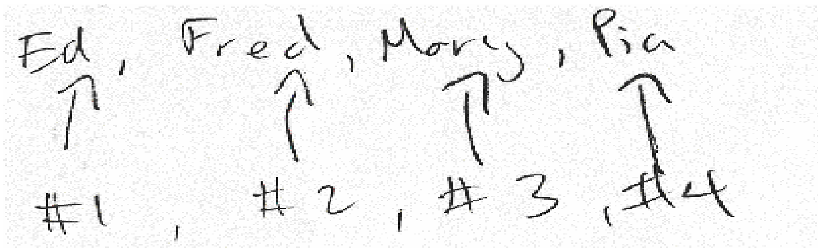

Figure 3. A participant's written work for typical version of the Permutation problem, illustrating use of a novel schematic representation

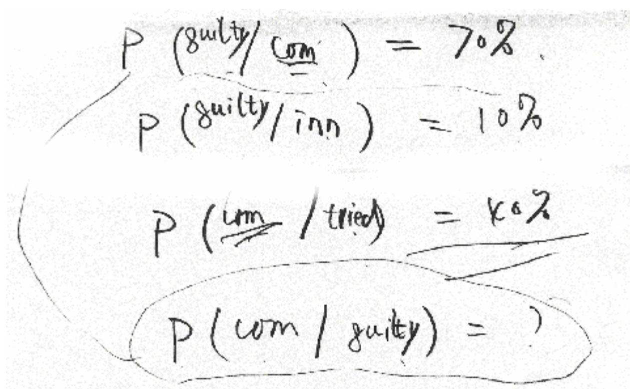

Figure 4. A participant's written work for the atypical version of the Conditional problem, illustrating use of spatial reorganization of given information

Computational method We also coded the *computational method* used by the problem solver, that is, the means by which the problem solver actually computes the answer required by the problem. We did this because in the course of coding the student protocols, we noticed that sometimes visual representations were used very early in the problem solving process, for example while the problem solver seemed to be still trying to understand the given problem information or to classify the problem, and sometimes later in the solution process, for example when the subject was trying to compute the actual numerical answer. In an effort to begin to understand what specific role or function external visual representations are serving in probability problem solving, we decided to separately code the method by which the problem solver actually computed the numerical answer required in each of these problems. We classified this later stage of each problem-solving protocol into three broad classes of computational method: *formula, procedural,* and *graphical*.

The computational method was characterized as *formula-based* problem solving if the participant wrote down an explicit (standard) formula, then substituted in quantities and solved the problem. An example of a formula-based computational method would be the use of the combinations formula followed by the necessary calculations: ${}_5C_2 = 5!/(2!)(3!) = (5)(4)/(2)(1) = 10$. This complete and rather formal method was distinguished from a *procedural* approach, which was used to code solutions carrying out a calculation involving only numbers without reference to any general formula or underlying principle. An example of a procedural approach would be if the participant calculated the probability of getting three heads in three flips of a coin by simply multiplying $(1/2)(1/2)(1/2) = 1/8$ without indicating any rationale for that procedure. A computational method was considered *graphical* if the subject used an external visual device to solve the problem, but only if the graphical device was judged to be instrumental to the method by which the student arrived at the actual numerical solution. An example of a graphical computational method would be if the subject multiplied two conditional probabilities that were taken from branches of a tree diagram.

No computational method was coded if the participant did not attempt to solve the problem. Occasionally, multiple computational methods were coded for a single problem. This occurred only when a participant attempted the problem, then abandoned that attempt, and attempted another computational method.

In order to assess reliability of the coding of the written protocols for external representations and computation method, a second rater coded all student solutions. Initial percent agreement between the two raters was over 90% for both external representation and computation method. Discrepancies were discussed by the two raters and the resulting consensus was used in all analyses reported.

Coding of audio protocols In order to better understand how the external visual representations are being used by problem solvers, we also transcribed and examined the audio portion of the videotapes capturing the participants' think-aloud protocols. Each utterance in a participant's audio transcript was coded to indicate if the participant was engaged in either of two broad phases or stages of problem solution: 1) a problem-representation phase that involves understanding the problem text and reformulating the problem in mathematical form; or 2) a solution phase, that involves selecting a solution strategy and implementing it. The video track of the tapes focused on participants' written inscriptions, including use of external representations. The video tapes were used to match uses of external visual devices with verbal statements by the participants about their thoughts and actions and the general phase of problem solving that they were engaged in: either problem representation or strategy selection and solution.

2.4. RESULTS

Preliminary analyses showed that individual problems and problem types varied considerably in difficulty. The rightmost column in the Appendix shows the proportion of subjects who correctly solved each problem. These proportions vary from a low of .08 for Problem P8 (*Conditional probability*, atypical) to .73 (for Problems P3 and P4, the typical and atypical *Sequential* problems). Regarding problem type, it was found that participants were most successful at solving the *Sequential* problems (.73 correct overall), followed by the *Permutations* problems (.48), then the *Combinations* problems (.29), and finally the *Conditional probability* problems (.25). These differences in solution rate among problem types were significant: in a log-linear analysis with factors Problem

Type, Typicality, and Correctness (the dependent variable), the Problem Type × Correctness association was significant ($\chi^2(3) = 31.56$, $p < .05$).

We were also interested in whether the atypical variant of each problem type was more difficult for problem solvers than the typical variant. Results indicated that typical and atypical variants did not differ in mean difficulty across all four problem types: mean proportion correct for the four typical problems encountered by each subject was .43, whereas for the four atypical problems it was .44. However, it is clear from the solution rate for individual problems (see Appendix) that the typical-atypical difference in solution rates varies across problem types. This interaction was tested by the three-way association of Problem Type × Typicality × Correctness in the loglinear analysis described above. This association was significant, ($\chi^2(3) = 17.00$, $p <.05$). Consequently, it is necessary to examine the effects of typicality separately for each problem type.

In particular, a difference in the expected direction was found for the *Conditional Probability* problems, with 42% of participants correctly solving the typical version of the problem versus only 8% for the atypical version. Unexpectedly, for the *Combinations* problem the solution rate for the atypical variant of the problem was much higher (at 73%) than for the typical variant (at 12%). In order to understand this unexpected result, we analyzed participants' written protocols to identify the specific solution methods used by participants for these problems. We found that most participants did not use the combinations formula at all to solve the atypical variant of the *Combinations* problem; rather they tended to solve this problem by treating it as a "sequential" problem involving sampling without replacement. For example, the problem can be solved using the formula: $P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1)$.

Inspection of the individual student protocols revealed that 92% (24 out of 26) of the participants selected this alternate method to solve the atypical *Combinations* problems. This probably occurred because the atypical problem is extremely difficult using the *Combinations* approach: the only two participants who tried this approach both failed to solve it. In contrast, exactly half of the 24 participants who adopted the sequential-events approach for the atypical variant succeeded in solving it. Note that many participants (73%) also tried to solve the typical *Combinations* problem using a sequential-events approach. However, all of these subjects failed to solve the problem, contributing to the overall low solution rate for the typical version. The difference in apparent difficulty of the sequential approach to these two problems probably involves that fact that in the typical *Combinations* problem, order is not important (but the use of the sequential solution method tends to elicit a solution attempt involving ordered pairs). Thus, many participants gave the answer 1/90 for this problem using the sequential approach, whereas the correct answer is 1/45. In the atypical variant, in contrast, there is a symmetry to the outcomes in the outcome space such that order is irrelevant.

What kinds of external visual representations are used? For each specific form of external visual representation, we calculated the percentage of participants who used that representation at least once. As shown in Table 2, we found that participants most often used *reorganization of the given information* (used at least once by 96.2% of the participants), followed by use of *picture*s (by 84.6% of the participants), *novel schematic representations* (65.4% of the participants), *trees* (53.8%), *outcome listings* (38.5%), *contingency tables* (7.7%) and finally *Venn diagrams* (3.8%).

Table 2. Frequency and percentage of participants using each type of external visual representation at least once, with frequency and percentage use of each representation across all problem solutions

	By participant (N=26)		By problem solution (N=208)	
Representation	Frequency	%	Frequency	%
Reorganize	25	96.2	72	34.6
Outcome Listings	10	38.5	20	9.6
Contingency Tables	2	7.7	6	2.9
Venn Diagrams	1	3.8	1	0.5
Trees	14	53.8	27	13.0
Novel Schematic	17	65.4	24	11.5
Pictures	22	84.6	64	30.8

Are different types of external visual representations used with different types of problems? We investigated the relationship between the type or topic of the probability problem (*Combinations, Sequential, Permutations,* and *Conditional*) and the type of representation that participants chose to use for it. In this analysis no distinction was made between the typical and atypical versions of each problem type. Table 3 summarizes how often each type of external representation was used for each type of problem. Because there were two problems of each type, each entry in this table is calculated across a total of 52 problem solutions.

Table 3. Frequency and percentage of problems of each type for which a given type of external representation was used (out of N=52 problem solutions), with χ^2 goodness-of fit tests evaluating differences in the frequencies of use of each representation across the four problem types

Representation	Combinations		Sequential		Permutations		Conditional		$\chi^2(3)$
	Freq	%	Freq	%	Freq	%	Freq	%	
Reorganize	26	50.0	4	7.7	4	7.7	38	73.1	47.6*
Outcome Listings	9	17.3	7	13.5	4	7.7	0	0.0	9.2*
Contingency Tables	0	0.0	0	0.0	4	7.7	2	3.8	---
Venn Diagrams	0	0.0	0	0.0	0	0.0	1	1.9	---
Trees	5	9.6	6	11.5	5	9.6	11	21.2	3.7
Novel Schematic	0	0.0	0	0.0	24	46.2	0	0.0	72.0*
Pictures	16	30.8	31	59.6	15	28.8	2	3.8	26.4*

*$p < .05$

For each row of the table, we used a chi-square goodness-of-fit test to test if each visual representation was used with unequal frequencies across problem types (i.e. columns). The chi-square goodness-of-fit test revealed that the frequency of use of *reorganization of given information* differed significantly across problem types ($\chi^2(3)$= 47.6, $p < .05$). This representation was used most often for the *Conditional Probability* problems (73.1% of the time) and the *Combinations* problems (50%). Usage was also distributed unequally across problem type for *outcome listings* ($\chi^2(3) = 9.2$, $p < .05$), with the most frequent use being for *Combinations* (17.3%) and *Sequential* (13.5%) problems. Use of *Novel schematic representations* was also distributed unequally across problem type ($\chi^2(3) = 72.0$, $p < .05$), because these representations were used only for the *Permutations* problems (46.2% of the time). *Novel schematic devices* may be tried especially often for the specific permutations problems used here because these problems are difficult for novices to recognize as permutation problems. That is because the cover stories for these particular permutation problems involve matching two sets of entities (e.g., tutors with students) rather than simply ordering one set of objects. This situation does not plug neatly into any formula or solution schema that students had been taught. This situation apparently spurred participants to try to understand these relatively unusual problems by inventing or adapting “novel” graphical representations.

Also, the use of *pictures* was distributed unequally across problem type ($\chi^2(3) = 26.4$, $p < .05$), due to very frequent use of pictures for the *Sequential* problems (in 59.6% of problem solutions), *Combination*s (30.8%), and *Permutations* (28.8%) and infrequent use (3.8%) for the *Conditional Probability* problems. The use of *trees* did not vary significantly across problem type ($\chi^2(3) = 3.7$, $p < .05$). Inspection of Table 3 reveals that trees were used about 10% of the time or more for all four problem types. This result seems to show that at least for these types of probability problems, trees were perceived by study participants as widely applicable. The use of *Contingency tables* and *Venn diagrams* was too infrequent to be tested in the manner.

Sometimes problem solvers used more than one form of external visual representation in a single problem solution. Figure 5 shows the percentage of use of single and multiple representations across all problem solutions, separately by problem and problem type. Across the eight problems, multiple external visual representations were used in 23.6% of the problem solutions. The most common combinations of multiple representations were *pictures* with *reorganization* (used in 13% of the problem solutions), and *pictures* with *trees* (used in 6%; see Figure 2 for an example). All other instances of multiple external representations occurred less than 2% of the time. Multiple representations were used most often for *Combinations* and *Sequential* problems. This may simply reflect the fact that pictures were used quite often for these problem types, as shown in Table 3.

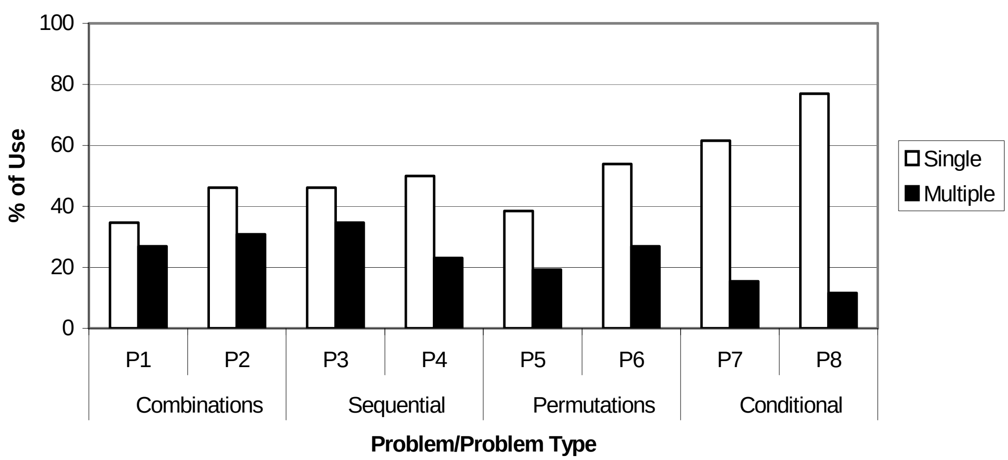

Figure 5. Percentage of use of single and multiple external representations in the problem solutions (N=208), by problem and problem type

The above results showing differences in frequency of use of specific representations across the four problem types demonstrate that participants are selecting representations based on the type of problem they are trying to solve, presumably reacting to differences in the problem schema for the four problem types. This suggests that participants' solution methods (at least, their use of external visual representations) vary depending on the problem's underlying schema or "deep structure." We return to this issue in the Discussion section.

Is solution success associated with use of external representations? If external visual devices do indeed serve some purpose for problem solvers, then we might expect an association between solution success and the specific external representation used (if any). We explored this idea by estimating the conditional probability of solution success given use of each type of external visual representation. The results, shown in Table 4, show that use of particular external visual representations was associated with higher rates of solution correctness for some problem types (compared to baseline performance for that problem type), and with lower rates of success for others. For example, the use of *reorganization* is associated with a higher rate of solution success for *Combinations* problems (.32 versus .29), but with a lower rate of success for *Sequential* (.50 versus .73), *Permutations* (.00 versus .48), and *Conditional* (.15 versus .25) problems.

Table 4. Proportion of correct solutions, given the use of a particular representation, separately by problem type. Dashed lines indicate a cell with fewer than four uses of that representation (i.e., n ≤ 3).

Representation	Combinations	Sequential	Permutations	Conditional
Reorganize	.32	.50	.00	.15
Outcome Listings	.43	.42	--	--
Contingency Tables	--	--	.00	--
Venn Diagrams	--	--	--	--
Trees	.60	.67	.25	.25
Novel Schematic	--	--	.44	--
Pictures	.23	.63	.59	--
Mean P(correct)	.29	.73	.48	.25

Table 4 shows that for the *Combinations* problems, use of *reorganization*, *outcome listings*, or *trees* were all associated with higher rates of solution success, whereas use of *pictures* was associated with a lower rate of success. Presumably, the first three types of representations are useful here because the essence of such combinatorics problems is to identify the number of outcomes in the outcome space. However, trees are not usually useful for problems involving simultaneous sampling of multiple objects (where order is not important). We therefore reexamined participants' solutions to try to understand this association. We found that trees were used in only five solutions for the *Combinations* problems, and all of these were cases where the problem solver was treating the problem as a sequential problem rather than using the combinations formula.

For the *Sequential* problems, use of any external visual representation was associated with a lower rate of solution success. *Sequential* problems were the easiest type of problem overall, with *P*(correct) = .73, so it may be that participants did not feel any need to call upon visual representations unless they were among the few who experienced difficulty with these problems.

For the *Permutations* problems, use of *reorganization*, *contingency tables,* and *trees* was associated with lower rates of solution success. Contingency tables in particular do not seem appropriate for permutation problems, which involve ordering a single set of objects. Trees are rarely used to represent sequential sampling without replacement, though in principle they could be applied. However, use of *pictures* was associated with a higher rate of solution success for these *Permutations* problems. *Pictures* may be especially useful for these particular (unusual) permutations problems, which are unusual in that they describe matching *two* sets of objects rather than ordering a single set. It may be that pictures facilitated the realization that the ordering of one of these sets is arbitrary.

Finally, for the *Conditional probability* problem, external visual representations were not often used. Use of *reorganization* was associated with slightly lower rates of solution success. It is surprising that trees were not often used for these problems because their use for such problems was explicitly described in the course.

Thus, this table seems to offer mixed evidence concerning the usefulness of external visual representations in probability problem solving. The positive associations found seem easily explainable. We argue that the few observed negative associations between external visual representations and solution success do not prove that use of external representations is harmful in probability problem solving. Rather, the negative associations may arise because external visual representations are more often called upon when the problem is especially difficult for the problem solver (cf. Hegarty & Kozhevnikov, 1999; Lowrie & Kay, 2001).

In addition to the analysis of solution correctness given the use of a particular external visual representation, we correlated the number of times a participant used *any* external visual representation (which ranged from 2 to 18 for participants) and the participant's overall solution success (defined as number of problems correct out of 8 for a participant). Results from this analysis show that there is a significant *negative* overall correlation between the use of an external visual representation and solution success ($r = -.40$, $p < .05$). We also found marginally significant negative correlations between use of certain specific representations and solution success. Specifically, the use of *reorganization of the given information* was negatively correlated with solution success ($r = -.37$, $p = .06$), as was the use of *outcome listings* ($r = -.37$, $p = .06$).

What computation methods are used in PPS? Are external visual devices used in computing problem solutions? We calculated how often each of the three computational methods (formula, procedural, graphical) was used for each problem. Results showed that students used the procedural computational method most often (on average in 5.5 out of 8 problems) followed by formula-based computational methods (1.19 out of 8 problems) and finally, graphical solutions (0.42 problems out of 8.) Thus, external visual representations were rarely used to compute solution. For 13.5% of problems overall, subjects did not complete the problem to the point of computing a solution. Multiple computation methods (coded only when the participant made multiple solution attempts) were observed only 2.4% of the time.

Are different computation methods used with different problem types? We calculated frequency and percentage of use of each of the three types of computation method across problem types. Note that more than one type of solution method could be coded for a given solution, and that if the problem solver did not attempt to compute a numeric solution no computation method was coded. Results showed that formulas were used most often for the *Conditional* problems (21.2% of the time) and *Combinations* (21.2%). A procedural computation method was used most often for *Sequential* problems (84.6%), *Combinations* (75.0%), and *Permutations* (75.0%). Finally, a graphical computational method was used most often with the *Conditional Probability* problems (9.6%). Note that for the *Conditional Probability* problems the observed student solutions were not purely graphical; rather the tree graphs were typically used in conjunction with procedural calculations. We also performed a chi-square goodness-of-fit test to determine if each computational method was used equally often across all four problem types. The results (Table 5) show differential use of the procedural computation method across problem types ($\chi^2(3) = 8.57$, $p < .05$). Frequency of use did not differ significantly across problem types for the procedural computation method ($\chi^2(3) = 6.05$, $p > .05$), nor for the graphical computation method ($\chi^2(3) = 5.21$, $p > .05$). The pattern of results in Table 5 suggests that a procedural solution method is used relatively less often for the *Conditional Probability* problems.

Table 5. Frequency and percentage of problem solutions (n = 52) of each type for which a given computational method was used, with χ^2 goodness-of-fit tests of differences in frequency of use of each method across the four problem types

Computational method	Combinations		Sequential		Permutations		Conditional		$\chi^2(3)$
	Freq	%	Freq	%	Freq	%	Freq	%	
Formulas	11	21.2	3	5.8	6	11.5	11	21.2	6.05
Procedural	39	75.0	44	84.6	39	75.0	21	40.4	8.57*
Graphical	3	5.8	3	3.8	0	0.0	5	9.6	5.21

*$p < .05$

Is solution success associated with computational method? We also checked for associations between solution success and computational method, separately by problem type (Table 6). Use of a formula-based computation method was associated with a higher rate of solution success only for *Combinations* problems. This makes sense, because the combinations problems are arguably best solved via formulas. Use of a procedural method was associated with the highest rates of solution success for the other three problem types. This is probably because if it is intuitively clear to a student how to solve a problem, only the computations need be written down (and the solution would be coded as a procedural one). Using a graphical method to aid in computing the solution was observed infrequently, except for the *Conditional probability* problems. For these problems, the tree can be used to organize the procedural calculations.

Table 6. Proportions of correct solutions given the use of a particular computation method, separately by problem type. Dashed lines indicate a cell with fewer than four uses of that representation (i.e., n ≤ 3).

Computational method	Combinations	Sequential	Permutations	Conditional
Formula	.38	--	--	.11
Procedural	.30	.77	.57	.46
Graphical	--	--	--	.30
Mean	.29	.73	.48	.25

Are there differences between typical and atypical problems in the use of external visual representations or computational method? We investigated whether there is a difference in the rates of use of visual representations for the typical and atypical problems. To test this, for each type of representation we compared the summed frequency of its use for the four problems presented in their typical versions to the summed frequency of its use for the four problems presented in their atypical versions. The results show that the only significant difference in use of an external representation between typical and atypical problems was for pictures (paired-sample $t(25) = -3.86$, $p < .05$). Specifically, *pictures* were used more often for atypical problems (for 38.5% of problems) than for typical problems (23.0%), as shown in Table 7. This result is not at all surprising – pictures may be used especially often to try to better understand or structure a difficult problem, especially one that does not plug in easily to a familiar solution schema. On the other hand, if the solution method is obvious, nothing is gained (and time and effort are expended) in drawing a picture. The only other representation used more often for atypical problems (though the difference is not significant) is *reorganization.* As for pictures, it can be argued that this is a general-purpose method that is often useful when the problem text itself is difficult to understand.

Table 7. Frequency and percentage use of each type of external representation, separately for the four typical and four atypical problems experienced by each participant. (= significant difference between the total number of uses for typical and atypical using a dependent samples t test, df = 25)*

Representation	Typical		Atypical	
	Freq	%	Freq	%
Reorganize	34	32.8	38	36.5
Outcome Listings	12	11.5	8	7.8
Contingency Tables	4	3.8	2	2.0
Venn Diagrams	1	1.0	0	0.0
Trees	15	14.5	12	11.5
Novel Schematic	13	12.5	11	10.5
Pictures*	24	23.0	40	38.5

*$p < .05$

Table 8 reports the rates of use of different computation methods for typical versus atypical problems. There were no significant differences in the use of different computation methods for typical and atypical problems.

Table 8. Percentage use of each type of computational method, summed across problems, separately for the four typical and four atypical problems experienced by each participant

Computational Method	Typical		Atypical	
	Freq	%	Freq	%
Formula	15	14.5	16	15.5
Procedural	73	70.3	70	67.3
Graphical	6	5.8	10	10.0

2.5. ANALYSIS OF AUDIO PROTOCOLS

As described in the Methods section, each utterance in the audio track of the session videotapes was coded as relevant to either the problem solvers' problem-representation phase or as part of the solution execution phase (cf. Novick & Hmelo, 1994). We also matched any use of an external visual representation in a solution (as captured in the video track) to any utterances made simultaneously. This enabled us to classify uses of external visual representations as being associated with either or both of these broad temporal stages of problem solving.

Of the 2,756 utterances in the audio transcripts, approximately 63% of them were coded as part of the Problem Representation phase (1,734 utterances) and approximately 32% of the utterances were coded as Solution Execution (881 utterances). The remaining utterances were not coded as part of either phase. This occurred if the utterance was a meta-comment or irrelevant, for example concerning how difficult the problem was or the temperature of the room in which the study was being conducted.

Of the 1,734 utterances that were coded as part of the Problem Representation phase, 417 (24.08%) of them were matched with the use of an external visual representation. Of the 881 utterances that were coded as involving the Solution Execution phase, 31 utterances (3.68%) were matched with the use of an external visual representation. Although participants did use external visual representations during both phases of problem solving, they tended to use them more often during the Problem Representation phase than during the Solution Execution phase.

These associations were broken down by type of representation used (Table 9). Results indicated that participants more often tended to use the external visual representations to help understand and organize the problem text (i.e., in problem understanding) than to select or execute solutions. This trend was especially strong for *pictures*, *reorganization of the given information*, and *novel schematic representations*. For example, *pictures* were used significantly more often during the problem representation phase because participants claimed that they helped them visualize the problem more clearly. One subject explained, "I drew the ten cookies because I needed literally to visualize it, and then based on what the information is in this problem, there's obviously….there's ten different types of cookies." (Subject #16)

Table 9. Total frequency of use of particular representations, by problem solving phase

Representation	Problem Representation	Strategy & Execution
Reorganize	169	9
Outcome Listings	17	3
Contingency Tables	11	3
Venn Diagrams	4	0
Trees	50	13
Novel Schematic	45	1
Pictures*	120	4
Total	417	32

*$p < .05$

3. DISCUSSION

Our results show that students sometimes choose to use self-generated external visual representations while solving probability word problems. Presumably, this is because problem solvers believe that these representations are useful in solving the problems, because in this study they were requested merely to "show their work," and not explicitly requested to produce any diagrams or other visual devices. A skeptical observer might worry that the verbal prompts occasionally issued by the experimenter here could have served as a general prompt to try alternative representations. However, Russell (2000) found similar levels and patterns of use of these types of representations in students' actual answers to course assignments, lending confidence to the conclusions that problem solvers choose to use such representations because they are thought to be useful.

The results also document what types of spatial and graphical devices are used in probability problem solving. Using a very broad definition of external visual representations, the types we identified included (in decreasing order of frequency of use): *reorganization of the given information, pictures, novel schematic representations, trees, outcome listings, contingency tables,* and *Venn diagrams* (cf. Russell, 2000). Of course, our reported relative frequencies of use for these representations may not generalize to other curricula, other specific sets of problems, and other problem solvers.

We found evidence that the frequency of use of each of these different external visual representations depends on the *type* of probability problem being solved. Type of problem refers to the basic problem schema, and not to surface characteristics of the problem. The types of probability problems studied here were *Conditional Probability, Combinations, Permutations,* and *Sequential* problems. Our results showed that *pictures* were used most often for *Sequential, Combinations*, and *Permutations* problems; *outcome listings* were used more often for *Combinations* and *Sequential* problems; *trees* were used most often for *Conditional* problems; *novel schematic representations* were used mainly with *Permutation* problems; and *reorganization of given information* was used more often with *Conditional* and *Combinations* problems. One way to interpret these findings is that problem solvers attempt to select representations appropriate to the problem's structure, and that the appropriateness of the representation is determined by the problem's solution schema, not by surface characteristics.

However, any conclusions as to *specific* associations between the type of visual device and the problem type must be tempered by consideration of the particular set of problems used here to represent these general types. First, the *Permutation* problems studied here were unusual in that they described situations in which two sets of entities (e.g., tutors and students) were to be matched in a one-to-one fashion, but the ordering of one set was arbitrary, making the problem isomorphic to an ordering problem. This "schema mismatch" may have made these problems particularly difficult for our problem solvers, spurring more attempts to use novel schematic representations and pictures. Second, problems were experienced with our manipulation of typical and atypical problems for the *Combinations* problems. The atypical *Combinations* problems were most often solved by an alternative method, using a sequential-events approach, that resulted in a higher rate of success than for the typical variant of this problem type. Thus, although it was our intention to manipulate problem typicality in such a way that the atypical problems were at least as difficult as the typical ones, this did not happen for this problem type. In future studies, we hope to more fully refine and explore the notions of problem typicality and difficulty, and to try to disentangle their effects experimentally by careful development and piloting of materials.

In future research, we also hope to more fully investigate aspects of the schematic devices that play a part in determining the appropriateness of a representation for a given problem. We believe that the seven types of external visual representations studied here differ in some important ways. Three of the visual representations (reorganization, outcome listings, and contingency tables) can be considered forms of *tabulation*. Another three (Venn diagrams, trees, and novel schematic representations) could be classified as *schematic devices*, and the final type (pictures) refers to iconic representations of concrete aspects of the problems. We term the second group of representations (Venn diagrams, trees, and novel schematic representations) *schematic* because structural aspects of the graphs symbolically represent meaningful aspects of the problem.

Novick and Hurley (2001) propose that different types of schematic devices (or "diagrams") have structural aspects or properties that determine their range of applicability. The associations we have found between use of the different types of representations and specific problem types suggest that properties of the diagrams and properties of the problem schema are being matched (though not always successfully) by participants. For example, trees seem naturally appropriate for sequential problems such as the results of multiple coin flips or successive spins of a spinner, whereas contingency tables and Venn diagrams are particularly appropriate for representing joint or compound events.

These seven identified types of external representations also differ in terms of their degree of structure. This has implications for how broadly or how narrowly the visual representation may apply. Specifically, we argue that *reorganization of the given information, pictures,* and *outcome listings* are relatively general representations that can be applied to a wide variety of problems, whereas *trees, contingency tables,* and *Venn diagrams* have more inherent structure, thus may be applicable to a more limited set of problems. Finally, the category of *novel schematic representations* is by definition not limited to any specific type of structure, thus this category of representation is also widely applicable (although any specific novel graph may have limited applicability). However, novel schematic representations seem to be used only when the problem solver encounters a very atypical or unusual problem that does not seem to plug into any familiar schema.

If we are right that *reorganization of the given information, pictures,* and *outcome listings* are very general tools, whereas the schematic devices (*trees* and *Venn diagrams*) and contingency tables are more limited in scope of application because they have more constrained structures, then those variations in scope of application ought to show up in our data. Calculating the average percentage of problems for which each type of representation was used (Table 2), provides some supporting results for this idea. The three types of representation argued here to be general ones (*reorganization, pictures,* and *outcome listings*) were used in 25% of problem solutions on average, whereas the three specific types were used in only 6% of problem solutions on average. However, the picture given by Table 3 is a bit less clear. Here it can be seen that uses of *reorganization of the given information* and *pictures* are spread across all four problem types, and *outcome listings* are used for three out of four types, whereas the more constrained types of representation *Venn diagrams* and *contingency tables* are used for only one or two types of problem. However*, trees* are used across all four problem types. Thus, except for *trees*, the predicted pattern does hold.

3.1. DO SPONTANEOUSLY SELF-GENERATED VISUAL REPRESENTATIONS HELP PROBABILITY PROBLEM SOLVERS?

The present study provides mixed evidence for the idea that external visual devices are used by probability problem solvers because they are helpful (i.e., they aid in solving the problem). For example, we found higher rates of solution success given use of *reorganization*, *outcome listings*, and *trees* for the *Combinations* problems. The finding regarding outcome listings makes sense intuitively because the essence of the combinations problems involves determining the number of outcomes in the outcomes space. Furthermore, use of trees is associated with a higher success rate (60%) for those solving the *Combinations* problems via a sequential approach, and use of trees is a relatively successful strategy (67% success rate) for the "true" *Sequential* problems. This finding seems easily interpretable, because these sequential events problems have structures that map directly onto tree diagrams. Specifically, the *Sequential* problems used here described a sequence of trials or events, each of which had several possible outcomes. Thus the process determining the outcome space can be described by a branching set of possibilities. In the corresponding tree, each node of the tree graph corresponds to one of the sequential events (e.g., one spin of the spinner), and the branches that ensue from that node represent the several possible outcomes of that uncertain event.

For the *Permutations* problems, the use of *pictures* led to a higher success rate. Our explanation for this finding is based on the point we have already made, that these particular permutation problems were atypical in that their semantic content (i.e., the real-world situation they described) describes a matching process between two sets of objects (e.g., tutors and students). Typical permutation problems the students had seen in their course consisted of problems in which a *single* set of objects is randomly ordered. Thus, it is only through a relatively sophisticated symmetry argument (requiring what is perhaps a rare or difficult insight) that the student was likely to see that one set of objects could be arbitrarily ordered, hence ignored, reducing the problem to one about ordering a single set of objects. For this reason we suspect that the increased solution success associated with use of pictures for this problem type may indicate a facilitative effect of pictures for *problem restructuring*. Such restructuring seems a necessary insight to deal with this relatively novel type of problem.

Although problem solvers showed relatively frequent use of *novel schematic representations* for the *Permutations* problems, these novel or invented types of external graphical representations were apparently not always useful, because their use did not lead to increased solution success.

Why are positive correlations between specific types of visual representations and specific problem types relatively rare in our data? As Novick and Hmelo (1994) observe, having an appropriate problem representation does not guarantee that the problem can be solved, because computational or other issues may intrude, lowering correlations between initial problem representations and solution success. Furthermore, even if graphics could be helpful, prior research shows that students are not always successful in finding correct representations for problems (Novick, 1990). Our data provide additional evidence that this is true. Additionally, some evidence from our study suggests that choosing an *inappropriate* representation might be harmful to a student's chance of successfully solving a problem. For example, for the *Permutations* problems solution success was negatively associated with use of *contingency tables* and with use of *reorganization of the given information.* The former finding can be explained because contingency tables are not appropriate for representing problems involving the ordered selection of a single set of objects. The latter finding can be explained by viewing the *reorganization* strategy as a response commonly chosen when the problem solver is confused. Thus, the negative association may indicate that when a student is stymied by a problem, rewriting the given information might be seen as a general-purpose strategy, to be tried if the student is merely casting about for any approach that might help.

We also found lower rates of success associated with use of *outcome listings* for the *Sequential* problems. Here, we suspect that the choice of representation could be based on a wrong understanding of the problem situation, or might just be an unfortunate (being potentially misleading) choice. For the *Sequential* problems used here, the listings seem to be appropriate, but they may cue (incorrect) approaches based on treating the outcome space as consisting of equally-likely outcomes.

In addition to lower rates of solution success associated with particular external visual representations, we also found a significant negative correlation ($r = -.4$, $p < .05$) between solution success and the overall use of external visual representation, suggesting that our participants were often using the external visual representations in futile solution attempts. Looking more closely at the correlations, we found that *reorganization of the given information* and *outcome listings* were marginally significantly correlated with solution failure. These two types of representations are very general tools for problem solving and participants may use these types of representations mainly when they are having trouble solving the problems. This tendency could produce such correlation with solution failure. A related possibility is that participants who are adept at solving probability problems may not need to use or report the use of a visual representation; whereas weaker problem solvers may choose very general external visual representations in the absence of insights that might allow them to select a more specific representation, leading to a lower rate of success overall for these lower-ability participants who needed to resort to external representations.

Prior research by Hegarty and Kozhevnikov (1999) suggests that there are two main types of external visual representations, schematic and pictorial. They found that the use of schematic representations was positively correlated with solution success and the use of pictorial representations was negatively correlated with success. In our data, solution success was negatively (but non-significantly) correlated with use of all types of external visuals, but the correlations were more negative with the use of *pictures*, *reorganization*, and *outcome listings* than with *contingency tables* and with the schematic representations (*Venn diagrams*, *trees*, and *novel schematic representations*), lending tentative support to the importance of this distinction.

In summary, appropriate use of a correct external visual representation may generally be helpful in problem solving, but this effect is difficult to measure in the present type of study, in which the student only generates an external representation if he or she so chooses. First, there is evidence (Lowrie & Kay, 2001; Table 7) that self-generated external visuals may be tried more often for difficult or novel problems, which have a lower solution rate in general. It should be easier to demonstrate facilitative effects of external visuals in less naturalistic studies in which the visual representations are provided to the student, or the student is explicitly asked to generate an appropriate representation before attempting to compute the answer. That type of study has been common in the literature on uses of visual representations in (non-mathematical) problem solving. However, the present data showing which types of graphical representations are spontaneously used for which types of problems may aid in designing such experimental studies and educational interventions.

Finally, facilitative effects of using visual representations may not be easy to detect in the present type of experiment because choosing the correct representation is a non-trivial task, and may require a certain level of problem understanding to accomplish (Novick, 1990; 2001; Novick & Hmelo, 1994; Novick & Hurley, 2001). With novice problem solvers, knowledge of why one representation is more appropriate than another may still be incomplete, because they have not yet mastered the appropriate schemas. Riley, Greeno and Heller (1983) found that failure to solve word problems might be caused more often by a lack of appropriate schemas than by poor arithmetic skills. They observed that problem solvers often carried out correct arithmetic procedures on incorrect representations of the problems. The negative associations we found between certain types of (presumably inappropriate) representations and solution success seem consistent with their conclusions. Interestingly, De Bock, Verschaffel, Janssens, Van Dooren, and Claes (2003) also found negative effects on solution success of asking students to generate *specified* diagrams for geometry problems, showing that not all experimenter-selected representations are useful as well (cf. Tversky, 2001; Mayer & Gallini, 1990; Scaife & Rogers, 1996), or perhaps merely indicating that not all student-generated visual representations are produced correctly, even when appropriately cued.

3.2. SOME FINAL ISSUES

One potential limitation of the present study is the question of how well the results will generalize to other populations of students. Participants in the present study were graduate students in social sciences and education, who were finishing or had recently finished an introductory course in probability and statistics. However, the participants were actually quite diverse in terms of mathematics background, including people who had not taken any mathematics in college and those who had taken a number of undergraduate or graduate mathematics and statistics courses. Thus, we believe that our results would generalize to other populations, such as high school students who had completed a similar probability course. But this question, and the question of how well the results would generalize to a wider set of probability problems, should be addressed by future research.

Another factor possibly affecting the generalizability of these results is that participants were taught probability problem solving using external visual representations, and associations between specific problem types and specific types of external graphical devices may have been implicitly or explicitly taught. Thus, the results of our study may just be a reflection of the instruction. However, the use of "novel" graphical representations by some participants indicates that although the participants were taught to use visual devices when solving probability problems, they do not always use the specific representations they were taught to use in class. This finding suggests that students believe that visual representations are useful, and try to use even representations that they have not been explicitly taught.

Studying how students solve probability problems (or any type of mathematics word problem) is a complex endeavor. One reason is that students can use any of several solution methods or strategies for many problems. Even worse, an individual student may switch approaches across similar problems, or even during the solution of a single problem. As an example of how multiple solution strategies can complicate the research process, we designed each of our probability problems with a particular formula or problem-solving schema in mind. However, in producing atypical problems for a given method, in at least one case (the atypical *Combinations* problem) we produced a problem that could easily be solved by another method entirely (treating the problem as involving sequential events), with a different appropriate external representation.

Thus, another limitation to the present study is that our manipulation of problem "typicality" was not fully successful, due to the use of alternative solution strategies by many participants. In a well-controlled experimental study with novice probability problem solvers, this problem could be avoided by introducing only one solution method or probability principle at a time. However, in a naturalistic study like the present one, where participants have been taught an array of probability problem-solving techniques, the problem of alternate strategy choice is difficult to avoid. Certainly such effects could be minimized by more careful piloting of materials in future studies.

Another issue deserving of future study is to more closely investigate the temporal process of probability problem solving. In the present study we have used a coding scheme that separates uses of external visuals for problem understanding and representation from the type of method used to compute the problem solution, but we still do not have a clear picture of the temporal stages of probability problem solving. We plan future studies that will use think-aloud protocols and structured interviews to try to distinguish sequential stages of probability problem solving, and that will examine specifically when and how external visual representations are used in the temporal process of PPS. We also plan to investigate the coordination of external visual representations with internal visualizations (cf. Scaife & Rogers, 1996). Results of these studies may bring us to a more complete understanding of the role played by visual representations and visualization skills in probability problem solving.

Results from the present study might be useful in improving instruction in the domain of probability problem solving. The very general types of external representations considered here (*pictures*, *reorganization*, and *outcome listing)* might be taught to students as general methods that can help them restructure particularly difficult problems. In contrast, the schematic representations studied here (*Venn diagrams* and *outcome trees*) and *contingency tables* could be taught as applicable to particular problem types. In line with the work of Novick and colleagues (e.g., Novick, 1990; Novick & Hmelo, 1994), abstract aspects or "features" of problems and of specific graphical representations could be taught to students, and it could be emphasized that a given representation will most likely be useful when these structural aspects of the problem and the visual device match. To some extent such principles may already be employed by instructors of probability courses, but future research should explore and better document the success of such practices.

REFERENCES

Batanero, C., Godino, J., & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education, 12*(1).
[Online: www.amstat.org/publications/jse/v12n1/batanero.html]

Chance, B., & Garfield, J. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal, 1*(1)*,* 38-41.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ1(2).pdf]

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modeling non-linear geometry problems. *Learning and Instruction, 13*(4), 441-463.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Gelman, A., & Nolan, D. (2002). *Teaching statistics: A bag of tricks.* Oxford, England: Oxford University Press.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective Probability* (pp. 129-161). Somerset, NJ: John Wiley & Sons.

Ginsburg, H. (1997). *Entering the child's mind: The clinical interview in psychological research and practice.* New York: Cambridge University Press.

Hall, V., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology*, *89*(4), 677-681.

Hampson, P., & Morris, P. (1990). Imagery and working memory. In P. J. Hampson, D. E. Marks, & J. T. E. Richardson (Eds.), *Imagery: Current developments* (pp. 78-102). New York: Routledge.

Hannafin, R. D., Burruss, J. D., & Little, C. (2001). Learning with dynamic geometry programs: Perspectives of teachers and learners. *The Journal of Educational Research, 94*(3), 132-144.

Hannafin, R. D., & Scott, B. (1998). Identifying critical learner traits in a dynamic computer-based geometry program. *The Journal of Educational Research, 92*(1), 3-12.

Hegarty, M., & Just, M. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717-742.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684-689.
Hollebrands, K. (2003). High school students' understandings of geometric transformations in the context of a technological environment. *Journal of Mathematical Behavior, 22*(1), 55-72.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge; New York: Cambridge University Press.
Kaufmann, G. (1990). Imagery effects on problem solving. In P. J. Hampson, D. E. Marks, & J. T. E. Richardson (Eds.), *Imagery: Current developments* (pp. 169-197). New York: Routledge.
Keeler, C. M., & Steinhorst, R. K. (2001). A new approach to learning probability in the first statistics course. *Journal of Statistics Education*, *9*(3). [Online: www.amstat.org/publications/jse/v9n3/keeler.html]
Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109-29.
Koedinger, K. R., & Anderson, R. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30-43.
Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*(1), 59-98.
Konold, C. (1995). Confessions of a coin flipper and would-be instructor. *The American Statistician*, *49*(2), 203-209.
Konold, C. (1996). Representing probabilities with pipe diagrams. *The Mathematics Teacher*, *89*(5), 378-384.
Konold, C., Pollatsek, A., Well, A. D., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*, 392-414.
Kozhevnikov, M., Hegarty, M., & Mayer, R. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, *20*(1), 47-77.
Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching*, *35*(1), 3-25.
Logie, R., & Baddeley, A. (1990) Imagery and working memory. In P. J. Hampson, D. E. Marks, & J. T. E. Richardson (Eds.), *Imagery: Current Developments* (pp. 103-128). New York: Routledge.
Lowrie, T., & Kay, R. (2001). Relationship between visual and nonvisual solution methods and difficulty in elementary mathematics. *The Journal of Educational Research*, *94*(4), 248-255.
Mayer, R. (1989). Systemic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*(2), 240-246.
Mayer, R. (1992). Mathematical problem solving: Thinking as based on domain specific knowledge. In R. E. Mayer (Ed.), *Thinking, problem solving, cognition* (pp. 455-489). New York: W. H. Freeman & Co.
Mayer, R. (2001). *Multimedia learning*. Cambridge: The Cambridge University Press.
Mayer, R., & Anderson, R. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, *83*(4), 484-490.
Mayer, R., & Anderson, R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*(4), 444-452.
Mayer, R., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715-726.

Mayer, R., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, *91*(4), 171-185.

Molitor, S., Ballstaedt, S. P., & Mandl, H. (1989). Problems in knowledge acquisition from text and pictures. In H. Mandl & J. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 3-35). North-Holland: Elsevier Science Publishers.

Nemirovsky, R. (1994). On ways of symbolizing: The case of Laura and the velocity sign. *Journal of Mathematical Behavior*, *13*(4), 389-422.

Novick, L. (1990). Representational transfer in problem solving. *Psychological Science*, *1*(2), 128-132.

Novick, L. (2001). Spatial diagrams: Key instruments in the toolbox for thought. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Volume 40 (pp. 279-325). San Diego, CA: Academic Press.

Novick, L., & Hmelo, C. (1994). Transferring symbolic representations across nonisomorphic problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1296–1321.

Novick, L.R., & Hurley, S. M. (2001). To matrix, network or hierarchy: That is the question. *Cognitive Psychology*, *42*(2), 158-216.

O'Connell, A. A. (1993). *A classification of student errors in probability problem-solving*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

O'Connell, A. A. (1999). Understanding the nature of errors in probability problem solving. *Educational Research and Evaluation*, *5*(1), 1-21.

O'Connell, A., & Corter, J. E. (1993, April). Student misconceptions in probability problem-solving. Paper presented at annual meeting of the American Educational Research Association, Atlanta, GA.

Penner, D. E., Giles, N. D., Lehrer, R., & Schauble, L. (1996). Building functional models: Designing an elbow. *Journal of Research in Science Teaching*, *34*(2), 125-143.

Pollatsek, A., Well, A., Konold, C., Hardiman, P., and Cobb, G. (1987). Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes*, *40*(2), 255-269.

Presmeg, N. C. (1986). Visualization in high school mathematics. *For the Learning of Mathematics – An International Journal of Mathematics Education*, *6*(3), 42-46.

Qin, Y., & Simon, H. (1995). Imagery and mental models in problem solving. In B. Chandrasekaran, J. Glasgow, & N. H. Narayanan (Eds.), *Diagrammatic reasoning* (pp. 403-434). Menlo Park, CA: The AAAI Press.

Reusser, K. (1996). From cognitive modeling to the design of pedagogical tools. In S. Vosniadou, E. De Corte, R. Glaser & H. Mandl (Eds.), *International perspectives on the design of technology–supported learning environments* (pp. 81-103). Mahwah, NJ: Erlbaum.

Riley, M., Greeno, J., & Heller, J. (1983). Development of children's problem-solving ability. In H. Ginsburg (Ed.), *The development of mathematical thinking*. New York: Academic Press.

Russell, W. E. (2000). The use of visual devices in probability problem solving. (Doctoral Dissertation, Columbia University, 2000). *Dissertation Abstracts International*, *61*, 1333.

Santos-Trigo, M. (1996). An exploration of strategies used by students to solve problems with multiple ways of solution. *Journal of Mathematical Behavior*, *15*(3), 263-284.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185-213.
Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, *20*(4), 457-497.
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129-184.
Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. *Instructional Science*, *28*(3), 227-262.
Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*(3), 380-400.
Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79-112). Cambridge, MA: MIT Press.
Zahner, D. C., & Corter, J. E. (2002, April). Clinical interviewing to uncover the cognitive processes of probability problem solvers. Paper presented at the 2002 Annual Meeting of the American Educational Research Association, New Orleans, LA.

James E. Corter
Doris C. Zahner
Department of Human Development
Teachers College, Columbia University
525 West 120th Street
New York, NY 10027

APPENDIX: PROBABILITY PROBLEMS

The eight probability problems (test form A) classified by topic (problem type) and level of typicality, with observed proportion correct for each item. Test form B counterbalanced surface content of the text and level of typicality for each problem.

Topic	Typicality	Problem Text	Proportion of correct responses
Combinations	Typical	There are 10 books on Mary's bookcase. She randomly grabs 2 books to read on the bus. What is the probability that the 2 books are *Little Women* and *War & Peace*? (Both these books are on her bookshelf.)	0.115
	Atypical	There are 10 cookies in a cookie jar. Three of the cookies are chocolate chip, seven are sugar. A child blindly picks 2 cookies from the cookie jar. What is the probability that both cookies are chocolate chip?	0.731
Sequential	Typical	There are three balls in an urn. One is red, one is white, and one is blue. Jane randomly draws a ball from the urn, then replaces it, three times in all. What is the probability that she draws a red ball on all three turns?	0.462
	Atypical	Three spinners are constructed. The first spinner has 2 equal areas (colored red and blue), the second has three equal areas (red, blue, and white), and the third again has two equal areas (red and white). All three spinners are spun and the result of each spin is recorded. What is the probability of getting 'red' on all three spins?	0.423
Permutations	Typical	An apartment building has four parking spaces in front (call them A, B, C, and D). There are four apartments in the building (#1, #2, #3, and #4), and each apartment has a single occupant with a single car. Every evening, all four occupants come home and park in a randomly chosen space. What is the probability that this evening they park so that the occupant of Apt #1 is in space A, the occupant of #2 is in space B, the occupant of #3 in space C, and the occupant of #4 in space D?	0.462
	Atypical	There are four math students (Ed, Fred, Mary, Pia) waiting to be randomly matched with four math tutors (#1, #2, #3, and #4). Each tutor works one-on-one with a student. What is the probability that Ed will be matched with tutor #1, and Fred will be matched with tutor #2?	0.731
Conditional Probability	Typical	Joe applies for a state-subsidized mixed-income housing project being built in his neighborhood. If he is classified as a low-income applicant, he has a 70% chance of getting an apartment. Applicants not classified as low-income have only a 10% chance of getting an apartment. Joe believes that on the basis of the records he is submitting that he has a 40% chance of being classified as low income. What is the probability that he gets an apartment?	0.500
	Atypical	Assume that in the city of Metropolis, if a criminal defendant in fact committed the crime, he has a 70% chance of being found guilty by the jury. A defendant who is in fact innocent has a 10% chance of being found guilty by the jury. Assume that 40% of defendants who are tried in Metropolis in fact committed the crime. We meet a Metropolis defendant in prison. What is the probability that he is fact committed the crime, given that we know he was found guilty by the jury?	0.077

STUDENTS' EXPECTATIONS OF INTRODUCTORY STATISTICS INSTRUCTORS

MARK A. EARLEY
Bowling Green State University
earleym@bgsu.edu

ABSTRACT

The purpose of this phenomenological study was to talk to students about their experiences taking introductory statistics. The author met with eleven students individually for four interviews throughout the semester, followed by a member-checking focus group during the last week of classes. One of the most salient themes to emerge was the students' reliance on their instructor for feedback about performance, directions on taking notes, and the creation of a classroom environment that motivated them to study. As part of the phenomenological tradition, the author presents his own reflections based on these students' comments. Conclusions include the encouragement of instructors to be more mindful of students' reactions to course content, and suggestions for developing a more learner-centered learning environment.

Keywords: *Statistics education research; Teaching statistics; Statistics classroom; Learning environment; Phenomenology*

1. INTRODUCTION

The current statistics reform paradigm stresses instructors teaching and students learning statistical concepts over mechanics (Chance & Garfield, 2002). One goal of this model is to help students develop *relational* or *structural knowledge* in addition to declarative and mechanical knowledge (Earley, 2001; Schau & Mattern, 1997). With relational knowledge, students go beyond just "knowing that" standard deviation is the square root of variance as well as "knowing how" to compute a standard deviation. How students connect standard deviation with concepts they already know demonstrates their relational or structural knowledge. The shift away from mechanics and toward understanding is one attempt to decrease students' anxiety levels, under the assumption that reducing the mathematical content and rote memorization of definitions and formulae reduces students' worries about course performance (Onwuegbuzie, DaRos, & Ryan, 1997). The most frequently cited implication is the need to develop class activities and assessment tools that are more concept-based and less calculation-based (Gal & Garfield, 1997). Carpenter and Lehrer (1999) discuss understanding in mathematics courses as a "mental activity" as well as something "*emerging or developing* rather than presuming that someone either does or does not understand a given topic, idea, or process" (emphasis added, p. 20). Garfield (1995) warns instructors, "no [teaching] method is perfect and will work with all students" (p. 32) and "teachers often overestimate how well their students understand basic concepts" (p. 31). These statements imply researchers and classroom teachers would be wise to invest time listening to individual students to get a

Statistics Education Research Journal, 6(1), 51-66, http://www.stat.auckland.ac.nz/serj

sense of how well their own teaching methods help students develop conceptual understandings of introductory statistics concepts.

What we are not seeing in the literature are discussions of how students' understandings are impacted by the classroom environment. What is also not heavily discussed in the literature is how students respond to more conceptually-based classrooms (Batanero, Garfield, Ottaviani, & Truran, 2000). Of course, this does not include references to the well known statistics anxiety phenomenon. Few researchers have explored any of our students' *experiences* in statistics courses – there is a more consistent focus on achievement outcomes (Becker, 1996). Gordon stated that "as in other fields, a major challenge currently facing researchers in statistics education is to improve our understanding of learning" (1995, para. 6) and "in order to teach statistics effectively, we must first understand the learners" (2000, p. 16). Gordon's work focuses on activity theory and the social, historical, and cultural aspects of classroom environments to help understand statistics learners (1995, 2000, 2004). Through this perspective, she concluded in a 1995 study that statistics classrooms must be supportive, instructors must provide guidance to students, and teaching must "build on the personal experience of the learner" (para. 47). Nine years later, Gordon (2004) indicated there is still a need for statistics education researchers and statistics educators to understand our learners. To this end, her interview work done for a 2004 report led to five categories of meanings students attach to statistics: statistics as having no meaning, statistics as processes, statistics as mastery, statistics as a tool, and statistics as critical thinking.

These meaning categories parallel those described by Reid and Petocz (2002), in which students described statistics as having (a) a focus on techniques, (b) a focus on using data, and (c) a focus on meaning. In a later report they redefined these conceptions as (a) doing, (b) collecting, (c) applying, (d) linking, (e) expanding, and (f) changing (Petocz & Reid, 2003). In both cases, the authors describe these conceptions as a continuum from limited to expansive. More limited conceptions, according to Gordon (2004), have the potential to limit students' study strategies and motivation for learning statistics. All of these studies suggest understanding students' descriptions of what "statistics" means is an important precursor for improving students' learning and ultimately their experiences in our courses. One question to address is how students create these meanings: Do they come in to class with them, do specific aspects of the classroom environment change initial meanings or create new ones (as Gordon suggests), and do these meanings change over the course of the instruction? In other words, what kind of impact does the learning environment created by instructors have on students' meaning-making?

Petocz and Reid (2002, 2003) have addressed some of these questions by connecting students' conceptions of statistics with expectations students have of statistics instructors. Again, the authors describe these conceptions as most limited to most expansive:

Conception 1: Providing materials, motivation, and structure
Conception 2: Explaining material and helping with student work
Conception 3: Linking statistical concepts and guiding learning
Conception 4: Anticipating student learning needs
Conception 5: Being a catalyst for 'open-mindedness'

The important addition in this work is the *blending* of students' conceptions of statistics and their expectations of instructors. As an example, Petocz and Reid (2004) describe one mix as students' conceptions of "doing" and "collecting" along with students' expectations of their instructors to provide materials, motivation, and structure (Conception 1). This would be the most limited blend of both sets. Although her interviews did not directly ask about students' expectations of their instructors, Gordon

(2004) did interview two instructors and also discusses their impression of the interactions between how students attach meaning to statistics and what goes on in the classroom.

One important aspect of these studies is their reliance on interviews and open-ended survey responses at *one point in time.* Any discussion of how classroom environments and instructors impact students' conceptions should incorporate a more longitudinal data collection strategy. For example, Murtonen and Lehtinen (2003) conducted a study with Finnish students in which they asked participants to write in "learning diaries" throughout the entire term. Students were to record difficult concepts they encountered as well as what they thought led to the difficulties. Students' reasons for difficulties fell into five categories: (a) superficial teaching, (b) difficulty linking the theory of quantitative methods with the doing, (c) unfamiliarity and difficulty with content (e.g., too many concepts to absorb), (d) lack of connections across bits of information, and (e) negative attitudes toward quantitative methods. Similar to the typologies of Gordon (2004) and Petocz and Reid (2003), Murtonen and Lehtinen's categories begin to blend students' understandings of quantitative material with how it is taught. Interestingly, they conclude "the experience of difficulty did not occur because of the major subject or any specific teacher, but because of some more general reason" (p. 182), indicating again that understanding students' backgrounds could be an important precursor to developing more effective classroom environments. What is also important from this work is its longitudinal design. Ultimately, the authors chose to combine responses from throughout the term, however, so we still have no presentation of how individual students' experiences change (if at all) as the course progresses. Petocz and Reid (2003) posit that students "would develop their approach to learning, and maybe change their conceptions of teaching from time to time" (p. 41), suggesting that rather than exploring class marks on assessments, an additional approach to data collection that also explores *individual students' experiences* would help statistics educators understand further what students take away from the course.

This expectation, along with Carpenter and Lehrer's (1999) assertion that mathematical learning is developmental, suggests that expanding this body of work through more longitudinal explorations of students' experiences *throughout* their course would be useful. Talking to students more frequently allows researchers to gather more information about students' experiences, such as: What do students do to prepare for class? What do students do while in class? What goes through students' minds during class? What aspects of class time help or hinder students' understandings of the material? What do students do when they leave class? For these reasons, the general purpose of the current phenomenological study was to begin this exploration with one very broad guiding question: "How do students describe their statistics course at different points *during the term in which they are taking the course*?"

2. METHODS

2.1. THE PHENOMENOLOGICAL TRADITION

The use of phenomenology allows us to understand student experiences differently than through surveys or other qualitatively-oriented traditions. Rather than "averaging" or grouping students' experiences together, phenomenology seeks to present "the qualitatively different ways in which a phenomenon is experienced" (Pietersen, 2002, para. 14). Van Manen (1990) stresses phenomenology as "human scientific study" with an emphasis on "explicat[ing] the meanings as we live them in our everyday experience, our lifeworld" (p. 11). Phenomenological research incorporates the researcher as someone learning about a phenomenon by "borrow[ing] other people's experiences and their reflections on their experiences" (Van Manen, 1990, p. 62). In phenomenology, it is the actual doing and living that constitutes knowledge – "a kind of knowing that can only be obtained through active engagement" (Ladkin, 2005, p. 116). Ultimately, phenomenological research attempts to present the unique lived experience of others so that the reader also develops a new understanding of the experience.

An essential first step in conducting phenomenological research is for the researcher to first understand his or her own experience with the phenomenon under investigation (commonly referred to as *bracketing* or *époche*) (Ladkin, 2005; Laverty, 2003). As one reads different descriptions of phenomenology, this process is part of what distinguishes the various approaches (Kerry & Armour, 2000; Laverty, 2003). Rather than completely "suspend" our own understandings, as Edmund Husserl's description of époche would have us do (Creswell, 1998), hermeneutic phenomenology, most commonly attributed to Hans-Georg Gadamer, (Kerry & Armour, 2000; Ladkin, 2005), incorporates these understandings as essential to the interpretive process to follow. In other words, because the researcher will interpret participants' experiences while developing a description of the phenomenon, an understanding of the researcher's own experiences is necessary so the reader has an idea of the researcher's interpretive lens. Van Manen (1990) describes this as the "intersubjective" nature of phenomenological research: "the human science researcher needs the other (for example, the reader) in order to develop a dialogic relation with the phenomenon" (p. 11).

This also means the researcher's own understanding develops during the research process – in his description of the phenomenological data collection process, van Manen refers to this as the "reflective data" (2002; see also Laverty, 2003) researchers collect alongside the "empirical data" (i.e., interview data) from our participants. van Manen (1990) indicates social science researchers "borrow" participants' experiences in order to develop a complete description of the experience. In this sense, the researcher is developing the description as his or her own developing understanding of the phenomenon under study. Kerry and Armour (2000) indicate "these personal histories lead to a unique perception of different experiences and that this personal history cannot be bracketed out; it is fundamental for interpretation" (p. 6). For example, Finney (2000) presents her own experience with success in statistics courses prior to detailing six other students' success experiences. The reader then understands how her interpretation and summary of the experience of success is shaped by her own initial perceptions. Mayhew (2004) also presents his experiences with spirituality prior to describing the meanings eight undergraduate students attribute to spirituality "to give the reader a sense of any prejudice or orientation that may have shaped [his] interpretation or approach to the study" (p. 656).

2.2. PARTICIPANT SELECTION

In phenomenological research, the researcher bases participant selection on finding and recruiting individuals who have experienced the phenomenon of interest (rather than selecting key informants in ethnography, or selecting individuals based on their life experiences for a life history) (Creswell, 1998; Laverty, 2003; Moustakas, 1994; van Manen, 1990). The quantity and depth of the data to be collected determines how many participants the researcher can recruit. In phenomenological inquiry, the primary data source is the interview, and depending on the phenomenon of interest, there may be more than one interview per participant. The researcher must have each interview transcribed for analysis, creating a large set of data to read through and analyze, so the number of participants is generally kept to a minimum (Creswell, 1998). There are no specific guidelines on what constitutes this "minimum" – "Boyd (2001) regards two to 10 participants … as sufficient to reach saturation and Creswell (1998, pp. 65 & 113) recommends … 10 people" (as cited in Groenewald, 2004, p. 11). Because I was interested in interviewing participants four times during the semester and interviewing all participants within a week of each other, a number around ten seemed manageable and was my goal.

Six departments across four colleges at our institution offered 40 sections of introductory statistics courses during the spring 2003 term (this includes multiple sections of the ten different courses). Students in this study came from multiple sections of four of these courses. All of these courses are traditional in scope, covering introductory descriptive and inferential statistical analyses in one term. The "Psychology 270" course is the required introductory course for psychology majors. The "Statistics 211" is the first in a required two-course sequence for all majors in the College of Business Administration. The "Math 115" course is offered to students not majoring in mathematics; many undergraduates take this course to fulfill the mathematics requirement for the university if a different course is not specified by their College. Mathematics majors take a different introductory statistics course, so the "Math 115" course is not calculus-based. The "EDFI 641" course is a required course for master's level students in the College of Education and Human Development. All of these departments offer more advanced courses, but only students in the College of Business Administration are required to take a second course.

My goal for recruiting participants was to get a mix of participants from across all of these courses. Eligible students included those over the age of 18 and enrolled in any introductory statistics course during the spring 2003 term. In January 2003 I posted flyers next to the doors of each classroom in which one of the 40 introductory statistics sections met. This method of recruitment does introduce typical volunteer bias, but the advantage is that if students are interested in talking about their experiences (whether good or bad), they will volunteer (a necessary participant characteristic in phenomenology). The flyer asked students to contact me via phone or e-mail if interested in taking part in my study, and I set up times to meet with the twelve who did so by the end of the third week of classes in order to have a first interview for each participant completed during the third or fourth week of the term. At our first interview, I informed participants of a $120 incentive to those who completed the study by returning for interviews throughout the semester. Although the issue of incentives is still controversial (Adler & Adler, 2002), I do believe in giving participants something in appreciation of their time (Weiss, 1994). I do not feel this had an impact on the quality of my participants' contributions because each participant was engaged during our conversations and each had much to say about their experiences in their statistics class.

To give the reader a general description of the participants, Table 1 lists demographic information for the final set of eleven students I interviewed (one student dropped out of the study after the first interview). I did achieve the variety I initially wanted: My group includes participants from freshman to senior with one graduate student, a grade point average (GPA) range of 2.00 – 4.00 (i.e., students fell across the general academic performance spectrum from "average" to "excellent" students), and students from courses in four of the six departments offering statistics courses in the spring (no students from the Sociology or Criminal Justice courses contacted me).

Table 1. Demographic summary

Participant [a]	Gender	Course Taken [b]	GPA [c]	Class
Aileen	Female	Psychology 270	3.60	Senior
Alastair	Male	Statistics 211	2.60	Sophomore
Brice	Male	Psychology 270	4.00	Senior
Brigit	Female	Math 115	3.77	Sophomore
Cecily	Female	EDFI 641	4.00	1st Year Graduate
Cedric	Male	Statistics 211	2.92	Freshman
Dillan	Male	Math 115	2.90	Sophomore
Eleyn	Female	Math 115	4.00	Freshman
Evan	Male	Statistics 211	2.50	Sophomore
Griffin	Male	Math 115	2.00	Freshman
Ian	Male	Statistics 211	3.66	Sophomore

[a] All names used in this report are pseudonyms.
[b] All courses are introductory-level statistics courses designed for non-statistics majors.
[c] GPA = Grade Point Average, calculated as an average of the students' grades for all of their coursework to date after converting the letter grades to "grade points": A = 4, B = 3, C = 2, D = 1, F = 0.

2.3. DATA COLLECTION

Phenomenological data collection proceeds in two parallel processes: (a) empirical data collection, typically in the form of in-depth interviews with participants, and (b) reflective data collection, in which the researcher reflects on his or her interactions with the empirical data (van Manen, 2002). The power of phenomenological inquiry is this interaction, where the researcher first presents his or her own experiences with the phenomenon of interest, and then *reflects on how his or her experience and understanding of the phenomenon changes and develops while listening to the participants' stories* (Moustakas, 1994).

I met with each student four times for one-on-one interviews in my office (the only exception was Ian, who started the study late and only met with me three times). We spaced interviews approximately 3 to 4 weeks apart depending on the student's schedule. I had an initial list of questions I wanted to discuss, based in part on my interests and in part on themes emerging from the previous interviews. Each interview began, however, with the general question, "So how is class going for you?" From there, we discussed whatever the student brought up as important at that time. Interview times ranged from 15 to 45 minutes, entirely dependent on how much the student had to say. I audio taped the interviews and took fieldnotes during our discussion. I then transcribed audiotapes, taking additional fieldnotes as I listened to our discussion again.

To increase the accuracy and representativeness of the themes I saw emerge (Maxwell, 2005), I met with nine of my participants in two member-checking focus groups during the last week of classes prior to final exams. "Member-checking" is described as a process of going back to the participants to have them review the researcher's interpretations of and conclusions drawn from the data. Through this process, participants comment on whether the researcher "got it right" (so to speak) (Creswell, 1998; Glesne, 1999; Maxwell, 2005). Due to the scheduling so close to the end of the term and final exams, I gave students two options: an afternoon time with lunch, selected by two participants, or an evening time with dinner, selected by seven. Two of my participants could not make these focus groups due to scheduling conflicts. I presented students with a summary of the major themes on slides, one at a time, and solicited group feedback on whether these were actually important to the students. To get this feedback, I asked students to think about what they would tell a group of statistics instructors about their experiences taking statistics. Students changed the wording on some (e.g., the course is "easier when instructors …" not "easy when instructors …," as I had first worded one of the themes), but otherwise agreed these were important ideas to them and ideas they would share with statistics instructors as important.

2.4. DATA ANALYSIS

To come to an understanding of how my participants experienced their statistics courses, I analyzed and reflected on the transcribed data throughout the study. After each interview, I took time to make notes on ideas the student seemed to stress or return to often. I also noted student comments that, for some reason, stood out for me. After each *round* of interviews, I explored major patterns across all 11 participants and developed questions for the next round of interviews. After the study ended, I also developed time-ordered displays (Miles & Huberman, 1994) of students' main comments from the beginning to the end of the term to explore changes in their descriptions of class. Because I developed interview protocols based in part on the ongoing analysis of each round of interviews (i.e., what did I find important or interesting that I wanted to explore again in the next interview?), I include here a list of the major "topics" for each round of interviews. I introduced these additional topics into the interview after the opening question about how class is going.

Interview 1: Discuss a general description of the class, what happens during class, what student does in between class meetings, other issues student brings up.

Interview 2: How does student take notes? What does student do with notes during and outside of class time? Why does student take notes? Does student only take notes in class, or does student read the textbook and take notes as well?

Interview 3: What would happen if there were no identifiable instructor? What would the student miss?

Interview 4: Summary of student's thoughts about course, additional comments, and comparison of their statistics course with another course they are taking (students chose one of their other courses and we discussed similarities and differences between that course and their statistics course).

All participants, after the first interview, came prepared to answer the "How is class going?" question – they clearly had topics they wanted to share with me. After this discussion, I introduced questions addressing each of the ideas listed above.

3. RESULTS

3.1. BRIEF REFLECTIONS ON MY OWN EXPERIENCE

My experiences taking statistics courses began with a required course for my (then) economics degree. I clearly remember a lot of frustration and confusion as the instructor filled blackboard after blackboard with Greek symbols and mathematical formulae. We used a computer software package for homework problems, without the benefit of a user's manual or any training on how to use the software (back then, the software required writing programs to run analyses as opposed to today's "point-and-click" platforms). My solution to this confusion and lack of understanding was to give in, earn a D-, and retake the course the following term.

This new instructor was so radically different from my first – primarily because he actually made class enjoyable. I *wanted* to be in class, and worked hard between class meetings in order to participate fully each week. We used computer software, but now he guided us through the programming during class time before sending us out on our own. I recall measuring 100 leaves from trees around campus for one of our early homework exercises – data we would analyze and discuss throughout the course. After being in class with this instructor (and concurrently taking an advanced economic theory course in which I was struggling), I switched my major to mathematical statistics at the beginning of my sophomore year and finally felt "at home" with my coursework.

Although I have had many (many) statistics courses since my start in 1989, I turn now to some brief reflections on my observations as an instructor of introductory statistics in education courses (note that none of the students in my classes participated in this study). Many of these informal observations are what led me to develop an interest in finding out what exactly my students experience when they take my course. Have I avoided creating a classroom environment similar to that very first environment in which I failed? Have I created a motivating and interesting classroom environment similar to my *second* experience (along with many other inspiring instructors and classes along the way)?

No matter what I do or say, students consistently come in to my course with the typical fear and anxiety. It seems at times that I am an actor on a stage, begging my audience to laugh and enjoy themselves. After careful attention to the textbook I use, the way I structure class meetings, the assessment system, and my own concern for students' success, some students still carry that fear and anxiety throughout the term. I hear from students that the textbook is confusing and impossible to read. I hear from students how they can follow along in class, but seemingly all is lost when they leave the classroom; they have no idea what they are looking at when they study at home. Student evaluations of the course consistently rate "relevance to my profession/degree program" the lowest of all evaluation items. Students come in the first day and announce this is their last course – all they want is their grade of "C" so they can graduate and move on.

The power of phenomenological inquiry is the "connections" researchers and participants can make through their shared experiences. Moving into this particular study, all of my own experiences, attitudes, beliefs, and values related to introductory statistics form the context in which I talk to participants, listen to their words, and try to make sense of what we discuss. Rather than considering this a major bias to the study, however, this context allows the reader to understand this narrative as *my understanding* of these students' experiences (as is the case with any naturalistic inquiry process). I have paid close attention to verbatim transcripts, and my reflections at the end of this narrative point to the "surprises" and discussion topics that made me reflect deeper on my role as instructor. The member-checking focus groups conducted with my participants helped validate the themes I saw emerge, to which I now turn.

3.2. THE VOICES OF MY PARTICIPANTS

Even with the variety of students I talked to, as well as the diversity of their classroom environments, three common themes emerged across all of our discussions: (a) *student behaviors and characteristics,* (b) *instructor behaviors and characteristics,* and (c) *resources students use for studying.* This report focuses on the second of these: the students' perceptions of the instructor and his or her behaviors. During the first two sets of discussions, this idea came across in a variety of my participants' responses to questions, regardless of whether I had directly asked about the instructor. Based on this preliminary analysis, I asked students the following question during our third discussion:

> *Imagine you had to take this course entirely online. There is no identified instructor now – you read texts and examples online, complete assignments and tests online, and receive scores via e-mail. What would that be like?*

My intent here was to focus on what students would miss about their instructors, having discussed them so much during our first two interviews.

Responses to this question, along with questions from earlier discussions, fell into three major sub-themes:

(1) The instructor's use of class time (including pace and teaching style),
(2) The instructor's role in the note-taking process, and
(3) The instructor's assessment strategies (including homework and exams).

I present next my interpretations of each of these themes along with representative quotes from my participants' responses.

The instructor's use of class time One of my early interests in conducting this study was to find out what it is like just to be in a statistics class: What are students thinking, what kinds of behaviors do they feel are important during and between class meetings, and most importantly, *why*? For most students, the response to what they do in class was "sit there and take notes." What thoughts did they have during class? For some, they were random because, as Brigit said,

Brigit: It's hard to stay focused on it because all she does is write definitions from the book onto the board and then just reads them and explains them a little bit. Very rarely we go through examples of things but mostly it's just definitions.

For Brigit and others, instructors organized class time around the traditional "lecture" approach with very little interaction among students. There were exceptions, in particular Griffin's instructor, who did no formal instruction but rather assigned group work from the text.

Griffin: He gives us problems to do from the book and then sits up front and watches us. He is there to answer questions. If we don't have any questions, we finish our exercises and leave.
Int: Do you like that approach?
Griffin: Well he seems pretty unmotivated to teach the class and not interested at all.
Int: What do you think of that?
Griffin: There is no reason for me to do any work outside of class. Why should I care if he doesn't?

When asked about the online course, only Griffin was excited at the prospect of "taking the class while [he stays] in bed." The remaining ten participants each expressed concern that this type of setting, most notably the absence of a live instructor, would never work for statistics. Cecily was most vocal (her eyes nearly popped out of her head when I asked her that question) in wanting to avoid this type of course.

Cecily: Wow … no way of asking questions? There's gotta be a way of asking questions … yeah I may be able to pass the class but I wouldn't know it. I wouldn't be able to use it.
Int: So you need the instructor there to ask questions to.
Cecily: Absolutely. I have a math background, I'm pretty good at math. But this is not just a

> math class … there's a lot of concepts here that just reading the book [won't work] …

Cedric also expressed concern over the "completely online" approach.

Cedric: I don't think that would work at all.
Int: Why not?
Cedric: Because, for example, I'll hit a stumbling block where I don't understand it, but then I talk to a peer and that sheds a different light on it. It makes it so much easier. Without that interaction between students it makes it more difficult, especially because the instructor can shed another different light on it.
Int: So having the different viewpoints are helpful.
Cedric: Yes. The textbook is good at telling you what to do, but not as good at telling you why this happens.
Int: So is that something the instructor has been helpful for? Presenting the whys?
Cedric: Yes.

Overall, students who had instructors giving *and explaining* examples in class were more comfortable with their experience – students like Cedric who were not getting these examples wished their instructors would do more of "the whys" instead of just "the whats."

I clearly recall my own desire to skip class meetings the first time I took statistics contrasted with my excitement to be in class the second time – student comments about how the instructor guided class time resonated strongly with me. As an instructor, I am keenly aware of students who do not appear engaged or "with me" – most likely a result of my own early experiences. My participants integrated the first four of Petocz and Reid's (2003) conceptions into their expectations of how class time is used: They expect instructors to be organized, present more than just an outline of the text, explain concepts beyond what the text offers, and be there to address student questions and provide alternative explanations when needed. This is different from Petocz and Reid's analysis, in which students could be described by one or two of the conceptions – all 11 of my participants discussed all four of these conceptions throughout the term.

The instructor's role in the note-taking process A natural progression from the first sub-theme on instructor behaviors and characteristics is how these behaviors influence students' note-taking processes. Students consistently reported relying predominately on their notes for studying course material. This is also how they each spent their time in class (again with the exception of Griffin noted above) – listening to the instructor and writing down whatever he or she wrote on the blackboard or overhead. I asked students how they decided what to write down, and most indicated it was "whatever [or everything] the instructor writes down." Only Brice and Cedric mentioned they also jot down additional notes to themselves about key formulae or concepts to remember.

Evan, Ian, and Alastair each had instructors who provided handouts with full or partial notes already on them. For these students, this meant they just followed along and maybe filled in an example or two – Evan called this "being spoon fed" the information.

Evan: He passes out sheets and he goes over the stuff in class … I remember a lot of it from last semester, so I'll go through and do the stuff that I know. And I'll just sit around and wait for a question that I didn't know and fill in the answer as he explains it.
Int: So then when you're in class you're going through and doing on your own the stuff that you remember and then waiting?
Evan: If I hear something that I didn't write down or don't remember from last semester,

I'll write that down.

Int: Do you take any other notes besides those on the sheet he hands out?

Evan: No he usually has a lot of space on the side to put extra stuff there. He spoon feeds us most of it.

With or without these handouts, students rely very heavily on their notes for studying outside of the classroom. Cedric further indicated that he has experienced the "false safety net" of feeling as though he is following along in class, not taking any notes as a result, and then being lost when he got home to study for the test (having no notes to refer to). Here again, students all mentioned either being happy with an instructor who walked through examples in class (so these could go into their notes) or being unhappy with an instructor who did not do so.

Three of my participants had sold their books back by the first interview (within a month of class starting), another two sold them back later in the term, and five of the remaining six students used their books for assigned homework problems only. Only Cecily read the textbook in between classes, though she indicated this was difficult to do. Brigit and Dillan struggled with this the most because their instructor "lectured" directly from the book.

Brigit: This class is different [from my other classes] because instead of having the professor's point of view on the notes, it's just point blank – I look at my book and I looked at my notes and it's word for word.

Int So what do you do then?

Brigit I have random thoughts during class. It's hard to focus when it's just like being dictated to. Plus I know if I miss anything it'll be in the book.

Dillan indicated it was almost "stupid" to go to class because he can sit at home and read the book – what he and other students want from instructors is explanations and "different viewpoints," not dictation. The textbook was not enough for them to learn the concepts, and the instructor in this case was not serving as an extra resource, which bothered them. For these students, and students who cannot or do not read the text even though they keep it, there is an even greater importance placed on course notes.

The instructor's assessment strategies There was an overwhelming consensus from all of my participants that regular homework and frequent assessment are necessary in order for students to know whether or not they are learning anything. As Brice put it:

Brice: Each person in the class was left on their own to make sense of what was being presented … maybe it felt like a struggle because if we'd had some more problems … we'd have some confirmation of whether or not we understood the material. Maybe it was a struggle because we just didn't know for sure if we understood it or not – we just really didn't have any confirmation of that.

Brice and others were concerned that with no practice, no regular homework, and no assignments to work on, preparing for exams was more difficult and led to an increase in anxiety prior to exams. Cecily mentioned that "a computer can give you the answer, but it doesn't give you how it got the answer," and so for her the additional need was for the instructor to *go over* the assessments when returning them to students.

Students also indicated that "what the instructor told us would be on the test" was their study guide, and any deviation from this on tests and quizzes was frustrating for students. Although I did not look at individual assessments unless the student asked me a question about them, what they perceived to be poor test development clearly impacted their class experiences. A major challenge was tests based on "applications" when instructors did not give students application models in class. Even though many students called for and recognized the need to understand what it is they were doing, they wanted instructors to guide them in this process. Cedric's first exam experience was particularly enlightening:

Cedric: I wasn't happy with my first test.
Int: What happened?
Cedric: Right answer, wrong work, so a lot of points were taken off. The one thing is, she doesn't accept just the answer. Even if you do some work, unless you do the right work it will be marked wrong. I can understand that if we're majors, but we're general business majors. Just getting the right answer is enough. Recognizing wrong answers is good, but not exactly knowing step-by-step.

This conflict between Cedric's realization that understanding is important (from earlier conversations) but not being comfortable when tested on this understanding is important to me. These data highlight the need for instructors to match more closely their instruction and assessment procedures so students do not get these conflicting messages. These data also suggest that instructors need to make explicit connections between instruction and assessment, as instructors' intentions are not always clear.

4. DISCUSSION

All of these sub-themes are intimately related to each other, as well as to the two other larger themes of student characteristics and behaviors and use of resources. Students in this study spoke frequently about the role the instructor plays in their class experiences, but for me the surprising aspect of all of our conversations was the *extent* to which students relied on their instructors. What we do in the classroom has far-reaching impacts beyond how students feel during class. When on their own, students need resources that will help them make sense of class material (Petocz and Reid's (2003) "Conception 3," p. 46). Those participants most comfortable with their experiences indicated they could take their notes home, complete homework assignments by referring to their notes as necessary, and received consistent feedback from their instructor as to whether they were indeed "getting it." Oathout's (1995) participants echoed "strongly favor[ing] frequent tests" (p. 50) and a lack of "mapping from lectures and assignments to test content and format [as] equally problematic" (p. 48).

Many students, however, did not do much work outside of class because graded assignments were either non-existent or infrequent. This meant their statistics course was the lowest priority class when studying during the week, and they rarely did anything with their statistics material outside of class time. This changed around exam time, when there was a need for notes to be clear and complete, and for these notes (typically verbatim from instructor notes in class) to be reflective of what would be assessed on the exams (similar to Petocz and Reid's (2003) "Conception 1" on teaching statistics, p. 44). The exams became the focal point for these students – they took notes so they would have something to study so they could do well on the exams. As Garfield (1995) states, "students learn to value what they know will be assessed" (p. 32) – Petocz and Reid (2003) discuss this as "doing" statistics (p. 42). Students in the current study usually waited until an exam was coming up to actually go back over their notes, so any questions they had could not be addressed and they were sometimes still confused and unsure going into the exam. How the instructor conducts class, how the students record classroom events and discussions in their notes, how the instructor assesses students in between exams, and how the instructor develops exams all become a single system to which instructors need to pay close attention.

Most importantly, students never talked about "getting it" in terms of understanding what they were doing. Although Cedric indicated that knowing why they are doing an analysis makes doing the analysis easier, students did not ever indicate they understood course material. Rather, they were happy if they could get problems correct and earn all of the possible points. Most of the instructors these students spoke of appear to encourage the "doing" aspect of statistics – solving problems is a large part of how these students spent their time both in and out of class. Students knew their tests would involve solving problems (based on what their instructors told them about the tests), so this was how students studied. If we want students to hold the more expansive conceptions of statistics that include "statistics as critical thinking" (Gordon, 2004) and a "focus on meaning" (Petocz & Reid, 2003), we need instructors to hold and espouse these conceptions as well. These data suggest that future researchers explore these same conceptions with instructors to discern their own conceptions as well as how they create learning environments that support students developing the more expansive conceptions.

5. REFLECTING ON MY OWN CLASSROOM

As I reflected on the discussions I engaged in with these students, my thoughts turned to my own teaching philosophy and classroom environment. Listening to them talk about their instructors and class experiences, I often asked myself, "Do I do that in *my* classroom?" Reflective questions about my own practices that developed because of these discussions, questions I believe worthy of further exploration both in my own practice and in the larger statistics education community, include:

(1) How much do I actually stress "doing" statistics over "understanding"?
(2) What are my expectations of students during and outside of class meetings? Am I clear in expressing these expectations to my students?
(3) What role does the textbook play in my course?
(4) How can I best assist students in taking complete and accurate notes?
(5) What is the connection between each class activity and/or discussion and my assessment of student understandings?

I firmly believe as instructors we need to spend more time reflecting on our own teaching philosophies and practices, and for me, this reflection requires student feedback. We can focus our reflections on each aspect of the classroom, understanding which piece(s) are working *for our students* and which are not. Statistics education researchers can also consider these questions, and begin to focus research more on these "emerging understandings." Through these explorations, we can continue to learn how students are responding to our desire for them to walk away from the course with some knowledge of what and why various concepts were covered, rather than walking away with the feeling that all they did was "plug and chug."

6. CONCLUSION

Although it appears intuitive that students rely on their instructors in any course, there are some implications that may not be as intuitive. Based on my own observations as an instructor, as well as comments provided by the students in this study, a primary motivation for students in introductory statistics (those who are not statistics majors) to do anything (take notes, work problems, study outside of class, etc.) is how it will impact their final grade (Garfield, 1995). Students further mentioned that they appreciate feedback on graded assignments as a way to make sure they know what the instructor wants them to know. Here we see a direct connection between how we assess students, the opportunities we provide for assessment (e.g. homework, in-class activities, quizzes, tests, projects), and *why* we even assess them (Chance & Garfield, 2002).

Assessment systems in any course gauge whether or not the students are meeting course objectives. As we work to develop introductory statistics courses oriented more toward understanding and less toward mechanics, our assessment systems must change as well. As we change our assessments, it is important to remember that our classroom environment must also change to encourage students' development of understanding over their mastery of mechanics (Treagust, Jacobowitz, Gallagher, & Parker, 2001). Demetrulias (1988) argues that an "opportunity to understand statistics from an integrated and flexible viewpoint *must go along with a classroom environment that rewards such exploration*" (emphasis added, p. 169). As students in this study report, what happens in class becomes their main resource for any work they do outside of class (Petocz & Reid, 2003). Spending time in class (and with assessments) on "how to understand" as much as "what to understand" becomes critical, and instructors must continue to develop and make use of delivery techniques, classroom examples, and homework activities that focus on this "how to understand" piece. I also encourage sharing these reflections and experiences in the literature, so that others may benefit from what we learn in our own classrooms. Results from this study suggest the need for researchers to encourage more personal accounts of actual experiences of instructors teaching, and students taking, statistics, as opposed to relying only on the more quantitative outcomes currently presented (Becker, 1996). Instructors and their students are "in the trenches," experiencing, reflecting on, and ultimately determining whether any particular classroom environment is or is not successful in increasing students' understandings (Batanero, Garfield, Ottaviani, & Truran, 2000).

ACKNOWLEDGEMENTS

I presented a preliminary version of this paper at the 2003 Annual Meeting of the Mid-Western Educational Research Association. My gratitude goes to Beverly Dretzke for her thoughtful feedback as discussant of that paper session and to Tony Onwuegbuzie for his insightful comments on a draft of this manuscript. I am also indebted to the editors and reviewers who carefully read and provided insightful feedback on earlier drafts of this manuscript. Of course, I owe much to my eleven participants for meeting with me throughout the semester and sharing their thoughts and experiences.

REFERENCES

Adler, P. A., & Adler, P. (2002). The reluctant respondent. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 515-536). Thousand Oaks, CA: Sage.

Batanero, C., Garfield, J. B., Ottaviani, M. G., & Truran, J. (2000, May). Research in statistical education: Some priority questions. *Statistical Education Research Newsletter, 1*(2), 2-6.
[Online: www.stat.auckland.ac.nz/~iase/serj/newsmay00.pdf]

Becker, B. J. (1996). A look at the literature (and other resources) on teaching statistics. *Journal of Educational and Behavioral Statistics, 21*(1), 71-90.

Boyd, C. O. (2001). Phenomenology the method. In P. L. Munhall (Ed.), *Nursing research: A qualitative perspective* (3rd ed., pp. 93-122). Sudbury, MA: Jones and Bartlett.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. A. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 19-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal, 1*(2), 38-44.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ1(2).pdf]

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks, CA: Sage.

Demetrulias, D. M. (1988). (Creatively) teaching the meanings of statistics. *The Clearing House, 62,* 168-170.

Earley, M. A. (2001). *The development of knowledge structures in introductory statistics*. Unpublished doctoral dissertation, University of Toledo, Toledo, OH.

Finney, S. J. (2000, April). *The meaning of success for students in statistical methods courses: A phenomenological study.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (Eric Document Reproduction Service No. ED441026)

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

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*(1)*,* 25-34.

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

Gordon, S. (1995). A theoretical approach to understanding learners of statistics. *Journal of Statistics Education,3*(3).
[Online: www.amstat.org/publications /jse/v3n3/gordon.html]

Gordon, S. (2000, May). Bibliography on activity theory and related research in education. *Statistical Education Research Newsletter, 1*(2), 16-22.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/newsmay00.pdf]

Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal, 3*(1), 40 - 59.
[Online: www.stat.auckland.ac.nz/~/iase/serj/SERJ3(1)_gordon.pdf]

Groenewald, T. (2004). A phenomenological research design illustrated. *International Journal of Qualitative Methods, 3*(1). Article 4.
[Online: www.ualberta.ca/~iiqm/backissues/3_1/pdf/groenewald.pdf]

Kerry, D. S., & Armour, K. M. (2000). Sport sciences and the promise of phenomenology: Philosophy, method, and insight. *QUEST, 52*, 1-17.

Ladkin, D. (2005). 'The enigma of subjectivity': How might phenomenology help action researchers negotiate the relationship between 'self', 'other' and 'truth'? *Action Research, 3*(1), 108-126.

Laverty, S. M. (2003). Hermeneutic phenomenology and phenomenology: A comparison of historical and methodological considerations. *International Journal of Qualitative Methods, 2*(3). Article 3.
[Online: www.ualberta.ca/~iiqm/backissues/2_3final/pdf/laverty.pdf]

Maxwell, J. A. (2005). *Qualitative Research Design: An Interactive Approach* (2nd ed.)*.* Thousand Oaks, CA: Sage.

Mayhew, M. J. (2004). Exploring the essence of spirituality: A phenomenological study of eight students with eight different worldviews. *NASPA Journal, 41*(3), 647-674.

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

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methodology courses. *Studies in Higher Education*, *28*(2), 171-185.

Oathout, M. J. (1995, April). *College students' theory of learning introductory statistics: Phase one.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (Eric Document Reproduction Service No. ED391841)

Onwuegbuzie, A. J., DaRos, D., & Ryan, J. (1997). The components of statistics anxiety: A phenomenological study. *Focus on Learning Problems in Mathematics, 19*, 11-35.

Petocz, P., & Reid, A. (2002). How students experience learning statistics and teaching. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/6b4_peto.pdf]

Petocz, P., & Reid, A. (2003). Relationships between students' experience of learning statistics and teaching statistics. *Statistics Education Research Journal, 2*(1), 39-55. [Online: www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf]

Pietersen, C. (2002). Research as a learning experience: A phenomenological explication. *The Qualitative Report, 7*(2), 1-13.
[Online: www.nova.edu/ssss/QR/QR7-2/pietersen.html]

Reid, A., & Petocz, P. (2002). Students' conceptions of statistics: A phenomenographic study. *Journal of Statistics Education, 10*(2).
[Online: www.amstat.org/publications /jse/v10n2/reid.html]

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 91-104). Amsterdam: IOS Press and the International Statistical Institute.

Treagust, D. F., Jacobowitz, R., Gallagher, J. L., & Parker, J. (2001). Using assessment as a guide in teaching for understanding: A case study of a middle school science class learning about sound. *Science Education, 85*(2), 137-157.

van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy.* New York: SUNY Press.

van Manen, M. (2002). *Phenomenology Online.*
[Online: www.phenomenologyonline.com]

Weiss, R. S. (1994). *Learning from strangers: The art and method of qualitative interview studies*. New York: The Free Press.

MARK A. EARLEY
Division of Educational Foundations & Inquiry
Bowling Green State University
554 Education Building
Bowling Green, Ohio, USA
43403-0246

PAST CONFERENCE

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

The second biennial United States Conference on Teaching Statistics (USCOTS 07) was held on May 17-19, 2007 at the Ohio State University in Columbus, Ohio, hosted by CAUSE, the Consortium for the Advancement of Undergraduate Statistics Education. The target audience for USCOTS is teachers of undergraduate and AP statistics, from any discipline or type of institution. Teachers from two-year colleges are particularly encouraged to attend.

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

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

Statistics Education Research Journal, 6(1), 67-75, http://www.stat.auckland.ac.nz/serj

FORTHCOMING IASE CONFERENCES

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

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

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

TOPICS

Submitted research papers address questions such as the following:

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

LOCAL SRTL-5 ORGANIZERS

Janet Ainley, janet.ainley@warwick.ac.uk
Dave Pratt, dave.pratt@ioe.ac.uk

For more information visit the SRTL-5 website: http://srtl.stat.auckland.ac.nz/

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

The meeting will be held on 19-21 August 2007 in Guimarães, Portugal, immediately prior to ISI 56 in Lisbon. The fascinating historic city of Guimarães is about 50 km from Portugal's second largest city, Oporto. This Satellite will involve papers on many aspects of assessing student learning in statistics. Over 40 papers will be presented along with a number of posters and discussions of examination questions. Proceedings will be available on CD and free at the publication page of IASE.

Non-participants must register by 31 May 2007.

CONFERENCE COMMITTEE

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

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

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

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

IASE SPONSORED IPMS (ORGANIZERS, PRESENTERS, DISCUSSANTS)

IPM37 *Research on Reasoning about Distribution*
Organizer: Joan Garfield, USA
Presenters: Rolf Biehler, Germany; Jane Watson, Australia; Chris Reading, Australia
Discussants: Roxy Peck and Beth Chance, USA

IPM38 *How Modern Technologies Have Changed the Curriculum in Introductory Courses*
Organizers: Lucette Carter, France; Catherine Pardoux, France
Presenters: Cecily Peters, Malaysia; Brigitte Chaput, France; Mathilde Mougeot, France
Discussants: Carmen Capilla, Spain; Robert Gould, USA

IPM39 *Preparing Teachers of Statistics*
Organizer: Allan Rossman, USA
Presenters: Carmen Batanero, Spain; Penelope Bidgood, UK; Verônica Yumi Kataoka, Brazil; Madhuri Mulekar, USA
Discussant: Jerry Moreno, USA

IPM40 *Research on the Use of Simulation in Teaching Statistics and Probability*
Organizer: Rolf Biehler
Presenters: Nicolas Christou, Ivo D. Dinov and Juana Sanchez, USA; Joachim Engel, Germany; Andrew Zieffler and Joan B. Garfield, USA
Discussant: Andrej Blejec, Slovenia

IPM41 *Optimizing Internet-Based Resources for Teaching Statistics*
Organizers: Roxy Peck, USA; Ginger Holmes Rowell, USA
Presenters: Mary Townsend, Canada; Iddo Gal and Dani Ben-Zvi, Israel; Dennis Pearl, USA
Discussant: Irena Ograjensek, Slovenia

IPM42 *Observational Studies, Confounding, and Multivariate Thinking*
Organizer: Milo Schield, USA
Presenters: Donald Rubin, USA; Nancy Wermuth, Sweden and David Cox, UK; James Nicholson, Jim Ridgway, and Sean McCusker, UK; John Harraway, New Zealand

IPM43 *Teaching of Official Statistics*
Organizer: Sharleen Forbes, New Zealand

IPM44 *Teaching of Survey Statistics*
Organizer: Steve Heeringa, USA
Presenters: Don Royce, Canada; James J. Brown, UK; Marie-Christine Ponsonnet, France; Giulio Ghellini, Italy
Discussant: Graham Kalton, USA

IPM45 *Studying Variability Through Sports Phenomena*
Organizer: Brian Phillips, Australia
Presenters: Tim Swartz, Canada; Stephen Clarke, Australia; Phil Everson, USA; Kaznori Yamaguch, Michiko Watanabe, and Fumitake Sakaori, Japan
Discussant: Larry Weldon, Canada

IPM46 *Use of Symbolic Computing Systems in Teaching Statistics*
Organizer: Zaven Karian, USA

IASE ORGANIZING COMMITTEE

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

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

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

The International Commission on Mathematical Instruction (ICMI, http://www.mathunion.org/ICMI/) and the International Association for Statistical Education (IASE, http://www.stat.auckland.ac.nz/~iase/) are pleased to announce the Joint ICMI/IASE Study Statistics Education in School Mathematics: Challenges for Teaching and Teacher Education. The conference is co-sponsored by the American Statistical Association and endorsed by the Interamerican Statistical Institute, Mexican Statistical Association and the International Statistical Literacy Project.

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

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

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

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

INTERNATIONAL PROGRAMME COMMITTEE

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

Chris Reading (Australia)

LOCAL ORGANISING COMMITTEE

Blanca Ruiz (Chair)
Ernesto Sánchez
Tomás Sánchez
Armando Albert

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

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

CONFERENCE THEME

The realization that data is preferable to anecdote or intuition as a basis for robust decision making is spreading through many professions and sections of society. More and more, people want to see "the evidence". Statistical methodology and modelling are increasingly pervading the research fabrics of all fields that advance by employing empirical enquiry. And because the root purpose of statistics is to extract insight and meaning about real contexts using data, statistics educators are increasingly realizing that this cannot be modelled by teachers without the use of rich, real contexts. It is important that data and contexts pervade statistical learning and teaching, to help students understand the nature and value of the statistical sciences, and to facilitate their learning. Successful learning processes involve data and contexts that are meaningful to students. These can be relevant to everyday life or to disciplines as varied as psychology, biology, business, sociology, engineering, the health sciences and statistics itself. But many questions remain about the myriad ways in which we can exploit context to achieve our educational goals. We also must look hard at how well we use the data and contexts that should be guiding our own educational practices.

EVIDENCE-BASED PRACTICE IN OTHER DISCIPLINES: SOME EXAMPLES

Statisticians are often essential contributors in research teams in many disciplines and examples drawn from these contexts can enrich and facilitate the teaching of statistics. Interaction between statistics educators, statisticians and researchers in a relevant specialization can contribute significantly to the rich, real contextual and data resources that are of such value in both motivating and assisting statistical learning. Trends in medicine and other health sciences are governed by data, and evidence-based medicine is taught now in all medical schools. Data from the biological sciences provide information for resolving problems on environmental and ecological issues. The six-sigma revolution uses statistical quality control methods to monitor and improve industrial and engineering processes resulting in evidence-based decision making in industry. National statistics offices and international agencies contribute to evidence-based decision making in government and on public policy by collecting, collating, analysing and presenting data to populations at large and to governments in particular.

EVIDENCE-BASED PRACTICE IN STATISTICS EDUCATION

Evidence-based practice should also be employed in statistics education itself. How do we use context when teaching about variability, probability, inference and modelling? How do we interpret data from surveys, questionnaires or interviews and how are these related to the research hypotheses? To what extent are conclusions valid and reliable? Are we dealing with and explaining risk appropriately? Only with the answers to these intriguing questions will we be able to make informed decisions as we strive to reach an evidence-based society. Education ideas are shared on the web, through international and national projects, programmes, workshops and conferences in statistics education where diverse innovations are shared. The impact and relevance of new ideas are assessed and often adopted by others in their own teaching. Reports on the successes of recent statistics education programmes in South Africa and Latin America reflect the impact of the two ICOTS conferences and provide helpful ideas for other countries.

THE INTERNATIONAL PROGRAMME COMMITTEE EXECUTIVE

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

LOCAL ORGANISING COMMITTEE

LOC Chair: Andrej Blejec

For more information visit the ICOTS-8 website: http://ICOTS8.org

OTHER FORTHCOMING CONFERENCES

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

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

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

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

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

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

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

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

The Mathematics Education into the 21st Century Project was founded in 1986 and is dedicated to the planning, writing and disseminating of innovative ideas and materials in Mathematics and Statistics Education. The next conference is planned for September 7 - 13, 2007 in Charlotte, North Carolina. The chairman of the Local Organising Committee is Dr. David K. Pugalee, of the University of North Carolina Charlotte. The title of the conference is "Mathematics Education in a Global Community." Papers are invited on all innovative aspects of mathematics education. Our conferences are renowned for their friendly and productive working atmosphere. They are attended by innovative teachers and mathematics educators from all over the world, 25 countries were represented at our last conference for example.

For more information:
Alan Rogerson, arogerson@inetia.pl
Website: http://math.unipa.it/~grim/21project.htm