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

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

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

ABSTRACT

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

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

1. THE IMPORTANCE OF UNDERSTANDING COVARIATION

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

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

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

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

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

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

2. REVIEW OF THE LITERATURE

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

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

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

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

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

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

2.2. COVARIATIONAL REASONING AND JUDGMENTS IN STATISTICS EDUCATION RESEARCH

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

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

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

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

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

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

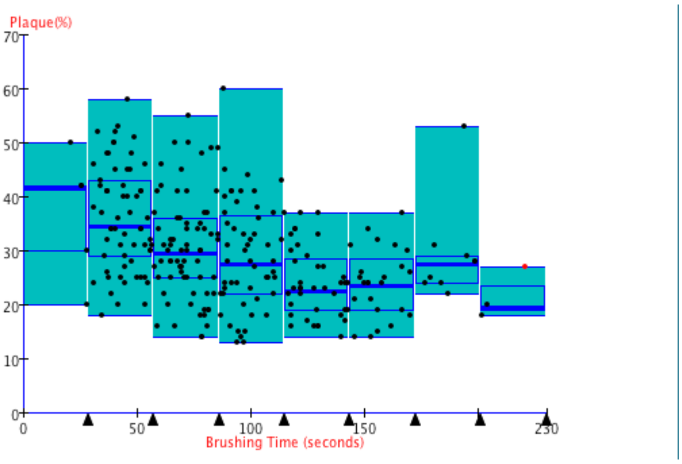


Figure 1. Minitools software allows students to start looking at the local variation for different values on the x-axis in addition to the global trend

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

2.3. UNIVARIATE DISTRIBUTION AS THE FOUNDATION FOR COVARIATIONAL REASONING

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

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

2.4. SUMMARY OF THE RESEARCH

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

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

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

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

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

3. METHODOLOGY

3.1. OVERVIEW OF STUDY

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

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

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

Table 1. The two sequences taught fall semester 2005

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

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

3.2. SETTING

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

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

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

3.3. INSTRUMENTS

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

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

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

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

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

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

3.4. DATA ANALYSIS

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

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

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

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

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

4. RESULTS

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

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

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

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

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

Note. The DR had a possible range of 0 to 10, with higher numbers indicating a higher perceived degree of reasoning. The BR had a possible range of 0 to 8, with higher numbers indicating a higher perceived degree of reasoning.
[a]Sequence 1 taught bivariate data early and inference later. Sequence 2 taught inference early and bivariate data later (see Table 1).

4.1. RESULTS OF FITTING THE UNCONDITIONAL MODEL

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

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

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

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

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

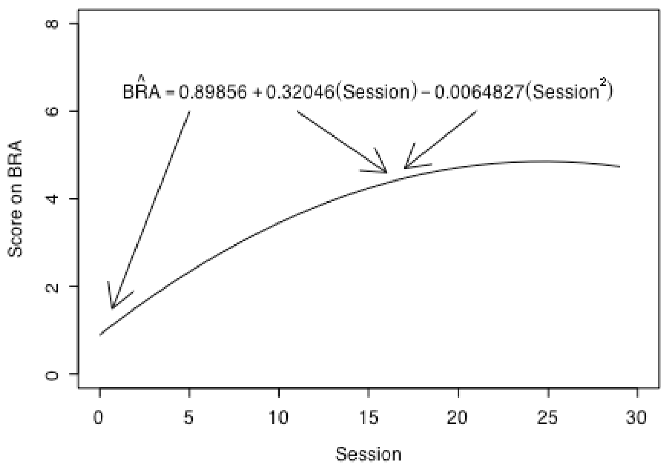


Figure 2. Predicted change in quantitative bivariate reasoning for an average student

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

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

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

4.2. RESULTS OF FITTING THE FIRST CONDITIONAL MODEL

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

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

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

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

4.3. RESULTS OF FITTING THE SECOND CONDITIONAL MODEL

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

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

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

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

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

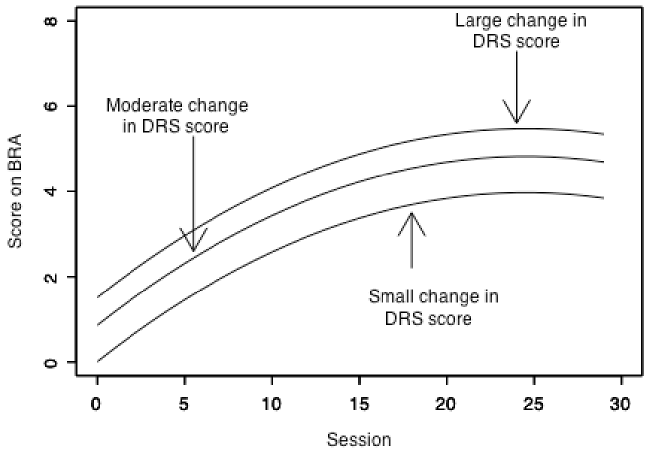


Figure 3. Predicted average change in quantitative bivariate reasoning for students with small, moderate, and large changes in their reasoning about distribution

5. DISCUSSION

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

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

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

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

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

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

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

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

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

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

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

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

5.4. LIMITATIONS TO THE STUDY

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

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

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

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

5.5. IMPLICATIONS FOR TEACHING

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

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

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

5.6. FUTURE RESEARCH

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

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

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

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

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

ACKNOWLEDGEMENTS

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX

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

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

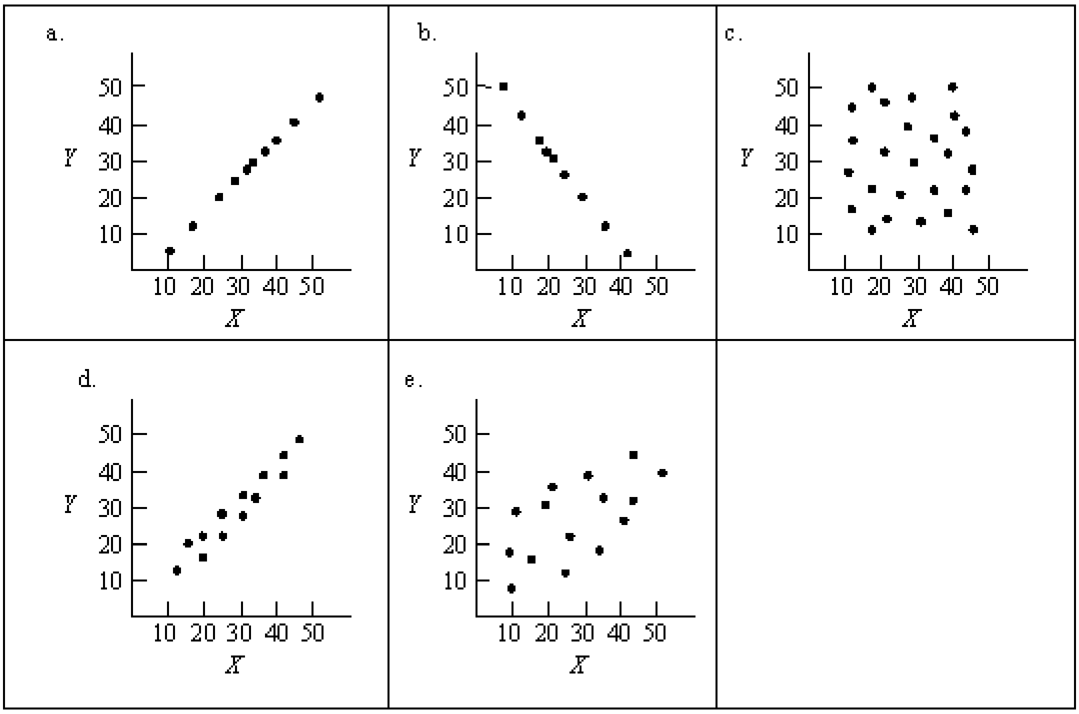
2. A student was studying the relationship between how much money students spend on food and on entertainment per week. Based on a sample size of 270, he calculated a correlation coefficient (r) of 0.013 for these two variables. Which of the following is an appropriate interpretation?
   a. This low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

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

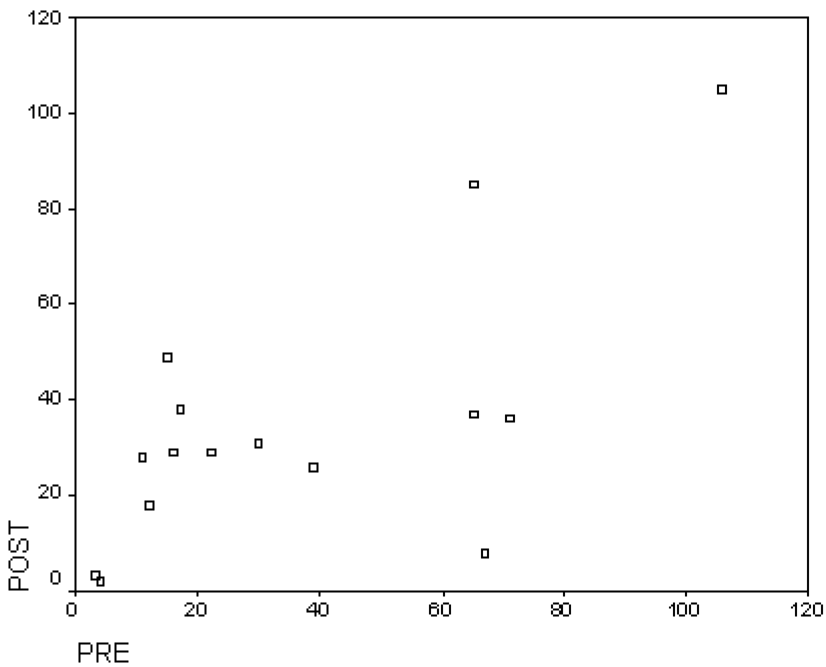
For items 4 and 5, select the scatterplot that shows:

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

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



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



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

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

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