



Statistical Reasoning Ability, Self-Efficacy, and Value Beliefs in a Reform Based University Statistics Course

**Olani, A., Hoekstra, R., Harskamp, E., &
van der Werf, G.**

Institute for Educational Research, University of Groningen,
Groningen

The Netherlands

Correspondence: Aboma Olani. Institute for Educational Research, GION, Grote Rozenstraat 3, 9712TG. The Netherlands E-mail: a.o.gicho@rug.nl

© Education & Psychology I+D+i and Editorial EOS (Spain)

Abstract

Introduction. The study investigated the degree to which students' statistical reasoning abilities, statistics self-efficacy, and perceived value of statistics improved during a reform based introductory statistics course. The study also examined whether the changes in these learning outcomes differed with respect to the students' mathematical background and perceived teacher support.

Method. Ninety-six first-year university students enrolled in an introductory statistics course were assessed both at the beginning and at the end of the course.

Results. The results showed that the students' statistical reasoning abilities and statistics self-efficacy significantly increased during the course. However, no significant changes were observed in their perceived value of statistics. The improvements in the students' statistical reasoning abilities were independent of their mathematical background or perceived teacher support. Larger positive changes in statistics self-efficacy were observed for students with favorable perceived teacher support.

Conclusions. This study concludes that students can attain important content related course goals in reformed statistics course regardless of their mathematics background. The attainment of other learning goals, particularly that of self- and value beliefs about statistics; however, are susceptible to teachers' support and encouragement.

Keywords: cognitive learning outcome; self-belief about statistics; prior mathematical skill; perceived teacher support.

Reception: 12/08/10 Initial acceptance: 12/15/10 Final acceptance: 17/03/11

Capacidad de razonamiento estadístico, auto-eficacia y creencias de valores en un curso universitario de estadística

Resumen

Introducción. El trabajo investiga la mejora de la capacidad de razonamiento estadístico, la auto-eficacia estadística y la percepción del valor de la estadística de los alumnos durante un curso introductorio de estadística. También se estudia si estos cambios en los resultados de aprendizaje difieren según el trasfondo matemático del alumno o la percepción de apoyo del profesor.

Método. Se evaluaron 96 estudiantes universitarios del primer curso, al principio y al final de una asignatura introductoria de estadística.

Resultados. Los resultados demuestran que la capacidad de razonamiento estadístico y la auto-eficacia estadística de los alumnos mejoró de forma significativa durante el curso. Sin embargo, no se observaron cambios significativos en su percepción del valor de la estadística. Las mejoras en la capacidad de razonamiento estadístico de los alumnos son independientes de su trasfondo matemático o su percepción de apoyo del profesor. Se observaron cambios positivos más grandes en la auto-eficacia estadística de aquellos alumnos que tenían percepción favorable del apoyo del profesor.

Conclusiones. Se concluye que los alumnos pueden lograr metas relacionadas con el contenido de la asignatura de estadística independientemente de su trasfondo matemático. El logro de otras metas de aprendizaje, concretamente las de creencias sobre uno mismo y sobre el valor de la estadística, son susceptibles al apoyo y aliento del profesor.

Palabras clave: resultado de aprendizaje cognitivo; auto-creencias sobre la estadística; capacidad matemática previa; percepción de apoyo del profesor.

Recibido: 08/12/10 Aceptación inicial: 15/12/10 Aceptación final: 17/03/11

Introduction

Over the past few decades, a widespread reform movement has emerged in statistics education at the tertiary level. The movement has given birth to the Guidelines for Assessment and Instruction in Statistics Education (GAISE College Report, 2005), which now serve as a blueprint for reform-oriented teaching of introductory statistics courses. The GAISE target the development of statistical reasoning abilities, thereby improving students' attitudes toward and self-beliefs about statistics. However, the question of whether these guidelines have indeed advanced these aims has received little attention. Most studies of the reform-based approach to teaching statistics (e.g., delMas & Lui, 2005; Zieffler, 2006) have merely examined changes in particular types of statistical reasoning ability (e.g., reasoning about sampling distribution, reasoning about standard deviation), resulting from instructions pertaining to one particular aspect of the GAISE (e.g., technology use, active learning methods). Other studies have focused on changes in students' attitudes and self-efficacy during statistics courses (Finney & Schraw, 2003; Schau, 2003). These issues have, to our knowledge, not been combined in a single study, particularly in an instructional setting in which principal aspects of the GAISE were implemented.

This study investigated changes in students' reasoning abilities, self-efficacy and value beliefs about statistics during an introductory statistics course that incorporated aspects of the GAISE in terms of goals, content, use of technology and assessment. The study also assessed whether the changes in these interdependent learning outcomes varied across students as a function of their mathematical background and their perceived teacher support. The results of this study may inform further research as well as classroom interventions aimed at guiding students in becoming competent individuals in the field of statistics.

Conceptual Framework

Over the years, the teaching of statistical skills or procedural knowledge has dominated the landscape of introductory statistics education. However, the traditional method of teaching theoretical statistics via mathematics rather than through applications has proven to be ineffective method for producing competent students in statistics. After completing a traditional statistics course, many students harbor misconceptions about basic statistical ideas, their reasoning skills in fundamental statistical concepts remain insufficient (Kahneman,

Slovic & Tversky, 1982), and their attitudes toward the course in particular or statistics in general remain negative (Schutz, Drogosz, White, & Distefano, 1998). This situation has given rise to the emergence of an important reform movement in statistics education.

The main aim of statistics education reform is to help students develop statistical reasoning abilities as well as positive self-beliefs about and attitudes toward statistics (Cobb, 1992; GAISE College Report, 2005; Moore, 1997). The literature provides several definitions of statistical reasoning ability. We used Garfield's (2003) definition: "...the way people reason with statistical ideas and make sense of statistical information" (p. 1). According to this definition, statistical reasoning refers to the ability to understand and integrate statistical concepts and ideas in order to interpret data and make decisions based on a given context. In addition, positive stimulation of students' self- and value beliefs about statistics has beneficial effects. These motivational factors influence the development of adequate statistical thinking during the teaching and learning process, the structural application of the knowledge obtained in real life situations, and a future interest in statistics (Gal & Ginsburg, 1994). This is because self- and task value beliefs play a central role in student's effort, persistence, and perseverance to study; and an individual who expects to do well in an activity and has high value for that task would be expected to choose that task over others and to display high effort and persistence for that task (Eccles & Wigfield, 1995; Lent, Brown & Hackett, 2000; Olani, 2009).

The instructional strategies pursued by the reform movement were primarily based on the theory of constructivism. The theory starts with the notion that people learn by actively constructing knowledge rather than by passively receiving knowledge provided by the teacher. The GAISE College Report has summarized the reform strategies as follows: (1) an emphasis on statistical reasoning and thinking, (2) the use of real data, (3) an emphasis on conceptual understanding rather than on mere knowledge of procedures, (4) an increased use of active learning methods, (5) the use of technology for clarifying abstract concepts and the automation of data analysis, and (6) the use of feedback-oriented assessments to evaluate and improve students' learning. The first three recommendations can be regarded as course aims and/or course content, whereas the latter three address the instructional pillars of pedagogy, technological use, and assessment, respectively. Additionally, GAISE has suggested teaching statistics in a manner in which the strong synergies among these strategies are promoted.

Review of Earlier Studies

Relatively few studies have examined changes in students' reasoning abilities in statistics during statistics instruction. Some have reported post-course improvements in the level of students' reasoning abilities (e.g., delMas, Garfield, Ooms & Chance, 2007; Derry, Levin, Osana, Jones & Peterson, 2000; Ragasa, 2008). Other research has shown no statistically significant changes in students' reasoning skills despite receiving good grades in the courses (e.g., Hirsch & O'Donnell, 2001). Many of these studies, however, concentrated on one particular aspect of the reform ideas. They measured, for example, the learning outcomes during or after the use of a specific active learning method or technology tool, such as computer-assisted instruction (e.g., Ragasa, 2008), simulation programs (e.g., delMas, Garfield & Chance, 1999), or cooperative learning (Magel, 1998). Similarly, the focus in these studies often was the measurement of one type of reasoning ability, for example, reasoning about sampling distributions (delMas et al., 1999), reasoning about standard deviations (delMas & Lui, 2005), or reasoning about bivariate data (Zieffler, 2006).

Studies that have explored the attainment of statistical reasoning abilities across a variety of topics in statistics courses are limited. delMas et al. (2007) examined the level of students' conceptual understanding with respect to general topics taught in a typical introductory statistics course rather than concentrating on a particular type of reasoning. A comparison of pre-test and post-test responses for a subset of participants in the study (763 students, taught by 22 instructors at 20 higher education institutions across the US) showed significant improvements in students' conceptual understanding. Although delMas et al. identified changes in students' abilities to reason across a breadth of topics in the introductory statistics course, the extent to which the teachers' instructional approaches deviated from the instructional and assessment reform guidelines was unclear. Furthermore, changes in affective learning outcomes were not assessed.

Only a few studies have reported positive pre-to-post-course changes in affective learning constructs (e.g., Harlow, Burkholder & Morrow, 2002; Schau, 2000). Other studies have reported either no changes (e.g., Green, 1993) or negative changes (e.g., Schau, 2003). As expected, research into courses that involve innovative instruction methods has shown better outcomes with respect to affective learning. Harlow et al., for instance, found signifi-

cant improvements in students' attitudes and self-efficacy accompanied by a decrease in anxiety. These results stemmed from a semester study of quantitative methods, during which a learning enhancement approach (peer monitoring, clear and not clear sessions, consult corners, and applied projects) was adopted. Finney and Schraw (2003) also observed an increase in statistics self-efficacy during an introductory statistics course. In contrast with many other studies, Finney and Schraw used task-specific self-efficacy measures (e.g., interpreting a standard deviation) rather than domain-specific self-efficacy measures (e.g., learning statistics). The closer correspondence between a task and self-efficacy assessment yields in better prediction of the task performance (Bandura, 1997; Pajares, 1996).

The studies of the changes in learning outcome during courses, in general, assume that instructional intervention improves a student's thinking, attitude, and self-confidence in statistics. However, attaining such learning outcomes is a complex endeavor due to the myriad factors that influence the process. Research on learning in general has consistently demonstrated that students are more likely to develop a deep understanding of topics, self-confidence, and positive attitudes when they feel they are supported (Fraser, 1994). Specifically, teacher support is one of the most important instructional elements that buffers against student anxiety and helps students become actively involved in course-related activities (Pintrich & DeGroot, 1990). The importance of support is paramount in stressor subjects such as statistics, in which students, particularly those studying social sciences, frequently harbor feelings of insecurity about their capabilities and the subject's usefulness in their professional and daily lives. Furthermore, prior mathematical skills impact both success in statistics and motivational variables (Schutz et al., 1998). Therefore, the attainment of course outcomes can vary depending on students' classroom behavior and background.

The literature provides considerable information about the acquisition of particular types of learning during statistics instruction that incorporate individual aspects of the reform recommendations. The GAISE, however, urge the development of both students' cognitive and affective behaviors through an integrated approach to teaching of statistics. Such an approach needs to incorporate at least the principal reform recommendations concerning goals, pedagogy, content, technology use, and assessment of learning, so that these elements can reinforce one another in a balanced manner to achieve the desired course objectives. At present, however, empirical studies on this topic are lacking.

The Present Study

Two research questions are central to this study: (1) Do students' statistical reasoning abilities, their statistics self-efficacy, and their perceived value of statistics increase during a reform-oriented teaching of introductory statistics course? (2) Do the outcomes differ with respect to students' mathematical background and their perceived teacher support? Based on previous research, we hypothesized that students' statistical reasoning abilities, their statistics self-efficacy, and their perceived valuation of statistics would improve during the course. In addition, we expected that the changes in these learning outcomes would vary across the students' mathematical background and their perceived teacher support. A larger positive influence was expected among students with a relatively strong mathematical background and a more positive perception of teacher support.

Method

Participants

The study participants were first-year undergraduate students who attended an introductory descriptive statistics course at a large university. The students were from the department of Pedagogical Science of the Faculty of Behavioral and Social Science. This department attracts more female than male students. A total of 157 and 127 students were assessed, respectively, before and after the course. Data from both assessments were obtained from 96 students of whom 97% were female. The average age was 19 years with a standard deviation of 1.5 years. It appeared that there were no significant differences in the average grades at the end of the semester (ranging from 0–40) among the students who participated in this study ($\bar{x} = 25.44$, $SD = 5.45$) and those who did not ($\bar{x} = 26.44$, $SD = 5.32$; $t(183) = -1.26$, $p = .21$).

The Course

The course under study was called Statistics 1, an introductory course that was mainly focused on descriptive statistics. It was a prerequisite for Statistics 2, which teaches inferential statistics. An outline of the course in terms of goals, content, pedagogy, technological use, and assessment vis-à-vis GAISE is given below.

Goal

The main goal of the course was to develop students' statistical reasoning abilities across various topics in the course, thereby enhancing their self-confidence for accomplishing statistical tasks and their beliefs about the usefulness of statistics. Accordingly, upon completion of the course, students were mainly expected to:

1. Able to select an appropriate measure of a distribution for a given data set and interpret its meaning;
2. Understand and interpret correlation and regression concepts;
3. Understand procedures related to data production;
4. Understand the usefulness of probability for decision-making.

Content

The students studied the first four chapters of the textbook *Introduction to the practice of Statistics 6th edition*, by David S. Moore, George P. McCabe, and Bruce Craig (2009). These chapters covered the topics of data distributions, the relationships between variables, data production, and introductory probability in relation to the study of random processes. In addition, two real datasets were used to help students make sense of the statistical concepts covered in real life situation.

Pedagogy

The course involved lectures and small-group sessions. The students could attend 12 class lectures and 4 small-group sessions (there were no small-group session in the first and the last week of the course). A session was 90 minutes each in both cases. The lectures were on Mondays and Fridays, and the small group sessions were scheduled between the lectures, on Tuesdays or Wednesdays. Thus, new course material could be explained during the Monday lecture and could be practiced in the small group sessions. The students were regularly given exercises related to the subject matter that had been covered in each lecture. These exercises were discussed in the subsequent small group sessions.

Lectures were offered by a faculty member, whereas a PhD student led the small group sessions. The lectures focused on conceptual discussions rather than on procedures. For in-

stance, when the concept of variance was explained, considerably more attention was paid to the questions of *what is the variance* and *why is it important*, than to *how is the variance calculated*. Fostering active learning in a lecture for a large group of students is relatively hard. However, the instructor tried to create interactive classroom atmosphere by posing questions to be discussed during lectures. For this purpose, multiple-choice questions selected from earlier exams were used. The instructor usually discussed a concept and displayed a question related to the concept on a PowerPoint slide. Students could then vote for alternative answers by raising their hands. If the voting showed that a substantial group had chosen a certain incorrect alternative, the teacher re-explained the concept, focusing on possible misconceptions that might led them to the wrong answer. Finally, the right answer would be given and the lecture would move on to the next concept. Roughly 30 minutes per lecture were used for such activities.

Small-group sessions consisted of not more than 16 students, and the students could work individually or in pairs. The focus of these sessions was to help students enrich their conceptual understanding of topics discussed in the previous lecture. This was achieved by helping students work with concepts discussed or on assignments given in the previous lecture, using one or more real dataset. At the end of the sessions, the PhD student always gave a short summary of the most important topics covered.

Technological Support

The lectures were presented using PowerPoint presentations. The PowerPoint slides were published on a Blackboard system (called Nestor) prior to the lecture to give students the opportunity to study them beforehand. Via a digital discussion board, to which the teacher had access as well, the students could discuss problems encountered in the course. Although the instructor mentioned the availability of this board, it was only used incidentally. SPSS and StatPlay were introduced in the small group sessions. SPSS is a system for analyzing data sets on a very basic level, and StatPlay is an interactive program through which students can learn about key concepts, such as distributions, variability between sample outcomes, and confidence intervals around sample outcomes. The program's aim is to visualize these concepts and help overcome potential misconceptions with respect to the concepts (Cumming & Thomason, 1998).

Assessment

Aside from asking questions during the lectures and small group-sessions, students could ask questions via e-mail. On average, one question per student was asked via e-mail. If a question was asked more than once, it was presented on Nestor together with its answer. Furthermore, three previous examinations as well as their answers were published on Nestor. Students were encouraged to do exercises in the textbook for which answers were provided for the odd numbered questions. These assessments were used to help students get continuous feedback on their learning. Grades in the course were assigned based on scores on a final examination consisting of 40 multiple-choice items.

Procedure

Participants were surveyed for their perceived self-efficacy, value of statistics, and statistical reasoning abilities, both before and after the course. Their perceived teacher support was assessed only at the end of the course. This was because students are, in our opinion, only able to form a balanced perception of their teacher at the end of a course. The survey questionnaires were administered in the students' native language. All questionnaires were included in a single booklet in which self-report scales preceded the statistical reasoning ability test; and students were instructed to complete the booklet in this order.

All items in the self-report scales were presented together and expressed using a common scale for both measurements. This organization was intended to minimize response bias that can occur if the scales are separated. The items in the scales were expressed in terms of a 4-point Likert-type scale: strong agreement (4), agreement (3), disagreement (2), and strong disagreement (1). We opted to use a forced-choice response scale with an even number of responses and no middle, neutral, or undecided responses because we assumed that students had formed an opinion about the attributes measured at the time of each test. Exposure to statistics coursework in secondary school mathematics studies would have helped them form certain opinions of the field of statistics, even before having started the course. Participation in the course enabled them to develop a well-defined view of the subject matter.

The pre-test was conducted during the second session of the course, and the post-test was conducted during the last session. The first session was used to introduce the course and to

highlight aspects of the planned course of study. Both the pre- and post-tests were conducted within the first half of the sessions to permit students to complete the scales without rushing. Students took between 25 and 35 minutes total to complete the questionnaire. Finally, in order to motivate the students to participate in the study, five participants in the group who took part in both measurement trajectories could win a coupon of 20 Euros.

Measures

Statistics Self-efficacy

Statistics self-efficacy was assessed using a scale consisting of 10 items that measured the students' confidence in their abilities to perform specific statistical tasks. The scale was developed for the purpose of this study. The self-efficacy items were selected with the help of the course instructor, who prioritized a list of specific tasks that are typically encountered during a statistics course. The top-10 conceptual (requiring no mathematical skills) tasks rated as strongly linked to course objectives were selected for the final scale. Bandura's (2006) Guide for Constructing Self-Efficacy Scales was consulted during the scale construction process. Sample items were: "I can explain what happens to a given distribution when the value of its standard deviation is changed." and "I can understand if a distribution is skewed when given the values of the mean and the median." A Cronbach's alpha of .73 and .84 were found in the pre- and post-test, respectively. Although prior validity data were not available for this scale, a similarly constructed statistics self-efficacy scale focusing on both descriptive and inferential statistical tasks was found to relate significantly to statistics anxiety, math self-efficacy, prior knowledge in mathematics, and performance in statistics (Finney & Schraw, 2003; Jimenez, Varga & Mestre, 2008).

Perceived Value of Statistics

Students' perceived value of statistics was measured using items adapted from the *Value* subscale of the Survey of Attitude Toward Statistics (SATS) (Schau, Stevens, Dauphinee, & Del Vecchio, 1995). The scale consisted of nine items that focused on students' attitudes toward the relevance, usefulness, and worth of statistics in their professional and personal lives. Sample items were: "Statistical skills will make me more employable" and "Statistics is relevant in my life." Adequate reliability and validity evidence is documented for this

scale. In this study, an internal consistency alpha of .70 and .71 were found in the pre- and post-tests, respectively.

Perceived Teacher Support

Students' perceived teacher support was measured using a 10-item scale that focused on perceptions of the teacher's helping behavior. We constructed items that measured teacher helping behavior with an emphasis on the extent to which a teacher supported and encouraged students toward engaging them in classroom activities. Sample items were: "The teacher emphasizes that all students who work hard can succeed in this course." and "The teacher always encourages us to ask questions during lectures." The scale yielded a Cronbach's alpha of .71. Although no prior validity data were available for this scale, a scale based on similar conceptual framework for small class size/seminars (a subscale of the College and University Classroom Environment Inventory (CUCEI)) relates significantly to many cognitive and affective learning outcomes (Fraser, Treagust & Dennis, 1986).

Statistics Reasoning Ability

Statistical reasoning ability was measured using a test consisting of 15 multiple-choice items with four alternatives each. Rather than focusing on computational and procedural knowledge, all items in the test dealt with the conceptual understanding of and reasoning about commonly encountered topics in descriptive statistics. Eight items were adapted from the Statistical Concept Inventory (SCI) (Allen, 2006) and seven items from the Statistical Reasoning Assessment (SRA) (Garfield, 2003). The items were selected with the help of the course instructor based on the goals and the content of the course. The learning outcomes measured, for instance, included the ability to compare the variability of distributions presented in a histogram, the ability to make a prediction about the occurrence of an event based on probability data, and the ability to identify an appropriate measure of central tendency by considering outliers. Internal consistency among the 15 items on the pre- and post-test produced Kuder-Richardson 20 (KR-20) values of .51 and .55, respectively. A reliability index exceeding .49 is considered reliable for short (10–15 items) achievement tests according to some references (e.g., Kehoe, 1995).

Mathematics Background

Mathematical background refers to students' reported secondary school mathematics levels and scores. The participants had studied mathematics in secondary school at a variety of levels; A1 (33%), A2 (37%), B1 (27%), and B2 (3%). These levels were arranged from the lowest (A1) to the highest level (B2), where level "A" referred to basic mathematics and level 'B' to advanced mathematics (algebra, geometry, and calculus). In subsequent analysis, students who had attended B2 secondary school mathematics level courses were grouped together with those who had attended B1 mathematics level courses because only three students had attended the B2 level. The students' secondary school mathematics scores were represented by their grades in their respective mathematics course level. The grading scale ran from 10 (excellent) to 1 (very poor), with 5.5 as a common minimum passing mark. This can be roughly equated to A⁺ to F in the letter grading system.

Results

Changes in Reasoning, Self-efficacy, and Value Beliefs

A paired t-test was carried out to examine the extent to which students' statistical reasoning abilities, their statistics self-efficacy and their perceived value of statistics changed during the course. Table 1 presents the mean and standard deviations. The results showed the presence of a statistically significant mean increase in statistical reasoning ability and statistics self-efficacy. This increase was 1.61 questions (95% CI: 1.15 to 2.10; $\eta^2 = .39$) for statistical reasoning ability and .17 (95% CI: .08 to .25; $\eta^2 = .13$) for statistics self-efficacy. On the other hand, no significant change was detected in the students' perceived value of statistics (mean difference = $-.04$, 95% CI: $-.10$ to $.01$; $\eta^2 = .03$).

Table 1. *T-tests for Differences in Variables Before and After the Course*

Learning Outcomes	Item No	Pretest		Post-test		Change		CI
		\bar{x}	SD	\bar{x}	SD	\bar{x}	SD	
Reasoning	15	9.07	2.19	10.68	2.31	1.61	2.38	[1.15, 2.01]
Efficacy	10	2.79	.35	2.96	.45	.17	.44	[.08, .25]
Value	9	2.85	.27	2.81	.31	-.04	.27	[-.10, .01]

Note. The scores for statistical reasoning ability could range from 0 to 15 correct questions. For statistics self-efficacy and perceived value of statistics, the average scores could range from 1 (strongly disagree) to 4 (strongly agree).
Change = post-test - pre-test. N = 96.

However, the mean increase, especially that of statistical reasoning ability, has to be interpreted with caution for two important reasons. First, a paired match t-test analysis of individual item results showed that only half 47 % (7 out of 15) of the items had a significant increase in the percentage of student with correct answer from the pre- to the post-course measurement. Second, destructor analysis of items in the reasoning test showed that considerable proportion of students had misconception/ misunderstandings of basic statistical concepts by the end of the course. Table 2 presents common misconceptions based on item-distracters towards which considerably large numbers of students were attracted irrespective of any visible ambiguity in the items and their alternatives.

Table 2. *Item Responses Showing Misconception/Misunderstanding*

Item	Misconception	% of students	
		Pretest	Post-test
3a	Fail to understand the effect of extreme values on mean	54	41
4c	Fail to apply ‘law of large number’ (not able to understand that sample size is related to the probability of occurrence of events)	43	40
9b	Fail to correctly apply probability laws for an independent events	40	31
11a	Fail to make correct prediction based on probability data	35	34
13c	Confuses mean and mode	23	26
2b	Confuse percentile and percentage	25	20

Changes across the Levels of Secondary School Mathematics

A one-way ANOVA was conducted to compare the pre-to-post course changes in the learning outcomes as a function of the participants' secondary school mathematics level (A1, A2, and B1 and B2). Table 3 presents the mean and standard deviations.

Table 3. *Mean and SD for Secondary School Mathematics Levels*

Secondary School Mathematics Level	N	Changes in Learning Outcomes					
		Reasoning		Efficacy		Value	
		\bar{x}	SD	\bar{x}	SD	\bar{x}	SD
A1	32	1.53	2.24	.20	.53	-.12	.32
A2	34	1.29	2.19	.13	.39	.00	.22
B1 & B2	30	2.10	2.41	.17	.24	-.02	.24

Note. Change = pre-test – post-test, A1 = lowest level secondary school mathematics, A2 = intermediate level secondary school mathematics, B1 & B2 = highest level secondary school mathematics. N = 96.

The results revealed no significant differences in the pre-to-post course changes in statistical reasoning abilities ($F(2, 93) = .98, p = .38, \eta^2 = .02$). This also applied to statistics self-efficacy ($F(2, 93) = .23, p = .79, \eta^2 = .004$) and perceived value of statistics ($F(2, 93) = 1.95, p = .15, \eta^2 = .04$). These findings suggested that the learning outcome did not vary as a function of the level at which students studied mathematics in secondary school.

Changes in Relation to Secondary School Mathematics Score and Perceived Teacher Support

A standard multiple regression was performed to explore whether or not the changes in learning outcome were linearly related to the students' secondary school mathematics scores or perceived teacher support. Table 4 shows the mean, the standard deviation, and Pearson correlation coefficients of the study variables.

Table 4. Means, Standard Deviations and Correlations among Study Variables

Variables	\bar{x}	SD	1	2	3	4	5
1. Change in Reasoning	1.61	2.38	-				
2. Change in Efficacy	.17	.44	.04	-			
3. Change in Value	-.04	.27	.10	.26*	-		
4. SSMS	6.62	1.03	-.08	.07	.07	-	
5. PTS	2.28	.29	.01	.25*	.22*	.12	-

Note. The scores for SSMS (secondary school mathematics scores) could range from 0 to 10. For PTS (perceived teacher support), the average scores could range from 1 (strongly disagree) to 4 (strongly agree).
Change = post-test - pre-test. N = 96. * $p < .05$.

Table 5. Summary of Multiple Regression Analyses of Changes in Learning Outcomes using Mathematics Scores and Perceived Support

Change in Learning Outcomes	Predictors	<i>b</i>	<i>SEb</i>	β	R	R ²	F ratio
Reasoning	Constant	2.36	2.74				
	PTS	.17	.84	.02	.08	.007	.31
	SSMS	-.18	.24	-.08			
Efficacy	Constant	-1.01	.50				
	PTS	.367	.15	.24*	.25	.06	3.14*
	SSMS	.02	.04	.04			
Value	Constant	-.69	.31				
	PTS	.20	.09	.21*	.22	.05	2.41
	SSMS	.01	.03	.05			

Note. Change = pre-test – post-test, SSMS = secondary school mathematics scores, PTS = perceived teacher support. N = 96, * $p < .05$.

The results showed that secondary school mathematics scores and perceived teacher support accounted for 0.7% ($R^2 = .007$, $F(2, 93) = .31$, $p = .73$) of the total variance in pre-to-post course changes in statistical reasoning ability. The factors also explained 6% ($R^2 = .06$, $F(2, 93) = 3.14$, $p = .04$) and 5% ($R^2 = .05$, $F(2, 93) = 2.41$, $p = .09$) of the total variance in pre-to post-course changes in statistics self-efficacy and perceived value of statistics, respectively. Perceived teacher support significantly predicted pre-to-post-course changes in self-efficacy ($\beta = .24$, $p = .02$), as did perceived value of statistics ($\beta = .21$, $p = .04$).

Further inspection of the partial correlation coefficients revealed that the perceived teacher support had the largest unique contribution to the variances in self-efficacy (5.8%, partial corr. = .242) and perceive value (4.4%, partial corr. = .210) models.

Discussion

This study showed that the ability to reason with statistical ideas and the efficacy for accomplishing such tasks increased after taking an introductory statistics course aligned with the GAISE recommendations. The observed increase in statistical reasoning ability during a reform-oriented statistics course, as well as the improvement in statistics self-efficacy, was in line with the results of other studies (e.g., delMas et al., 2007; Harlow et al., 2002). However, no clear changes were observed in the perceived value of competence in statistics. Although this result contrasted with our expectation, some previous studies have reported similar results. Schau (2000) reported no significant change in attitudes toward usefulness of statistics in a study in which positive changes were observed in other dimensions of attitudes. It might be the case that a single course in statistics is insufficient to impact students' perceived value beliefs in statistics. In addition, unlike statistical reasoning and self-efficacy measures that were formulated such that they correspond to specific course objectives, the perceived value of statistics was measured using a scale that tended to be relatively general. The items in the *value* scale measured students' perceptions about the usefulness of the general field of statistics (e.g., "statistics is relevant in my life") rather than their perceptions about the usefulness of the statistical skills they acquired during the course. Thus, this might be because this measure could not capture the real impact of the course on students' value beliefs at a specific task level.

The changes in the learning outcomes occurred regardless of students' mathematics background. This implied that a mathematical background may not always determine the learning outcome in a statistics course, particularly when no specific basic prerequisites are necessary, as was the case for the course examined in this study. Because the students came from different streams of secondary school programs that taught at a variety of mathematics levels, the course was designed for a diverse audience, and no specific prerequisites were set. Additionally, the course in this study focused on statistical reasoning ability rather than on statistical computation. The former demands less mathematical skills than the latter. This

might give mathematical skills a less important role in the attainment of learning outcomes pursued by an education in statistics.

Furthermore, students who perceived their teacher as supportive became more self-efficacious and appreciative of statistics than those who experienced their teacher as less supportive. This finding was in line with those of previous studies (e.g., Pintrich & DeGroot, 1990). This finding implied that statistics teachers' support and encouragement can be used as an important input to lay the motivational foundations for cultivating students' self-beliefs about the statistics course, which is usually perceived as a very difficult and unlikeable subject by many students. However, it looks that perceived support tend to relate more to students self-and task value beliefs than their actual performance. This is more likely because support is in a better position to stimulate beliefs than actual performance in the statistical reasoning tasks. These tasks are higher order cognitive tasks which are relatively complex and stable than the affective behaviors, which have a fluctuating nature depending on changing circumstances and as a result are more susceptible to environmental percepts.

There are some limitations that must be acknowledged. First, the lack of a control group in this study made it impossible to draw definitive causal conclusions. The overall design of this study was mainly correlation focusing on the changes (and variation) in the three learning outcomes over a course described vis-à-vis the GAISE. Second, the participants in this study were predominantly females, and generalizing the results to males may be problematic. This is because females tend to differ in their statistical reasoning ability and self-beliefs about statistics from their male counterparts (e.g., Tempelaar, Gijsselaers, & Schim, 2006). In future research, the present study suggests the use of an experimental design in which a proportional number of male and female participants are present. In addition, it would be useful to focus on the interplay between statistical reasoning ability and affective course outcomes in relation to course grades. In this study, scores on the statistical reasoning ability test were not graded. This might have induced students not to perform their best on this test. Finally, we measured statistical reasoning ability and statistics self-efficacy using indicators associated with specific statistical tasks. However, we did not use such indicators when measuring students' perceived value of statistics. It is worth pointing out the appropriateness of using specific items when measuring effects of instruction on components of affective behavior in statistics education.

In summary, students can attain important course goals in statistics regardless of their mathematics background. The attainment of such learning goals, particularly that of self- and value beliefs about statistics; however, are susceptible to teachers' support and encouragement. Therefore, in addition to instructional upgrades, teacher support and encouragement can play a vital role in a better attainment of course goals by buffering some of the negative feelings associated with statistics and offsetting the risk of low engagement in the course.

References

- Allen, K. (2006). *The statistics concept inventory: development and analysis of a cognitive assessment instrument in statistics*. Unpublished doctoral dissertation, University of Oklahoma, Norman.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Adolescence and education: Vol. 5. Self-efficacy and adolescence* (pp. 307-337). Greenwich, CT: Information Age Publishing.
- Cobb, G. (1992). Teaching statistics. In L. Steen (ed.), *Heeding the call for change: Suggestions for curricular action*, (pp. 3-23). Washington, D.C.: Mathematical Association of America.
- Cumming, G., & Thomason, N. (1998). StatPlay: Multimedia for statistical understanding, In L. Pereira-Mendoza, L. Kea, T. Kee, and W-K. Wong (Eds.), *Statistical education: Expanding the network. Proceedings of the Fifth International Conference on Teaching of Statistics* (pp. 947-952). Voorburg, The Netherlands: International Statistical Institute.
- delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education* 7(3). Retrieved from <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm>.
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6 (2), 22-38.
- delMas, R. C., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55-82.
- Derry, S. J., Levin, J. R., Osana, H. P., Jones, M. S., & Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal*, 37(3), 747-773.
- Eccles, J., & Wigfield, A. (1995). In the Mind of the Actor: The Structure of Adolescents' Achievement Task Values and Expectancy-Related Beliefs. *Personality and Social Psychology Bulletin*, 21(3), 215-225.

- Fraser, B. J. (1994). Research on classroom and school climate. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 493–541). New York: Macmillan.
- Fraser, B.J., Treagust, D.F., & Dennis, N.C. (1986). Development of an instrument for assessing classroom psychosocial environment in universities and colleges. *Studies in Higher Education, 11*(1), 43-54.
- Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161-186.
- GAISE (2005). Guidelines for assessment and instruction in statistics education (GAISE) college report. The American Statistical Association (ASA). Retrieved from <http://www.amstat.org/education/gaise/GAISECollege.htm>.
- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education, 2*(2). Retrieved from <http://www.amstat.org/publications/jse/v2n2/gal.html>.
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 22-38.
- Green, K. E. (1993, April). Affective, evaluative, and behavioral components of attitudes toward statistics. Paper presented at the meeting of the American Educational Research Association, Atlanta.
- Harlow, L. L., Burkholder, G. J., & Morrow, J. A. (2002). Evaluating attitudes, skill, and performance in a learning-enhanced quantitative methods course: A structural modeling approach. *Structural Equation Modeling, 9*, 413-430.
- Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education, 9*(2). Retrieved from <http://www.amstat.org/publications/jse/v9n2/hirsch.html>.
- Jimenez, J. M., Vargas, M. V., Mestre, A. B. (2008). Measuring attitude toward statistics: the influence of study process. *Electronic Journal of Research in Educational Psychology, 6* (3), 729-748.
- Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation, 4*(10). Retrieved from <http://PAREonline.net/getvn.asp?v=4&n=10>.

- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Lent, R. W., Brown, S. D., & Hackett, G. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology*, 47(1), 36-49.
- Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education*, 6(3). Retrieved from <http://www.amstat.org/publications/jse/v6n3/magel.html>.
- Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistics Review*, 65, 123-165.
- Olani, A. (2009). Predicting first year university students' academic success. *Electronic Journal of Research in Educational Psychology*, 7 (3), 1053-1072.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40.
- Ragasa, C. (2008). A comparison of computer- assisted instruction and the traditional method of teaching basic statistics. *Journal of Statistics Education*, 16(1). Retrieved from <http://www.amstat.org/publications/jse/v16n1/ragasa.pdf>.
- Schau, C. (2000, November). Teaching and learning in statistics courses: Importance of students' attitudes and beliefs. Paper presented to the faculty at Arizona State University.
- Schau, C. (2003, August). Students' attitude: The "other" important outcome in statistics education. Paper presented at a joint meeting, San Francisco.
- 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, 868-875.
- Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences*, 10, 291-308.

Tempelaar, D. T., Gijssels, W. H., & Schim van der Loeff, S. (2006). Puzzles in statistical reasoning. *Journal of Statistics Education*, 10(1). Retrieved from <http://www.amstat.org/publications/jse/v14n1/tempelaar.html>.

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, Minnesota.