AN EXPLORATION OF STUDENT ATTITUDES AND SATISFACTION IN A GAISE-INFLUENCED INTRODUCTORY STATISTICS COURSE

WARREN PAUL
Department of Ecology, Environment and Evolution, La Trobe University, Wodonga, Victoria 3689, Australia
w.paul@latrobe.edu.au

R. CLARE CUNNINGTON
Academic Language and Learning, La Trobe University, Wodonga, Victoria 3689, Australia

ABSTRACT

We used the Survey of Attitudes Toward Statistics to (1) evaluate using pre-semester data the Students' Attitudes Toward Statistics Model (SATS-M), and (2) test the effect on attitudes of an introductory statistics course redesigned according to the Guidelines for Assessment and Instruction in Statistics Education (GAISE) by examining the change in attitudes over the semester and, using supplementary data from an annual Student Feedback Survey, testing for a change in overall satisfaction following implementation of the redesigned course. We took an exploratory rather than confirmatory approach in both parts of this study using Bayesian networks and structural equation modelling. These results were triangulated with analysis of focus group discussions and the annual Student Feedback Survey.

Keywords: *Statistics education research, Survey of Attitudes Toward Statistics (SATS), Students' Attitudes Toward Statistics Model (SATS-M), Student feedback survey, Bayesian networks and structural equation modelling*

1. INTRODUCTION

Teaching statistics to non-statistics majors is challenging, and it would be beneficial for students and teaching staff alike if the experience was more enjoyable. But how can we make statistics more interesting and enjoyable for students that have an aversion to numbers, let alone maths? It is widely recognised that most of our students will become consumers rather than producers of statistics (Gal, 2002; Rumsey, 2002), and that we should focus on learning outcomes that fit with the needs of the majority of students, while maintaining interest and pathways for other students that want a broader and deeper understanding of statistics (Ramirez, Schau, & Emmioğlu, 2012). The general thrust of research in statistics education and the subsequent reform movement has been toward demonstrating the value of statistics in the everyday and professional lives of students, by shifting the emphasis from manual calculations and procedural matters to activities that develop statistical literacy and thinking (Cobb, 1992). This reform culminated in the Guidelines for Assessment and Instruction in Statistics Education (GAISE; Aliaga et al., 2005). It seems reasonable to expect that the implementation of these guidelines will see a general improvement in students' attitudes toward statistics, but there needs to be ongoing

Statistics Education Research Journal, 16(2), 487-510, http://iase-web.org/Publications.php?p=SERJ

evaluation to establish what, if any, aspects of this reform improve the experience and outcomes for students.

The GAISE college report (Aliaga et al., 2005) made six recommendations for transforming statistics education in universities: 1) emphasize statistical literacy and develop statistical thinking; 2) use real data; 3) stress conceptual understanding, rather than mere knowledge of procedures; 4) foster active learning in the classroom; 5) use technology for developing conceptual understanding and analysing data; and 6) use assessments to improve and evaluate student learning. Olani, Hoekstra, Harskamp, and Van Der Werf (2011) noted that the first three recommendations relate to course aims/content, whereas the latter three address pedagogy, technology, and assessment. In addition to these, Gal, Ginsburg, and Schau (1997) contend that teachers should also consider students' attitudes toward statistics in the design of courses and evaluation of student learning. They give three reasons for considering and monitoring attitudes in statistics courses: 1) their role in influencing the teaching/learning process (process considerations); 2) their role in influencing students' statistical behaviour after they leave the classroom (outcome considerations); and 3) their role in influencing whether or not students will choose to enrol in a statistics course later on, beyond their first encounter with statistics (access considerations).

One of the more commonly used instruments for measuring students' attitudes toward statistics is the Survey of Attitudes Toward Statistics (SATS; Schau, Stevens, Dauphinee, & Del Vecchio, 1995). The SATS-36 contains 36 questions grouped according to six attitude components—affect, value, cognitive competence, difficulty, interest, and effort—as well as items that assess student characteristics such as gender and age, and students' previous achievement-related experiences (see Table 3 for the complete list of questions used in the SATS-36). The development of SATS was influenced by results from earlier attitudinal surveys and Eccle's expectancy-value theory (EVT). In the context of mathematics attitudes (Eccles & Wigfield, 1995; Wigfield & Eccles, 2000) EVT posits that students tend to value an activity when they think they are good at it, and they are less likely to believe they are good at something if they believe it is difficult.

Schau and her colleagues (Ramirez et al., 2012; Schau, 2003) also proposed the Students' Attitudes Toward Statistics Model (SATS-M) which suggests that students' characteristics (e.g., gender and age) influence previous achievement-related experiences and that both impact attitudes toward statistics; all three then influence course outcomes. Ramirez et al. note that empirical research generally supports the SATS-M, more so for the effects of prior achievement-related experiences on statistics attitudes and attitudes on performance outcomes than for the effects of student characteristics. They leave the intra-attitudinal relationships largely unspecified in the SATS-M because of the variety of causal relationships that would be consistent with EVT. Nonetheless, two studies (Hood, Creed, & Neumann, 2012; Sorge & Schau, 2002) have tested causal models that were based on EVT and that used measures of achievement-related experiences, performance outcomes, and attitudes toward statistics, including four components of the SATS-M (affect, cognitive competence, difficulty, and value). Both models indicate that past performance directly affects cognitive competence and achievement, and that difficulty directly affects cognitive competence, but other than that the models are quite different. For example, the model by Hood et al. predicts that value will indirectly affect achievement, through its effect on expectancies, but the model by Sorge and Schau has no direct or indirect causal path between value and achievement. We may one day be able to design interventions that target particular attitudes (e.g., value), but a greater understanding of how these attitudes are inter-related is needed before we can predict with any certainty the holistic consequences of such interventions.

Despite efforts to understand the causes and consequences of students' attitudes toward statistics, the various interventions to improve attitudes that have been attempted have had disappointing results, with none of the activities resulting in evidence of an improvement in mean attitudes after participation (e.g., Brandsma, 2000; Carnell, 2008; High, 1998; Rhoads & Hubele, 2000). There are a number of possible explanations for these results. Olani et al. (2011) contend that isolated activities may have no effect on attitudes, and a combination of reforms, as recommended in the GAISE report (Aliaga et al., 2005), may be required. As alluded to earlier, these reforms would need to be implemented within a safe and supportive environment (i.e., the process considerations mentioned by Gal et al., 1997). In addition, Gal and Ginsburg (1994) note that attitudes can be expected to be stable over time, which suggests that any pre-post changes that do occur will be small. An explanation for the stability of students' attitudes is offered in the literature on consumer satisfaction, including university student satisfaction, where expectancy disconfirmation theory predicts that disconfirmation of prior attitudes influences satisfaction decisions, and in turn those decisions influence post-attitudes (Athiyaman, 1997, 2004; Oliver, 1980). Basically, if a student perceives the quality of a course to be less than expectations, then negative disconfirmation occurs and attitudes are revised unfavourably; if perceived quality matches expectations, then confirmation arises and there is no change in attitudes; and if perceived quality exceeds expectations, then positive disconfirmation occurs and attitudes are revised favourably. However, Athiyaman (2004) notes that assimilation-contrast theory posits that a quality perception differing only slightly from one's expectation tends to result in movement of perceptions toward expectations (assimilation) whereas a large difference from expectation tends to be magnified (contrast). This suggests that any disconfirmation would need to be appreciable before having a discernible effect on satisfaction and post-attitude.

It is worth noting that most pre-post studies that have been conducted only asked whether there was evidence of a change in mean attitudinal scores in the population, and sought to answer this question via univariate paired-samples *t*-tests, for example. However, there is reason to believe that attitudinal changes may be inter-correlated (Olani et al., 2011), and this correlation structure together with focus group discussions (Gal et al., 1997) could provide insight into why attitudes improve for some students but not others.

In the present study we examine the causes and consequences of students' attitudes toward statistics, with particular attention paid to the intra-attitudinal relationships, and we ask whether implementing the GAISE recommendations in a statistics course can make a difference to attitudes and satisfaction. We used the SATS-36 to measure attitudes toward statistics of students taking an introductory statistics course that was redesigned according to the GAISE recommendations. A brief description of the course before and after the redesign, including the GAISE-related features incorporated in the new course, is provided in the Methods, with further details about the redesigned course provided in the Appendix. In accord with the SATS-M, we hypothesised that students' prior mathematics achievement-related experiences influence attitudes, and that attitudes in turn influence outcomes. However, given the ambiguity concerning intra-attitudinal relationships, we took an exploratory rather than confirmatory approach by using algorithms for learning Bayesian networks to learn the causal structure from the pre-semester data, with a follow up exploratory analysis using classical structural equation modelling (SEM) to further assess and diagnose the fit of the causal model. We also hypothesised that the redesigned course would have a positive effect on attitudes. This was tested at the population level using repeated measures (permutational) MANOVA. Unlike other pre-post studies of attitudes, we also explored correlations among pre-post attitudinal changes of individual students using Bayesian networks, and interpreted the results using feedback from focus groups that was obtained during the semester. Lastly, we hypothesised that student satisfaction would improve along with attitudes, and we used supplementary data from the annual Student Feedback Survey, which is undertaken as part of the University's quality assurance program, to examine the change in overall satisfaction between the four years before and four years after the implementation of the redesigned course.

2. METHODS

2.1. DESCRIPTION OF THE INTRODUCTORY STATISTICS COURSE

Prior to the redesign, the introductory statistics course had a "traditional" format. The course began with an explanation of what statistics is about, followed by lessons in descriptive statistics, with detailed procedures for calculating sample statistics and producing graphs (e.g., histograms), and it ended with procedures for calculating test statistics, *p*-values, and confidence intervals. In the middle of semester, four weeks were devoted to probability theory and details about continuous and discrete probability distributions. Anecdotal evidence, and a marked decrease in attendance at lectures, suggests it was here that any gains made in explaining and demonstrating the value of statistics were lost for most students. In addition to the 2-3 hours of lectures each week, students undertook a 1-hour tutorial that was spent on solving textbook problems, and 1-hour computer laboratory which involved basic instruction in SPSS. According to responses to open questions on the annual Student Feedback Survey, students generally found the work difficult, "dry," and to be lacking relevance.

The course was redesigned to help students understand why we need data in our daily and professional lives, how data are produced, and how they are transformed using statistics into information that empowers us to make decisions and solve problems. It was our intention to help students appreciate how numbers reported in the media shape debates about social, economic, and environmental problems and policies (e.g., see Best, 2005, 2008). Importantly, we wanted students to recognise that numbers reported in the media can often be misleading, and how they, as consumers of statistics, can use their knowledge of statistics to critically evaluate the information they receive through the media. The emphasis in the redesigned course is on key concepts, rather than mathematical calculations and procedural issues, and learning statistics by doing it with the SPSS statistical software package and real data from their discipline area (i.e., science, psychology, or business).

All six GAISE recommendations were designed into the course through the constructive alignment of intended learning outcomes (ILOs), learning activities, and assessments (see Biggs 1996). The ILOs for the redesigned course are:

1. Interpret and convey statistical information using the terminology and important concepts of statistical design and analysis.
2. Apply the appropriate descriptive tools of statistics to summarise data and reveal important features that will help to answer questions.
3. Apply the appropriate inferential tools of statistics to weigh up evidence and draw conclusions.
4. Critically evaluate the statistics presented in the media in order to assess the value of the information.

The first ILO addresses the need to emphasize statistical literacy and statistical thinking, and stress conceptual understanding, while the remaining ILOs are intended to develop students' statistical thinking using news and graphs in the media, real data, and technology. Further details on the content, activities, and assessments of the redesigned course are given in the Appendix.

The transition from the old to the new course required a greater emphasis on plain-language or diagrammatic (as opposed to mathematical) explanations of the *why*, not just the *how*, statistical design and analysis is done, as well as a much reduced content to avoid cognitive overload impeding students' learning of the concepts. Importantly, to accommodate the shift in emphasis toward statistical thinking and concepts, the fourth ILO (critical evaluation of statistics reported in the media) was introduced, and the learning outcome and content associated with formal probability and probability distributions were removed. In agreement with Moore (1997), it is our opinion and experience that only an informal grasp of probability is needed to understand the reasoning behind statistical inference. A summary of the main differences between the old and the new course is given in Table 1.

Table 1. Summary of the main differences between the old and the new course

Component	Changes made
ILOs	Removed the ILO of "Apply the laws of probability to find the probability of an event from a given probability distribution," and added the ILO of "Critically evaluate the statistics presented in the media in order to assess the value of the information."
Content	Removed the material on probability laws, calculating the probability of an event, the Bernoulli and binomial probability distributions, paired-samples designs and tests, tests of directional hypotheses (preferring instead to use two-tailed tests supplemented by confidence intervals), and manual calculations of tests statistics and confidence intervals. Added material on how to think critically about numbers reported in the media and additional material on how data are produced, with an emphasis on the concepts of selection bias, causality (see Schield, 1995) and confounding bias, and techniques to minimise bias.
Activities	Removed the 1-hour tutorial (each week) which involved manual calculations and calculations using a graphing calculator, extended the computer lab from 1-hour to 2-hours to spend more time analysing real data using SPSS, and added weekly online quizzes that address concepts. The 3-hour lecture was reduced to 1-2 hours (each week).
Assessments	Kept the two assignments (one on analysing real data using descriptive statistics with SPSS, and one on using inferential statistics) and the 2-hour open-book exam, but replaced the weekly tutorial questions with weekly online quizzes.
Contact hours	Reduced from 4-5 hours to 3-4 hours per week (for 12 weeks).

2.2. PARTICIPANTS

The University's Human Ethics Committee granted approval for this project. The study used a single-group pre-post design. Students taking Critical Thinking with Statistics at a small regional campus in Australia in 2012 and 2013 were invited by email to 1) complete an online survey of attitudes toward statistics at the beginning (within the first two weeks) and end (the last two weeks) of their course, and 2) participate in a focus group discussion held in the second last week of semester (after submitting their final assignment). There were 208 students enrolled in the course (105 in 2012, and 103 in 2013), with an approximately equal proportion of students from business, psychology, and life sciences degree programs. The majority of students have English as their main language. One-hundred and seven (51%) students participated in the pre-semester survey, and 27 (13%) also participated in the post-semester survey. Seventeen students participated in the focus group discussions. Of the students that participated in the pre-semester survey, 65% were female and 82% were aged between 18-25 years old. The final mark in the course for all students enrolled had a mean of 64% and median of 70%. For the students that participated in the pre-semester survey the average final mark was 66% and the median was 72%. The students that participated in the post-semester survey had a mean final score of 75% and median 78%. These statistics suggest that the students taking part in the pre-semester survey were representative of all students enrolled, while those that also participated in the post-semester survey may not be. With regard to the focus group, participants were anonymous and could not be traced to individual survey responses or final marks; however, as noted in the Results, these students appeared to exhibit the full range of attitudes and changes in attitudes observed in the pre- and post-semester surveys, which suggests their views and experiences were at least indicative of all students enrolled in the course. The responses from the surveys and focus groups from both years, 2012 and 2013, were combined for the data analysis.

2.3. MEASURES

The SATS-36 (Ramirez et al., 2012; Schau, 2003; Schau et al., 1995) contains 36 Likert-type items for assessing six attitude components: affect, value, cognitive competence, difficulty, interest, and effort. Students rate each item on a scale from 1 (strongly disagree) to 7 (strongly agree). The pre-semester and post-semester items are identical except for tense. A further eight items are included that assess student characteristics such as gender and age, and students' previous achievement-related experiences. Because higher item responses reflect more positive attitudes, responses to items that are negatively worded are reversed before combining students' responses into component scores. Thus, higher scores for every component (except for difficulty) also reflect more positive attitudes. Students with higher difficulty scores believe that statistics is easier whereas those with lower scores believe that it is harder. Several studies have shown that scores from the SATS have good to excellent psychometric properties (Ramirez et al., 2012). The SATS-36 can be obtained from Candace Schau at www.evaluationandstatistics.com.

2.4. DATA ANALYSIS

Pre-semester data Bayesian networks were used to learn the causal structure from the pre-semester data. A Bayesian network consists of a factorisation of a probability distribution over a set of ordered variables $\{X_1,...,X_p\}$, as in (1) where PA_j represents the parents of X_j,

(1) $$p\left(x_1,...,x_p\right)=\prod_j p\left(x_j \mid pa_j\right)$$

and a directed acyclic graph (DAG) corresponding to the factorisation (Pearl, 2000). The nodes in a DAG represent variables and the edges (arrows) between the nodes represent causal connections between the variables. The conditional independencies that may be derived from the factorisation correspond to the *d*-separation statements in the DAG. Bayesian networks can be used in a confirmatory way to test an a priori composite hypothesis by proposing a causal structure and testing the concomitant conditional independencies, or in an exploratory manner to learn the causal structure from data. Structure learning algorithms fall generally into three categories (Koski & Noble, 2012): *constraint based* methods, where the independence relationships established through a series of conditional independence tests are used to produce an equivalent DAG; *score based* algorithms that attempt to find the structure that maximises a score function; and *hybrid* algorithms that use both constraint based and score based techniques. All three methods are implemented in the Bayesian Network Structure Learning (bnlearn) package in R (R Development Core Team, 2013; Scutari, 2009), which was used to learn the causal structure from the pre-semester data.

Classical, or covariance-based, SEM was used to further assess and diagnose the fit of the Bayesian network. Classical SEM differs from Bayesian networks primarily in that a model is tested using a global comparison of the observed and model-implied covariance matrices, rather than on the basis of local conditional independence tests. Most SEM packages include a variety of indices for assessing goodness of fit and modification indices for diagnosing lack of fit, as well routines for estimating direct and indirect effects (Byrne, 2010; Kline, 2005). IBM SPSS Amos version 22 was used for this part of the analysis.

Comparison of pre and post-semester data Of the twenty-seven students that completed the pre- and post-semester SATS only twenty answered all questions in the survey. Given that MANOVA is known to be sensitive to violations of multivariate normality for small samples that are grossly nonnormal (Ito 1980, Gonzalez & Manly 1998), and notwithstanding the data showing no evidence of gross nonnormality from univariate and bivariate plots, permutational MANOVA (PERMANOVA) was used to test the null hypothesis of no change in mean attitudinal scores using PERMANOVA+ for PRIMER (Anderson, Gorley, & Clarke, K. R, 2008). The only assumption of the permutation test is that the observations are exchangeable under a true null hypothesis (Anderson 2001), which is equivalent to the assumption of independent and identically distributed observations. This assumption (i.e., the assumption of homogeneity of variances) appeared reasonable based on univariate plots (e.g., Figure 4) and a nonmetric multidimensional scaling plot (not shown) of the data. In the repeated measures PERMANOVA, participant ID was included as a between-subjects random effect and time as a within-subjects fixed effect, and p-values were based on 9999 permutations of residuals under a reduced model (Anderson, 2001; Anderson & ter Braak, 2003). In addition, Bayesian network learning algorithms were applied to the changes in attitudinal scores to learn the relationships among them and their relationship to final course mark.

Focus group discussions Data were gathered at 1-hour focus group discussions with students held in the second last week of semester, 2012 and 2013. The aim of the focus group was to elicit more information about student experiences and insight into reasons for their attitudes toward learning statistics. There were 17 participants in total, with 7 students (5 female, 1 male, and 1 email gender unknown) in 2012 and 10 students (7 female, 3 male) in 2013. Participants were informed about the project and given the option to withdraw; their verbal consent was obtained and recorded. They were encouraged to be open and constructively critical, using the following prompts to guide their discussion:

- What students liked/did not like about the subject.
- Advice they had for students/lecturers in this subject in the future.
- What learning strategies they used in this subject.
- Whether their attitude to statistics altered during the semester, or not, and why this was the case.

The comments were analysed to identify emerging themes used to illuminate findings in the quantitative data from the surveys. In the free flowing discussion of the focus groups it is not possible to tally numbers of responses – instead agreement is gauged by the tenor of the comments. The focus groups were led by an independent staff member from the Academic Language & Learning Unit.

Supplementary results from the annual Student Feedback Survey An annual Student Feedback Survey is undertaken as part of the University's quality assurance program. Students are asked to complete an online questionnaire which includes the following eight questions, each of which is rated on a 5-point Likert scale from strongly disagree to strongly agree:

- The learning outcomes of the subject were made clear to me.
- The subject enabled me to achieve the learning outcomes.
- I found the subject to be intellectually stimulating.
- I found the resources provided for the unit to be helpful.
- I received constructive feedback on my work.
- The feedback I received was provided in time to help me improve.
- The overall amount of work required of me for this subject was appropriate.
- Overall I was satisfied with the quality of this subject.

To assist our understanding of whether the redesign of the course had changed students' attitudes, the results for the last question on overall satisfaction were analysed here for the four years before and after the redesign. A weighted regression of the mean satisfaction scores against period (before, after) was done with the linear model (lm) function in R (R Development Core Team, 2013), with weights set equal to the number of respondents in each year because these are inversely proportional to the variance of the annual means.

3. RESULTS

3.1. PRE-SEMESTER DATA

The Spearman correlation matrix given as Table 2 of pre-semester attitudinal variables (using the short names of affect, cognitive, value, difficulty, interest, and effort), prior mathematics achievement (mathshis), mathematics cognitive competence (maths), statistics cognitive competence (confidence), expected grade (expgrade), gender, and final course mark (finmark) indicates that there are clear associations among the variables. Table 3 lists the questions associated with each variable (note that the first six variables correspond to the six attitude components, with the corresponding 36 questions, used in the SATS-36). A scatter plot matrix of the variables (not shown) indicated that the assumption of multivariate normality—required for the partial correlation tests of conditional independence of continuous variables in bnlearn, and for covariance-based tests in Amos—is reasonable. The sample mean and standard deviation for each of the attitude components given in Table 2 suggests that students were generally feeling confident and positive about the course at the beginning of semester, despite initial perceptions that it will be difficult (lower difficulty scores equate to the course being perceived as difficult).

Various algorithms were used to learn the causal structure compatible with these data, and it is important to note that this approach is exploratory and that the results are necessarily tentative. The algorithms within the bnlearn package (Scutari, 2009) can deal with a mixture of discrete and continuous variables. The score-based algorithms, *tabu* and *hill climbing*, appeared to better capture a structure that was consistent with SATS-M. Constraint-based algorithms produced sparser structures, which may be a consequence of the low power associated with the conditional independence tests (Fast, 2010). *Hill climbing* is a greedy search on the space of the directed graphs, and *tabu* is a modified hill climbing search that is able to escape local optima by selecting a network that minimally decreases the score function (Scutari, 2009). The *tabu* algorithm consistently produced the same structure, whereas *hill climbing* produced different structures from run to run. Bayesian networks produced by *hill climbing* and *tabu* are shown as Figure 1 and Figure

2. These two structures were indistinguishable on the basis of the BIC scores, which were -1792.8 for the *hill climbing* algorithm and -1794.6 for the *tabu* algorithm (note that higher values are better with the re-scaled scores used in bnlearn).

Table 2. Mean, standard deviation (SD), and Spearman correlations for attitude components from pre-semester data (note that gender was coded as 1 for male and 2 for female)

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Affect	4.32	1.40	--											
2. Cognitive	4.92	1.27	$.84^{**}$	--										
3. Value	5.32	1.04	$.41^{**}$	$.43^{**}$	--									
4. Difficulty	3.65	1.00	$.64^{**}$	$.65^{**}$	.16	--								
5. Interest	4.93	1.36	$.57^{**}$	$.48^{**}$	$.63^{**}$	$.27^{**}$	--							
6. Effort	6.28	0.81	.07	.09	$.22^{*}$	-.19	$.41^{**}$	--						
7. Maths	4.91	1.59	$.47^{**}$	$.52^{**}$	.07	$.36^{**}$	$.23^{*}$	-.04	--					
8. Mathshis	4.93	1.78	$.48^{**}$	$.48^{**}$	-.0004	$.32^{**}$	$.21^{*}$	.01	$.91^{**}$	--				
9. Confidence	5.14	1.55	$.60^{**}$	$.75^{**}$	$.34^{**}$	$.63^{**}$	$.44^{**}$	.09	$.55^{**}$	$.49^{**}$	--			
10. Gender	n/a	n/a	$-.27^{**}$	-.11	-0.09	-.18	$-.20^{*}$	-.003	-.09	-.11	-.17	--		
11. Expgrade	2.34	0.76	$-.38^{**}$	$-.49^{**}$	$-.37^{**}$	$-.27^{**}$	$-.31^{**}$	-.07	$-.24^{*}$	$-.22^{*}$	$-.46^{**}$	.18	--	
12. Finmark	66.23	19.66	$.22^{*}$	$.28^{**}$	.16	.08	$.26^{**}$	$.24^{*}$	.07	.05	$.24^{*}$	.06	$-.32^{**}$	--

**p*-value < 0.05, ***p*-value < 0.01

Using the structure produced by the *tabu* search (Figure 2) as a starting point, a structural equation model was fit to the data using Amos. This was initially done without gender in the model, and based on significance tests for the regression parameters, the arrow from expected grade to effort was removed (p-value = 0.653). Further, based on modification indices, the double-headed arrows confidence ↔ expgrade and confidence ↔ maths were added to allow these variables to covary. This seemed reasonable because these variables possibly represent the same underlying construct of expectancy/ability (see Eccles & Wigfield, 1995). The goodness-of-fit statistics for the initial model were $\chi^2(38) = 117.882$ with p-value < 0.001, comparative fit index (CFI) = 0.880, and root mean square error of approximation (RMSEA) = 0.150, and for the second model $\chi^2(37) = 64.659$ with p-value = 0.003, CFI = 0.959, and RMSEA = 0.089. A multiple-group analysis was then done to compare the structural weights, covariances, and residuals between males and females. No statistically significant differences were found, so it reasonable to conclude that the structure is invariant to gender. Finally, gender was added to the model, with an arrow between it and affect (as in the Bayesian network in Figure 2). The goodness-of-fit statistics were $\chi^2(47) = 77.254$ with p-value = 0.004, CFI = 0.955, and RMSEA = 0.083. The chi-square test result indicates a statistically significant difference between the observed and model-implied covariance matrices, but this is a common occurrence and may be attributable to the sensitivity of the test to sample size and the fact that the test is based on a central χ^2 distribution which assumes the model fits perfectly in the population (Byrne, 2010). The CFI and RMSEA values, on the other hand, suggest a well-fitting model (Byrne, 2010; McDonald & Ho, 2002). We note that of the 66 standardised residual covariances only four were larger than 2 and the largest residual was -2.325, the residual covariance between expected grade and final mark, which suggests there are no serious discrepancies between the observed and model-implied covariance matrices (Byrne, 2010;

McDonald & Ho, 2002). The final path diagram, together with standardised parameter estimates, is shown as Figure 3. The unstandardised direct, indirect, and total effects for this model are given in Table 4.

Table 3. Descriptions of the names of variables

Variable name	Question(s) asked
affect	I will like statistics.
	I will feel insecure when I have to do statistics problems.
	I will get frustrated going over statistics tests in class.
	I will be under stress during statistics class.
	I will enjoy taking statistics courses.
	I am scared by statistics.
cognitive	I will have trouble understanding statistics because of how I think.
	I will have no idea of what's going on in this statistics course.
	I will make a lot of math errors in statistics.
	I can learn statistics.
	I will understand statistics equations.
	I will find it difficult to understand statistical concepts.
value	Statistics is worthless.
	Statistics should be a required part of my professional training.
	Statistical skills will make me more employable.
	Statistics is not useful to the typical professional.
	Statistical thinking is not applicable in my life outside my job.
	I use statistics in my everyday life.
	Statistics conclusions are rarely presented in everyday life.
	I will have no application for statistics in my profession.
	Statistics is irrelevant in my life.
difficulty	Statistics formulas are easy to understand.
	Statistics is a complicated subject.
	Statistics is a subject quickly learned by most people.
	Learning statistics requires a great deal of discipline.
	Statistics involves massive computations.
	Statistics is highly technical.
	Most people have to learn a new way of thinking to do statistics.
interest	I am interested in being able to communicate statistical information to others.
	I am interested in using statistics.
	I am interested in understanding statistical information.
	I am interested in learning statistics.
effort	I plan to complete all of my statistics assignments.
	I plan to work hard in my statistics course.
	I plan to study hard for every statistics test.
	I plan to attend every statistics class session.
maths	How good at mathematics are you? (1 = very poor to 7 = very good)
mathshis	How well did you do in mathematics subjects you have taken in the past? (1 = very poorly to 7 = very well)
confidence	How confident are you that you can master introductory statistics material? (1 = not at all confident to 7 = very confident)
gender	What is your gender? (1 = male, 2 = female)
expgrade	What grade do you expect to receive in this subject? (1 = A to 5 = fail)

finmark	Final course mark (%)

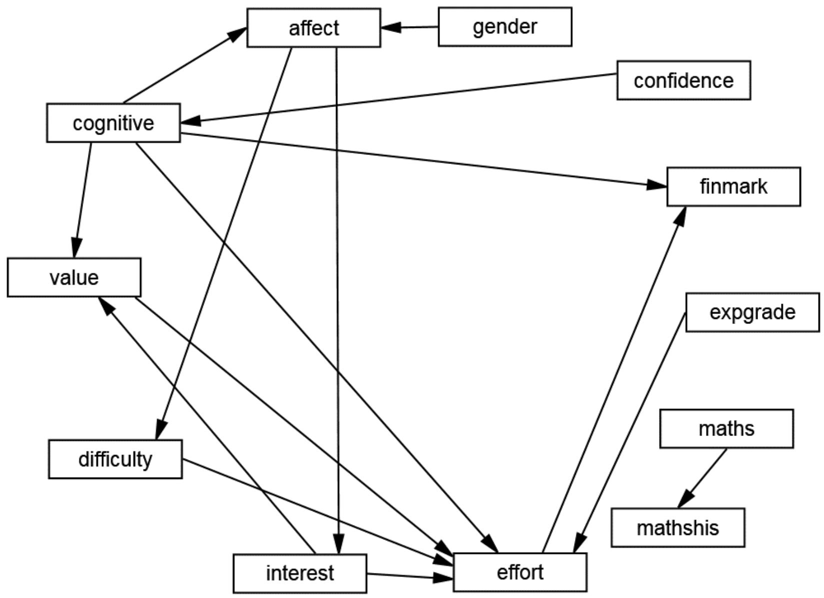

Figure 1. Bayesian network from the pre-semester data with hill climbing search and score algorithm

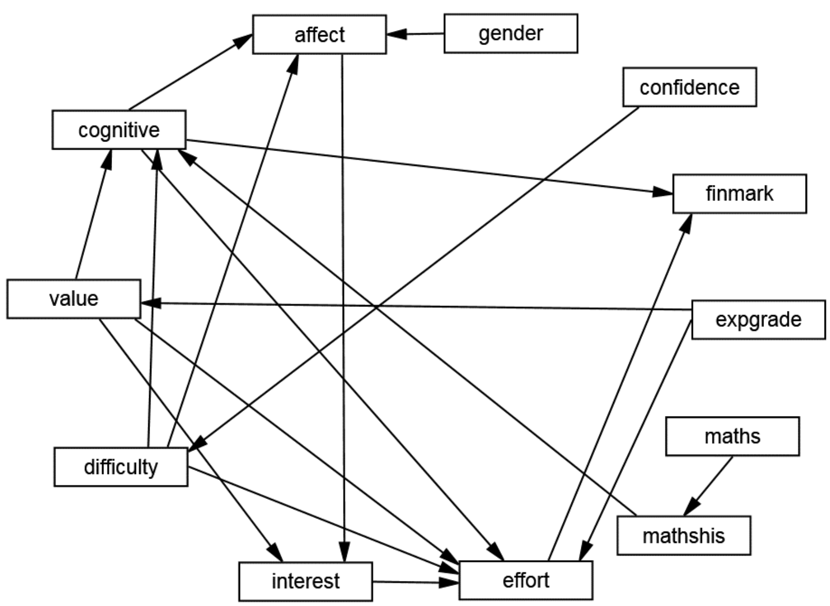

Figure 2. Bayesian network from the pre-semester data with tabu search and score algorithm

These results indicate that students with higher expectancies (expected grade) see more value in statistics, and those that had performed well in past mathematics courses (mathshis) had a greater sense of cognitive competence and were therefore prepared to put more effort into their studies. In turn, those that valued statistics and felt competent to succeed also believed they would like the subject and find it interesting. Ultimately, the students that felt cognitively competent had better course outcomes (finmark). Gender had an indirect effect on course outcomes via affect, with males liking statistics more and outperforming females, but the total effect of gender was small in comparison to the effects of other variables in the model (Table 4).

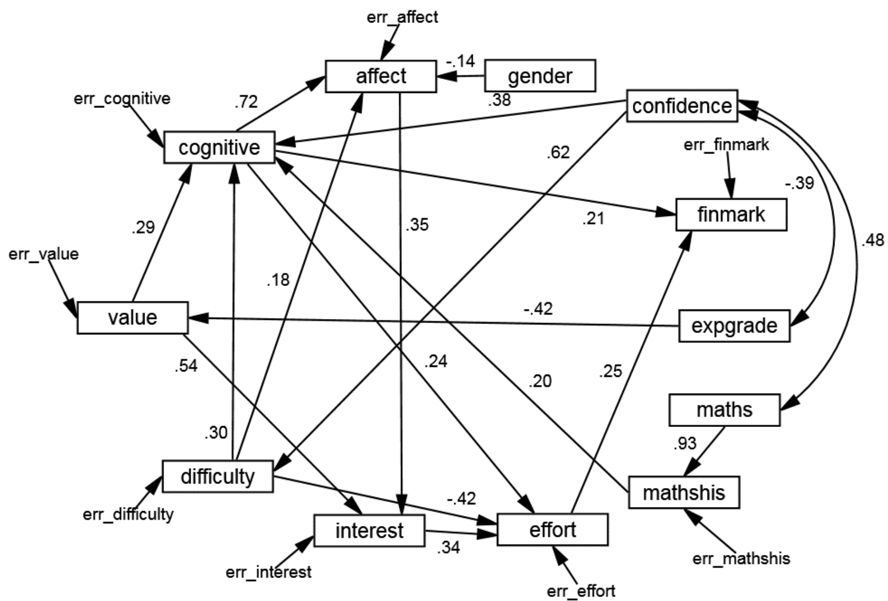

Figure 3. Path diagram for pre-semester data, with standardised parameter estimates shown

Table 4. Unstandardised direct, indirect, and total effects in the final model from pre-semester data

Variable	Effect	Gender	Maths	Exp-grade	Confi-dence	Maths-his	Value	Diffi-culty	Cogni-tive	Affect	Interest	Effort
Mathshis	Direct		1.02**									
	Indirect											
	Total		1.02**									
Value	Direct			-0.57**								
	Indirect											
	Total			-0.57**								
Difficulty	Direct				.37*							
	Indirect											
	Total				.37*							
Cognitive	Direct				.30**	.13*	.32*	.40*				
	Indirect		.13*	-.18*	.14*							
	Total		.13*	-.18*	.44*	.13*	.32*	.39*				
Affect	Direct	-.39*						.27*	.80**			
	Indirect		.11*	-.15**	.45*	.10*	.26*	.31*				
	Total	-.39*	.11*	-.15**	.45*	.10*	.26*	.58*	.80**			

Interest	Direct						.68*			.35*		
	Indirect	-.14**	.04*	-.44**	.16*	.04*	.09**	.20*	.28*			
	Total	-.14**	.04*	-.44**	.16*	.04*	.77*	.20*	.28*	.35*		
Effort	Direct							-.34**	.15		.19**	
	Indirect	-.03**	.03*	-.11*	-.03	.03*	.19*	.10*	.05**	.07**		
	Total	-.03**	.03*	-.11*	-.03	.03*	.19*	-.24**	.20*	.07**	.19*	
Finmark	Direct								3.25**			6.23
	Indirect	-.16**	.60*	-1.28*	1.25	.59*	2.24*	-.23	1.25*	.41**	1.17**	
	Total	-.16**	.60*	-1.28*	1.25	.59*	2.24*	-.23	4.51*	.41**	1.17**	6.23*

*p-value < 0.05, **p-value < 0.01

3.2. COMPARISON OF PRE AND POST-SEMESTER DATA

Twenty-seven students completed the pre- and post-semester SATS, but only twenty of these answered all questions in the survey (6 males and 14 females). As noted in Section 2.2, the mean course mark for the students that participated in the post-semester survey was approximately 10 percentage points higher than that of all enrolled students, and this suggests that this sample of students may not be representative of the population of students.

The means and standard deviations for the pre-semester, post-semester, and change in attitudinal scores are given in Table 5. In addition, the pre- and post-semester attitudinal scores were plotted for the twenty individuals to check for patterns within the data (e.g., see the plot for cognitive competence in Figure 4). It is worth noting that of these twenty students only two revised their expected grade for the subject: one down from A to C, and the other up from C to B.

Table 5. Means and standard deviations for pre-semester, post-semester, and change in attitude components (paired samples)

	Pre-semester		Post-semester		Change	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Affect	4.21	1.66	4.61	1.32	0.56	1.18
Cognitive	4.81	1.31	4.92	1.24	0.11	1.31
Value	5.55	0.88	5.39	1.21	-0.12	0.77
Difficulty	3.44	1.08	3.55	1.02	0.10	1.15
Interest	5.34	1.26	5.05	1.54	-0.31	1.64
Effort	6.24	0.78	5.99	0.80	-0.25	0.97

A repeated measures PERMANOVA was conducted for these twenty, and whereas the variation among participants was statistically significant, with a p-value < 0.001 for the random effect, the effect of time was found to be nonsignificant, with p-value of 0.167. Although this result indicates

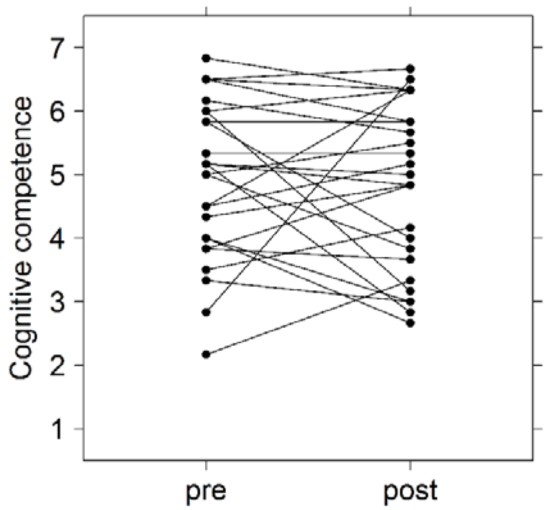

Figure 4. Pre- and Post-semester scores for individuals' cognitive competence

there was no change in attitudes at the population level, there was clearly some structure among the changes in attitudes as evidenced in the correlation matrix in Table 6. To explore this structure further the *tabu* and *hill climbing* score-based algorithms were applied to the data to learn the Bayesian network. The two score-based algorithms produced identical structures (Figure 5). Cognitive competence is the root node in this structure, suggesting that positive changes in cognitive competence hada positive effect on other attitudinal components and course outcomes.

Table 6. Pearson correlations among the changes in attitudinal scores and the final mark (Note: the attitudinal variables represent the change in scores from pre- to post-semester)

	1	2	3	4	5	6	7
1. Affect	--						
2. Cognitive	.75**	--					
3. Value	.34	.41	--				
4. Difficulty	.77**	.72**	.50*	--			
5. Interest	.22	.40	.54*	.31	--		
6. Effort	.09	.19	.23	.06	.49*	--	
7. Finmark	.68**	.55*	.64**	.61**	.50*	.35	--

*p-value < 0.05, **p-value < 0.01

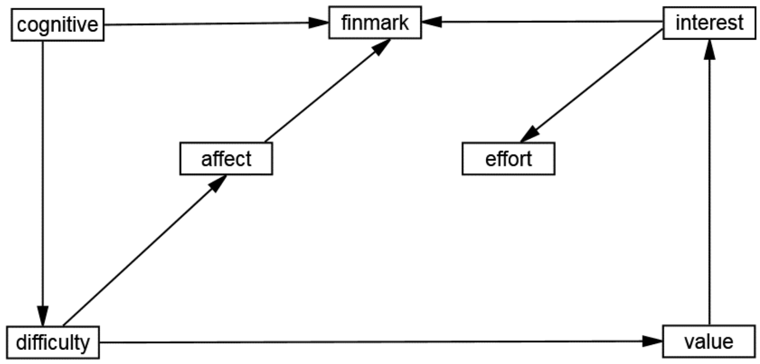

Figure 5. Bayesian network obtained for the changes in attitudes with tabu and hill climbing search and score algorithms (Note: the attitudinal variables represent the change from pre- to post-semester)

3.3. FOCUS GROUP DISCUSSIONS

Discussions indicated that participants had a wide range of attitudes to this course, both positive and negative, which suggests that they represented the views of all enrolled students despite there being only 17 students in the sample. The focus groups addressed many aspects of the course, but we concentrate here on what students liked or did not like about the course and the reasons given for attitude changes from the beginning to the end of the semester. The content of the discussion is presented according to the emergent themes of mathematical content, relevance to professional and everyday life, statistical literacy, co-operative learning, and teacher support.

A dominant theme in the discussions concerned mathematics. Some participants held negative attitudes from the start of the course and remained so. They had either failed a previous statistics course, or were returning to study after a break and were concerned about their maths skills. These students tended to see the course as being "forced" on them and their aim was to "get through it." Some students, particularly those with recent experience in maths or statistics, were positive from the start and remained so as indicated by this comment: "… this has been like a refresher for me cos I did it at school not long ago. This is more in depth than school but I had the fundamentals." However, many students expressed an initial negative attitude to statistics which altered as the course progressed and they realised that maths skill would not present a barrier to them: "At the start I was really worried because I have never been good at maths, but as we went along and [the lecturer] explained things I settled down and felt more comfortable and confident. It was not as scary as I thought it was going to be. The worked examples and online quiz helped and [the lecturer] explained that it was not just going to be maths and equations." There were also students, including students with English as an Additional Language (EAL), who reported that they looked forward to this course because they have strong maths skills and expected that assignments would be focused on maths rather than writing. These students were disappointed to discover that writing short and long answers was required in the assessments.

Another important theme from the discussions concerned the applicability of what they had learned to their other courses and everyday lives. Students appreciated that the examples used to present or expand statistical concepts were relevant to their disciplines. They spoke about transferring the skills gained in this subject to their discipline subjects such as genetics or psychology. One student commented, “I had probably never used stats and did not really know how I was going to go. I found we use stats a lot in our environment and genetics class so it made it easier to see what was going on. What I did in stats carried and grew into those subjects.” Students spoke about their excitement at learning new ways to analyse and summarise data, and the way studying this subject had broadened their knowledge on how to gather information using “the right statistic.” They also liked “learning about how to read results in journal articles,” and they discussed their growing ability to “conceptualise the meaning of data in the media.”

A number of students commented on challenges faced when learning statistical language. One student spoke about “scary words” rendering them “frozen.” Another student spoke about being unable to read their own lecture notes, despite careful preparation for the open-book exam, “… it was like reading Chinese.” There was agreement that even asking a question was a language challenge as is indicated by the comment: “I was glad when someone posted questions on [Learning Management System]. I wanted to know about that topic too but I did not know how to even start writing my question.”

There was considerable agreement that this course could have been more fun if learning had been more collaborative. Some students wanted to cooperate with peers in class and in a limited way for assessment as is evident in these two comments: “Get people into groups to discuss the assignment” and “For the assignment, I would have liked to have a group assignment… i.e., an assignment that is designed for group work—not one that I did as a group.” It was interesting to note that those who wanted to engage with peers wanted explicit instruction to do so. For example, one student explained, “… we did it [group discussion] in the first day but it was not reinforced again, we did not know if we could work together.”

Teacher support and feeling safe to ask questions emerged as an important theme in discussions about what they liked about the course. Students commented that the teachers’ approach was helpful to learning statistics, their “knowledge and personable teaching style has been outstanding.” They reported that they could easily approach their statistics teachers, and they could subtly indicate their need for clarification—“ …[tutor]… noticed my questioning face”—and receive help at the point of need. One student felt confident that s/he would not be made to feel stupid: “The dumber the question is the more [lecturer] winds it back—I could ask anything.”

All five themes have a bearing on students’ feeling of cognitive competence, and all but teacher support relate directly to the recommendations in the GAISE report. The reduced mathematical content lessened the anxiety that many students felt at the beginning of semester, and the emphasis on statistical thinking gave them the confidence to apply their learning in contexts outside of their statistics classes. Teacher support gave students the confidence to explore ideas and ask questions. Cooperative learning, which many of the participants wanted more of, has been reported to have a positive effect on mathematics self-concept (Townsend, Moore, Tuck, & Wilton, 1998). Interestingly, the greater emphasis on statistical literacy (including critiquing news articles) meant that assessments were now based on written answers, and some students lacked the confidence to use statistical language in written communication. The focus group discussions indicate that our particular implementation of the GAISE recommendations had both positive and negative impacts on students’ attitudes toward statistics, and this observation is consistent with our findings from the comparison of pre- and post-semester survey data (Section 3.2), where changes in attitudes ranged from positive to negative, and appeared to be driven by changes in cognitive competence.

3.4. SUPPLEMENTARY RESULTS FROM THE ANNUAL STUDENT FEEDBACK SURVEY

A time series plot of annual mean satisfaction with introductory statistics is given as Figure 6. The data indicate an upward shift in level following the implementation of the redesigned subject in 2012; specifically, the mean overall satisfaction score increased from 3.44 before the redesign to 4.08 after, on a 5-point Likert scale, with $t(6) = -3.971$ and p-value of 0.007. Although acknowledging the potential for confounding bias (see Discussion), these results suggest that the redesign contributed to an improvement in satisfaction.

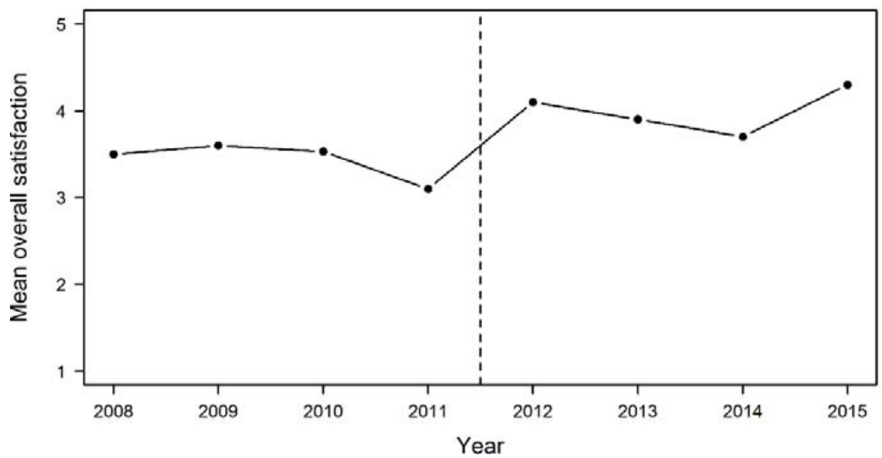

Figure 6. Mean satisfaction score for the four years before and after redesigning the course (the dashed vertical line represents the time of transition)

4. DISCUSSION

With respect to the causes and consequences of students' attitudes toward statistics, results from the pre-semester data are consistent with EVT (Eccles & Wigfield, 1995) and the SATS-M (Ramirez et al., 2012; Schau, 2003). Broadly speaking, our model predicts that students with higher expectancies (expected grade) see more value in statistics, and in turn they have a greater interest and sense of cognitive competence which, through a series of direct and indirect effects on other attitudinal components, leads to better course outcomes. In addition, prior achievement in mathematics was found to be related to attitudes toward statistics; specifically our model suggests it determines students' feeling of cognitive competence. The relationship between prior achievement-related experiences and cognitive competence was also reported by Sorge and Schau (2002) and Hood et al. (2012), with similar standardised path coefficients: ours was 0.2, Hood et al. estimated it to be 0.19, and from Sorge and Schau it was 0.13. Also similar to the findings reported by Hood et al., our evidence indicates that course outcomes (students' final course mark) are determined by cognitive competence and effort. Their standardised path coefficient for the effect of effort was 0.27, and ours was 0.25. Where our models differ is that we found a direct effect of cognitive competence on final course mark and an indirect effect mediated by effort, whereas Hood et al. predict only an indirect effect that is mediated by expectancies. Neither Sorge and Schau nor Hood et al. included gender in their model, but consistent with SATS-M we found a direct effect of gender on affect, specifically males generally liked statistics more than females.

With regard to intra-attitudinal relationships there were few similarities between our model and the models of Sorge and Schau (2002) and Hood et al. (2012), and this is clearly an area in need of further research. The most notable among these was the direct effect of difficulty on cognitive competence, with standardised path coefficients of 0.23 (Hood et al.), 0.3 (ours), and 0.67 (Sorge & Schau). Similar to Sorge and Schau we found a direct effect of cognitive competence on affect, while Hood et al. found the reverse relationship. We found a direct effect of value on cognitive competence, but again Hood et al. found the reverse relationship.

On the matter of whether our implementation of the GAISE recommendations made a difference to attitudes and satisfaction, the results are varied. Observational data from annual Student Feedback Surveys indicated that mean overall satisfaction increased after redesigning the subject, from 3.44 to 4.08 on a 5-point Likert scale in the four years before and after the redesign. According to expectation disconfirmation theory an increase in satisfaction should lead to a positive change in attitudes, and empirical evidence obtained in variety of settings, including in studies of student satisfaction in a higher education setting, supports a link between satisfaction and post-attitudes (Athiyaman, 1997, 2004; Oliver, 1980). However, similar to other researchers that have examined the effects of interventions on students' attitudes toward statistics (e.g., Brandsma, 2000; Carnell, 2008; High, 1998; Rhoads & Hubele, 2000), we found no evidence of a change in mean attitudes after having completed the course. The failure to detect a change in mean attitudes may be related to the small sample size for the pre-post component of this study.

Despite the failure to detect a change in mean attitudes toward statistics over the semester, we did notice that the changes in attitudes were correlated with one another, and we explored this correlation structure using Bayesian networks. The root node in the resulting model was the change in the feeling of cognitive competence, and it had both a direct effect and indirect effect via a change in perceived difficulty, value, interest, and affect on final course mark. This suggests that the course improved the feeling of cognitive competence for some but not all students, and this in turn had a bearing on other attitudes and course outcomes. Again we acknowledge the small sample size for the pre-post study presented here, and accordingly our conclusions should be regarded as tentative and as a possible starting point for future research.

In an attempt to understand why attitudes improved for some students but not others we looked to the results of the focus group discussions. On the one hand it was clear that many students felt more cognitively competent when they realised that this was not a mathematics course. This was certainly expected because it was one of the motivations for focusing on statistical thinking in the GAISE recommendations (Aliaga et al., 2005). Others, on the other hand, felt less competent when they realised that written answers were required, and this seems to be connected to statistical literacy and the challenges associated with using statistical language. The problem seems to involve students' ability to use statistical terms in written prose. In this course we emphasize the meaning of statistical terms and concepts, as well as misconceptions where statistical terms have quite different meanings than the colloquial sense of the word, and we model written answers. There is certainly scope for other activities; for example, group projects that give students practice at communicating using statistical language (Aliaga et al., 2005), answering cloze questions (Taylor, 1953) where students are required to replace missing words in a passage of text, and activities that require students to deconstruct conclusions from statistical analyses to explain the meaning and order of the parts. Another important theme from the focus groups was the desire for collaborative projects. We left the choice of group work up to the students, but some students expressed a preference for explicit instruction to work in groups. Students are familiar with group work in their other courses, and they clearly benefit from cooperative or collaborative projects. Springer, Stanne, and Donovan (1999) note that meta-analyses have demonstrated the favourable effects of small-group work on achievement and productivity, psychological health and self-esteem, intergroup attitudes, and attitudes toward learning, and Townsend, Moore, Tuck, and Wilton (1998) have shown that cooperative learning can have a positive effect on mathematics self-concept.

As noted already, there were some limitations in the current study. First, it is important to appreciate that learning Bayesian networks from data is an exploratory process, and whereas the outputs can be very informative further research is needed to test the proposed structures. It is not possible for the score-based algorithms used here to search the entire space of directed graphs, and the outputs can be the result of local maxima being achieved. Even if the skeleton of the structure (i.e., the undirected version of the graph) was correct, there can be numerous directed graphs that fall within the same equivalence class (Koski & Noble, 2012; Pearl, 2000) where the directions of some arrows may be reversed. Second, the low participation rate in the focus groups and low response rate in the post-semester survey may limit the generalisability of these findings, but we note that in both we seemed to capture the full gamut of attitudes and changes in attitudes. The problem with the response rate is related, not surprisingly, to the time of semester when these components of the study were conducted. Reminders were emailed to students at least twice, and time was set aside in computer labs for students to complete the post-semester questionnaire, but at this end of the semester students have numerous assessments due, and they are already asked to complete a Student Feedback Survey for each of the courses in which they are enrolled. Further, we recognise the potential for confounding bias in the before-after comparison of overall satisfaction from Student Feedback Surveys, and we note two events in particular that may have had a bearing on the increase in satisfaction. The Student Feedback Survey changed from a paper-based questionnaire to an online questionnaire in 2013 (a year after implementation of the redesigned course). This seemed to reduce the response rate from 30-40% before 2013 to about 10% in 2013 and 2014, but steps were taken to address this and the response rate was 35% in 2015. There have also been different tutors over the eight year period, but responses to open-ended questions on the Student Feedback Survey typically praise the efforts of tutors. The most important factor influencing the change in overall satisfaction, we believe, was the incorporation of the GAISE recommendations into the course. We can certainly attest to a positive change in the “vibe” in classes since implementing the new course.

The insights obtained by the combination of approaches used here suggest that future attempts at interventions to improve attitudes toward statistics should move beyond an assessment of mean changes in attitudes, toward mixed methods that seek to understand individuals’ change in attitudes using quantitative and qualitative methods (see Jick, 1979), and incorporating data such as Student Feedback Surveys if they are available. Indeed it would be worth considering a combination of attitudes and satisfaction in survey instruments in future studies. The literature on consumer satisfaction is extensive, but few studies have considered models of satisfaction in the higher education sector and its influence on attitudes, enrolment intentions, and word-of-mouth recommendations.

ACKNOWLEDGEMENTS

We are grateful to all of the students that participated in this research. We also thank the editors and an anonymous reviewer whose valuable comments helped to substantially improve this manuscript.

REFERENCES

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., . . ., Utts, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report*. American Statistical Association. Alexandria, VA: American Statistical Association.
[Online: http://www.amstat.org/education/gaise]

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46.

Anderson, M. J., Gorley, R. N., & Clarke, K. R. (2008). *PERMANOVA+ for PRIMER: Guide to software and statistical methods*. Plymouth, UK: PRIMER-E Ltd.

Anderson, M. J., & ter Braak, C. J. F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, *73*(2), 85–113.

Athiyaman, A. (1997). Linking student satisfaction and service quality perceptions: The case of university education. *European Journal of Marketing, 31*(7), 528–540.

Athiyaman, A. (2004). Antecedents and consequences of student satisfaction with university services: A longitudinal analysis. *Academy of Marketing Studies Journal, 8*(1), 89–105.

Best, J. (2005). Lies, calculations and constructions: Beyond "How to lie with statistics." *Statistical Science, 20*(3), 210–214.

Best, J. (2008). Beyond calculation: Quantitative literacy and critical thinking about public issues. In B. L. Madison & L. A. Steen (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 125–135). Washington, DC: Mathematical Association of America.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364.

Brandsma, J. A. (2000). *Data collection and analysis: Examining community college students' understanding of elementary statistics through laboratory activities.* (Doctoral Dissertation), North Carolina State University. Retrieved from ERIC Document Reproduction Service No. ED 461396.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education, 16*(1), 1–15.
[Online: http://ww2.amstat.org/publications/jse/v16n1/carnell.html]

Cobb, G. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3–43). Washington, DC: Mathematical Association of America.

delMas, R., Ooms, A., Garfield, J., & Chance, B. (2006). Assessing students' statistical reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online:
https://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/6D3_DELM.pdf]

Eccles, J. S., & 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.

Fast, A. S. (2010). Learning the structure of Bayesian networks with constraint satisfaction. *Open Access Dissertations*, Paper 182.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education, 2*(2), 1−15.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal, J. Garfield, & Y. Gal (Eds.), *The assessment challenge in statistics education* (pp. 37−51). Amsterdam, The Netherlands: IOS Press.

Gonzalez, L., & Manly, B. F. (1998). Analysis of variance by randomization with small data sets. *Environmetrics, 9,* 53–65.

High, R. V. (1998). Some variables in relation to students' choice of statistics classes: Traditional versus computer-supported instruction. Adelphi University. Retrieved from ERIC Document Reproduction Service No. ED 427 762.

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal, 11*(2), 72–85.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Hood.pdf]

Ito, P. K. (1980). Robustness of ANOVA and MANOVA test procedures. In P. R. Krishnaiah (Ed.), *Handbook of statistics 1. Analysis of variance* (pp. 199–236). Amsterdam, North Holland.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*(4), 602–611.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Publications.

Koski, T. J., & Noble, J. M. (2012). A review of Bayesian networks and structure learning. *Annales Societatis Mathematicae Polonae. Series 3: Mathematica Applicanda, 40*(1), 53−103.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64.

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

Olani, A., Hoekstra, R., Harskamp, E., & Van Der Werf, G. (2011). Statistical reasoning ability, self-efficacy, and value beliefs in a reform based university statistics course. *Electronic Journal of Research in Educational Psychology, 9*(1), 49–72.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramirez, C., Schau, C., & Emmioğlu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal, 11*(2), 57–71.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Ramirez.pdf]

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10*(3), 6–13.
[Online: http://ww2.amstat.org/publications/jse/v10n3/rumsey2.html]

Rhoads, T. R., & Hubele, N. F. (2000). Student attitudes toward statistics before and after a computer-integrated introductory statistics course. *Education, IEEE Transactions on, 43*(2), 182−187.

Schau, C. (2003). *Students' attitudes: The "other" important outcome in statistics education.* Paper presented at the Joint Statistical Meetings, San Francisco, CA.

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal, J. Garfield, & Y. Gal (Eds.), *The assessment challenge in statistics education* (pp. 91–104). Amsterdam, The Netherlands: IOS Press.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*(5), 868–875.

Schield, M. (1995). Correlation, determination and causality in introductory statistics. *Proceedings of the ASA Section of Statistical Education* (pp. 189–194). Alexandria, VA: American Statistical Association.

Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software*, *35*(3), 1–22.
[Online: http://www.jstatsoft.org/v35/i03]

Sorge, C., & Schau, C. (2002). *Impact of engineering students' attitudes on achievement in statistics*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
[Online: http://evaluationandstatistics.com/AERA2002.pdf]

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30,* 415–433.

Townsend, M. A. R., Moore, D. W., Tuck, B. F., & Wilton, K. M. (1998). Self-concept and anxiety in university students studying social science statistics within a co-operative learning structure. *Educational Psychology, 18*(1), 41–45.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.

WARREN PAUL
Department of Ecology, Environment and Evolution, La Trobe University,
Wodonga, Victoria 3689, Australia

APPENDIX: DESCRIPTION OF THE REDESIGNED STATISTICS COURSE

The list of topics and specific learning outcomes in the course are given in Table A1. One topic is covered each week in a 1-2 hour lecture and 2 hour computer laboratory. Time is made available in lectures for students to ask questions or discuss and answer concept-related questions, either via a show of hands or using the online student response system Socrative (http://www.socrative.com/) which students access through a mobile device or laptop computer. The online student response system is also used to collect and display data on the students during class. Considerable use is made of freely available Java applets for demonstrating concepts relevant to statistical inference, such as sampling distributions, confidence intervals, and power and sample size calculations. Computer laboratory sessions usually involve analysis of real data with SPSS, where students are guided through the process of identifying the number and types of variables that are relevant to the question, choosing and performing an appropriate analysis, and writing a brief conclusion to answer the question. These sessions sometimes involve other activities such as experimenting with a power and sample-size calculator (http://homepage.stat.uiowa.edu/~rlenth/), completing select-and-fill concept maps (Schau & Mattern, 1997), or evaluating news stories (Best, 2005, 2008). All materials are available online via the University's learning management system, including links to short (5-10 minute) videos for each topic, such as those produced by Annenberg Learner for the *Against All Odds* series (http://www.learner.org/resources/series65.html) and relevant YouTube videos.

Assessments include weekly online quizzes, two assignments, and a final open-book exam. The online quizzes are adapted from or modelled on the multiple choice questions that are available from the Assessment Resource Tools for Improving Statistical Thinking website (ARTIST, https://apps3.cehd.umn.edu/artist/, see delMas, Ooms, Garfield, & Chance, 2006) and sources such as the GAISE College Report (Aliaga et al., 2005). The first assignment involves two tasks. The first task, which is started in the first computer laboratory, requires each student to explore possible applications of statistics in a field that is of interest to them (such as child psychology, neuroscience, ecology, environmental management, business management, market research, etc.), and to find specific examples that illustrate how statistics is used to search for patterns (i.e., trends or associations) in data, and how these patterns in turn have helped to identify a problem, find the causes of a problem, and develop a solution to a problem. The second task involves analysing a real data set with SPSS to answer a specific question using descriptive statistics. The second assignment also involves two tasks: critically evaluating a news article, and analysing a real data set in SPSS to answer a question using both descriptive and inferential statistics. The final open-book exam involves some multiple choice questions, identifying various scenarios as surveys, observational studies or experiments, and choosing the appropriate descriptive and inferential analyses for the situation, and interpreting the SPSS output from an analysis and writing a brief conclusion.

Table A1. List of topics and specific learning outcomes for the redesigned course

Topic	Specific learning outcomes
1. What is Statistics?	Explain why we need data. Explain what statistics is used for. Explain the meanings of the terms observation unit, variable, and data. Classify a variable as either qualitative (categorical) or quantitative.
2. Descriptive Statistics for One Variable 3. Descriptive Statistics for Two Variables	Choose an appropriate graph or numerical summary for data on one or two variables. Construct a graph or numerical summary using SPSS. Describe the important features of a graph or numerical summary and suggest possible reasons for those features.
4. Producing Data	Recognise the difference between 1) surveys and causal studies, and 2) experiments and observational studies. Explain the terms selection bias, measurement bias, and confounding bias. Explain why random sampling is important. Explain why random assignment is important.
5. Thinking Critically about Statistics in the Media	Critically evaluate the statistics presented in the media in order to assess the value of the information.
6. Introduction to Inferential Statistics	Explain the difference between descriptive statistics and inferential statistics. Explain why sampling distributions are so important for statistical inference. Calculate a t-score and use it to look up probabilities in a table of the t distribution.
7. Hypothesis Testing for One Quantitative Variable 8. Confidence Intervals for One Quantitative Variable 9. Hypothesis Tests and Confidence Intervals for One Quantitative and One Qualitative Variable 10. Chi-square Tests for One of Two Qualitative Variables 11. Correlation and Regression for Two Quantitative Variables	Choose an appropriate inferential analysis (hypothesis test and/or confidence interval) for data on one or two variables. Perform an inferential analysis using SPSS. Report conclusions from an inferential analysis. Determine the sample size required for a study.