STUDENT ATTITUDES TOWARD STATISTICS
FROM A RANDOMIZATION-BASED CURRICULUM

Todd Swanson[1], Jill VanderStoep[1] and Nathan Tintle[2]
[1]Hope College, Holland, Michigan USA
[2]Dordt College, Sioux Center, Iowa USA
nathan.tintle@dordt.edu

Recently, a large national (US) sample was used to evaluate attitudes toward statistics among undergraduate students. The majority of the courses in the sample used a similar (AP Statistics) curriculum. Recent interest in the use of randomization-based methods in introductory statistics raises substantial questions about the conceptual effectiveness and attitudes of students in such courses. To begin to better understand the advantages and disadvantages of a randomization-based curriculum, we evaluated student attitudes in courses using randomization curricula and compared these to students using a traditional curriculum. Overall, there were only small, statistically and practically insignificant, differences in student attitudes between the two samples. While randomization approaches remain new, our analysis suggests that these curricula may not be harming nor improving students' attitudes toward statistics relative to traditional courses in undergraduate courses.

INTRODUCTION

A consensus curriculum for the AP Statistics-equivalent, algebra-based introductory statistics course for undergraduate students has persisted since at least the late 1990s (hereafter, Stat 101; Schaefer, 1997). The traditional curriculum begins with descriptive statistics for one and two-variables, often transitions briefly to questions of study design, and then spends time on basic probability theory as a foundation for teaching sampling distributions and the central limit theorem. The course ends considering statistical inference using both confidence intervals and tests of significance. A consensus curriculum, along with the burgeoning field of statistics education gave rise to a period of reflection about the consensus standards for pedagogy in Stat 101. This focus culminated in the GAISE guidelines which were endorsed by the American Statistical Association in 2005 (GAISE, 2005). However, as argued elsewhere, substantial changes in the use of and access to computational technology was rapidly changing the way statistics was being conducted in practice, as well as the potential avenues for introducing students to statistical concepts (Cobb, 2007).

Cobb (2007) and others argued that these computational breakthroughs and changes in statistical practice showed a need for serious reconsideration of the topics and the order in which they were presented in Stat 101. In particular, Cobb argued for potential use of simulation, bootstrapping and permutation tests (hereafter termed "randomization methods") in Stat 101 as an alternative to the introduction the concepts of statistical inference (the logic and scope of inference) using asymptotic tests alone (e.g., t-tests). For example, it may be more intuitive for students to understand the logic behind and algorithm for a permutation test comparing two group means than the two-sample t-test. This is especially true when the permutation test is introduced using tactile strategies, like shuffling notecards containing the quantitative values for each observation. Likewise, students may be able to more easily understand the idea of a binomial distribution by simulating data using coin flips and spinners than discussing the motivation behind the formula for the binomial distribution probability mass function. In short, the arguments for randomization methods came down to an argument for an alternative path to both deeper and broader understanding of both the logic and scope of inference, while potentially not sacrificing the ability for students to understand and conduct traditional (asymptotic) tests.

During the last few years, increasing attention has been paid to these arguments with numerous panels and presentations at national and international conferences. In general, momentum is growing for the use of these methods for a mix of quantitatively justified and anecdotal reasons. Notably, an early version of such a curriculum showed improved student understanding of key components of the logic of inference, while sacrificing little performance in other areas on a valid and reliable test of student learning outcomes in Stat 101 (Tintle et al., 2011).

These students also showed improved retention in these same areas (Tintle et al., 2012); results which have been maintained at other institutions and in more recent versions of the curriculum (Tintle 2014). Furthermore, numerous curricula using these methods have been published (e.g., Lock et al. 2013) or have matured to a stable curriculum (e.g., Garfield et al., 2012, Tintle et al., 2014).

The arguments for a randomization-based curriculum go beyond merely improving student's conceptual understanding. Anecdotal evidence is mounting that students may be more engaged in the course through its integral use of tactile simulations and other hands-on active learning strategies (e.g., a permutation test or simulation of a binomial distribution is easily adapted to an in-class, hands-on activity, while a t-test may be less so). Furthermore, the approach keeps students "closer" to the data by using resampling techniques (bootstrapping and permutation), instead of heading down the often challenging path of abstract probability theory.

In short, proponents of the randomization-based approach to statistical inference argue that the approach may offer substantial benefits over the consensus curriculum due to its inherent affinity toward the GAISE guidelines, without sacrificing the outcomes desired by many client departments ("teach my students a t-test!"). However, to date, no published results on students' attitudes in a randomization-based curriculum exist.

In fact, it is only recently that large-scale consideration of student attitudes in Stat 101 and related courses have been assessed at the national level in the United States, using a valid and reliable instrument (Schau, Miller & Petocz, 2012). Recently, a series of articles were published involving the Survey of Attitudes Toward Statistics (SATS) with approximately 2200 introductory statistics (Stat 101) students (Schau & Emmioglu, 2012). Schau and Emmioglu found that, for students mainly taking the consensus curriculum, students attitudes generally stayed the same or decreased over the course of the semester, across six different components of the SATS. Given the importance of attitudes of students toward statistics as related to one of the goals of Stat 101 (Ramirez, Schau & Emmioglu, 2012), these results were not particularly encouraging.

In this paper we consider the attitudes of students before and after a randomization-based Stat 101 course using SATS, as well as their change during the semester. Results are qualitatively and quantitatively compared to the results of Schau and Emmioglu (2012) for US Stat 101 students who experienced the traditional curriculum.

METHODS

Instrument

The Survey of Attitudes Toward Statistics (SATS) was used to assess the attitudes of statistics students in 15 sections of statistics, taught by 14 different instructors at 11 different institutions. In total, 425 students all using the Fall 2013 edition/version of Tintle et al. (2014), took SATS both pre-course and post-course. Administration of the tests varied between instructors but was generally during the first week of classes for the pre-test and the week before or during finals week for the post-test.

The SATS is comprised of six subscales, along with an example item, are described briefly here. See Schau and Emmioglu (2012) for further description, and a review of other related literature.

- **Affect** (6 items) – students' positive and negative feelings concerning statistics – "I am scared by statistics."
- **Cognitive Competence** (6 items) – students' attitudes about their intellectual knowledge and skills when applied to statistics – "I can learn statistics."
- **Value** (9 items) – students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life – "I use statistics in my everyday life."
- **Difficulty** (7 items) – students' attitudes about the difficulty of statistics as a subject – "Most people have to learn a new way of thinking to do statistics."
- **Interest** (4 items) – students' level of individual interest in statistics – "I am interested in using statistics."

- **Effort** (4 items) – amount of work the student expends to learn statistics – "I plan to work hard in my statistics course"

The SATS consists of the items above with responses coming in the form of students responses to 7-point Likert scales (1=Strongly disagree, 4= Neither disagree nor agree, 7=Strongly agree), with approximately half of the items positively worded, and half negatively worded. The items are averaged together within each scale to yield a single value for each student for each subscale.

Curriculum

All students in this sample used the Fall 2013 edition of the *Introduction to Statistical Investigations* (ISI) textbook (Tintle et al., 2014). We provide a brief overview of the text here. A more detailed description is provided in a companion paper (Tintle, 2014). The ISI approach starts by teaching students about the four pillars of inference: significance (hypothesis tests), estimation (confidence intervals), generalizability (principles of good sampling) and causation (principles of good experimental design). The first three pillars are introduced by way of tests on a single proportion and rely on simulation of binomial distributions using tactile strategies (coin flips) and easily accessible free web applets. Students are also introduced to a theory-based approach to both testing and confidence intervals as a mathematical prediction of what you will obtain if you simulate. Tests and confidence on a single mean are also introduced in these first three chapters. The fourth pillar of inference (causation) begins discussions evaluating relationships between two variables. Evaluating relationships between two variables is the theme throughout the remainder of the course (Ch. 5-10), with chapters covering two proportions, two means, paired data, multiple proportions, multiple means and regression/correlation. Each of these chapters follows a similar format: Descriptive statistics, Simulation-based inference and Theory-based (asymptotic) inference. An emphasis on guided discovery, the use of real, published research data and conceptual understanding are maintained throughout. More detailed descriptions are available elsewhere (Tintle et al., 2014, Tintle, 2014).

RESULTS

Table 1 shows the values of Cronbach's alpha for each of the six subscales at both the pre- and post-test administrations. All subscales showed sufficient reliability (alpha>0.70) using the same criteria as Schau and Emmioglu (2012).

Table 1. Reliability (Cronbach's alpha)

	Pre-course	Post-course
Affect	0.83	0.89
Cognitive	0.87	0.88
Difficulty	0.86	0.90
Effort	0.71	0.80
Interest	0.88	0.91
Value	0.82	0.80

Table 2 shows the average value of the subscales both pre- and post-course, as well as the change during the semester. Overall, students entered the course with generally neutral Affect and Difficulty attitudes (between 3.5 and 4.5), while Cognitive Competence, Interest and Value were positive (between 4.5 and 5.5), and Effort was very positive (above 5.5).

Table 2. Overall pre- and post-course student attitudes (n=425)

	Pre-course	Post-course	Change
Affect	4.06 (0.89)	4.19 (1.17)	0.13 (1.05)*
Cognitive	4.66 (0.86)	4.80 (1.03)	0.14 (0.94)**
Difficulty	3.69 (0.56)	3.90 (0.78)	0.21 (0.69)***
Effort	6.27 (0.81)	5.73 (0.97)	-0.54 (1.00)***
Interest	4.64 (0.90)	4.22 (1.09)	-0.41 (0.92)***
Value	5.07 (0.73)	4.84 (0.88)	-0.24 (0.78)***

*$p<0.05$, **$p<0.01$, ***$p<0.001$

By the end of the course, attitudes exhibited small, but significant changes in all cases. In particular, students reported better affective feelings toward statistics and better cognitive competence about statistics at the end of the course than at the beginning. Furthermore, students reported being less interested in statistics and valuing it less at the end of the course than at the beginning. Finally, students reported having put less effort into the course and that it was harder than they anticipated at the beginning of the semester. These changes were significantly different across instructors (details not shown).

Table 3 shows change scores between the randomization curriculum compared to the large national sample using the consensus curriculum (Schau & Emmioglu, 2012). While there were no statistically significant differences ($p>0.05$ in all cases) overall when comparing the changes on the six subscales of the SATS between the randomization curriculum and the national sample (Schau & Emmioglu, 2012) we comment briefly on the overall trends. There was more improvement in students about cognitive competence in the randomization sample, more increase in students' perceptions of the difficulty of statistics in the randomization sample, more decrease in anticipated effort, less decrease in student interest and less decrease in perceived value of statistics in the randomization sample.

Table 3. Comparison vs. national sample using traditional curricula

	Change (post-pre)		
	Randomization (n=425)	Traditional (n=2200)	Difference[1]
Affect	0.13 (1.05)	0.13 (1.23)	0.00
Cognitive	0.14 (0.94)	0.10 (1.06)	0.04
Difficulty	0.21 (0.69)	0.15 (0.84)	0.06
Effort	-0.54 (1.00)	-0.48 (1.14)	-0.06
Interest	-0.41 (0.92)	-0.50 (1.25)	0.09
Value	-0.24 (0.78)	-0.32 (0.96)	0.08

[1] P-value from independent samples t-test
*$p<0.05$, **$p<0.01$, ***$p<0.001$

CONCLUSION

Overall, we found very similar attitudes pre- and post- course comparing the traditional (consensus) Stat 101 curriculum to the randomization curriculum. In particular, we found that students Affect, Cognitive competence and perceived Difficulty significantly increased, though none of the changes was particularly large. While the change in scores for both Cognitive competence and difficulty were larger for the randomization curriculum as compared to the traditional curriculum, neither change was statistically significant. While neither of these changes was statistically significant, we note that a perceived improvement in cognitive competence is potentially in line with findings showing actual improvement in conceptual understanding (Tintle et al., 2011, 2012; Tintle, 2014), however, importantly, perceived improvement in cognitive ability does not always translate into actual ability. Interestingly, students also had a larger increase in the perceived difficulty level of the course with the randomization curriculum.

Additionally, students had less interest in, perceived less value of and put in less effort by the end of the course, as compared to the beginning with both randomization and traditional curricula. Declines in interest and value were both less with the randomization curriculum, while declines in effort were more. Again, however, these results were not statistically significant.

Further studies, with larger sample sizes, and more sophisticated analyses which consider the many potential covariates affecting student attitudes which are potentially different between the two sample sizes are needed before conclusive statements about the findings in these areas can be drawn. However, a few general conclusions can be drawn.

First, we do not observe substantial differences between the attitudes of students in randomization-based curricula as compared to students taking the consensus curriculum. This suggests that the randomization-based curriculum is neither a panacea nor does it harm student attitudes any differently than does the traditional curriculum. Given the consistent improvement in some areas of conceptual understanding, there may be an overall net gain for students from a randomization-based curriculum. Second, many of the instructors in this sample were relatively new to teaching with the randomization-based curriculum. This and numerous other pedagogical, institutional and other factors should be accounted for in future analyses and samples comparing the two curricula. We hope to have completed further analyses in this regard prior to the July 2014 presentation in Flagstaff, but future studies will be needed which can more comprehensively address this issue.

REFERENCES

Cobb, G.W. (2007). The Introductory Statistics Course: a Ptolemaic Curriculum? *Technology Innovations in Statistical Education, 1*(1).

Guidelines for Assessment and Instruction in Statistics Education (GAISE) (2005). Retrieved from http://www.amstat.org/education/gaise.

Garfield, et al. (2012). *CATALST*. http://www.tc.umn.edu/~catalst/.

Lock, R., Lock, P.F., Lock-Morgan, K., Lock, E.F., & Lock, D.F. (2013) *Statistics: Unlocking the Power of Data*. John Wiley & Sons.

Ramirez C., Schau C., & Emmioglu, E. (2012). The Importance of Attitudes in Statistics Education. *Statistics Education Research Journal, 11*(2), 57-71.

Schau C., Miller M., & Petocz P. (2012). Special Issue: Research on Attitudes Towards Statistics: Editorial. *Statistics Education Research Journal*, *11*(2), 2-5.

Schau C., & Emmioglu, E. (2012). Do introductory statistics courses in the United States Improve Students' Attitudes? *Statistics Education Research Journal*, *11*(2), 86-94.

Schau, C. (2003). Survey of Attitudes Toward Statistics (SATS-36). [Online: http://evaluationandstatistics.com/] .

Scheaffer, R. (1997). Discussion to new pedagogy and new content: The case of statistics. *International Statistics Review*, *65*(2), 156-158. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/isr/97.Moore.pdf

Tintle, N. L. (2014, July). Quantitative evidence for the use of simulation and randomization in the introductory statistics course. *Proceedings of the International Conference on Teaching Statistics.* Flagstaff, Arizona.

Tintle, N. L., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2014). *Introduction to Statistical Investigations, Preliminary edition.* http://math.hope.edu/isi.

Tintle, N., Topliff, K., VanderStoep J., Holmes V-L. & Swanson T. (2012). Retention of Statistical Concepts in a Preliminary Randomization-Based Introductory Statistics Curriculum. *Statistics Education Research Journal, 11*(1), 21-40.

Tintle, N., VanderStoep, J., Holmes, V-L., Quisenberry, B. & Swanson, T. (2011). Development and Assessment of a Preliminary Randomization-Based Introductory Statistics Curriculum. *Journal of Statistics Education*, *19*(1).