7-2014

Quantitative Evidence for the Use of Simulation and Randomization in the Introductory Statistics Course

Nathan L. Tintle
Dordt College, nathan.tintle@dordt.edu

Ally Rogers
Dordt College

Beth Chance
California Polytechnic State University - San Luis Obispo

George Cobb
Mt. Holyoke College

Allan Rossman
California Polytechnic State University - San Luis Obispo

See next page for additional authors

Recommended Citation

In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute. iase-web.org [© 2014 ISI/IASE]

Quantitative Evidence for the Use of Simulation and Randomization in the Introductory Statistics Course

Abstract
The use of simulation and randomization in the introductory statistics course is gaining popularity, but what evidence is there that these approaches are improving students' conceptual understanding and attitudes as we hope? In this talk I will discuss evidence from early full-length versions of such a curriculum, covering issues such as (a) items and scales showing improved conceptual performance compared to traditional curriculum, (b) transferability of findings to different institutions, (c) retention of conceptual understanding post-course and (d) student attitudes. Along the way I will discuss a few areas in which students in both simulation/ randomization courses and the traditional course still perform poorly on standardized assessments.

Keywords
statistics course, student learning outcomes, simulation, randomization, pedagogical strategies, statistical reasoning

Disciplines
Curriculum and Instruction | Education | Statistics and Probability

Comments
Invited paper and slides from presentation at the 9th International Conference on Teaching Statistics, Flagstaff, Arizona, July 13-18, 2014.

Authors
Nathan L. Tintle, Ally Rogers, Beth Chance, George Cobb, Allan Rossman, Soma Roy, Todd Swanson, and Jill VanderStoep

QUANTITATIVE EVIDENCE FOR THE USE SIMULATION AND RANDOMIZATION IN THE INTRODUCTORY STATISTICS COURSE

Nathan L. Tintle[1], Ally Rogers[1], Beth Chance[2], George Cobb[3],
Allan Rossman[2], Soma Roy[2], Todd Swanson[4], Jill VanderStoep[4]
[1]Dordt College, Sioux Center, Iowa, USA
[2]California Polytechnic University, San Luis Obispo, California, USA
[3]Mt. Holyoke College, South Hadley, Massachusetts, USA
[4]Hope College, Holland, Michigan, USA
nathan.tintle@dordt.edu

The use of simulation and randomization in the introductory statistics course is gaining popularity, but what evidence is there that these approaches are improving students' conceptual understanding and attitudes as we hope? In this talk I will discuss evidence from early full-length versions of such a curriculum, covering issues such as (a) items and scales showing improved conceptual performance compared to traditional curriculum, (b) transferability of findings to different institutions, (c) retention of conceptual understanding post-course and (d) student attitudes. Along the way I will discuss a few areas in which students in both simulation/randomization courses and the traditional course still perform poorly on standardized assessments.

INTRODUCTION

While the use of simulation, bootstrapping and permutation tests (hereafter: randomization methods) in the practice of statistics have a longer history, substantial technological advances over the last three decades have led to the current, widespread use of these methods. In the realm of statistics education, increasing discussion has taken place with regards to the use of randomization methods to introduce students to the logic and scope of inference (Cobb, 2007). With this increased focus, more and more educators are considering the use of these methods in their courses, and numerous related curriculum projects are underway for the introductory statistics classroom (e.g., Garfield et al., 2012; Lock et al., 2013; Tintle et al., 2014).

Recently, numerous panels and presentations at statistics conferences have provided largely anecdotal support of the use of methods in the classroom reinforcing the initial claims made by Cobb (2007). In particular, arguments have been made that these approaches help students better understand the logic of inference (significance testing; interval estimation) through early introduction of inferential concepts via intuitive tactile and computer-based randomization techniques. Early introduction of these methods with students is facilitated by their intuitive nature requiring less formal training in probability and sampling distributions before they can be used by students. Furthermore, advocates of the use of randomization argue that student understanding of the scope of inference (generalizability and causation) can also be enhanced via these methods, due to the increased focus on connections between data production and data analysis.

Recently, two papers exploring students' growth in conceptual understanding and retention using an early version of a randomization curriculum yielded promising outcomes (Tintle et al., 2011; Tintle et al., 2012). In Tintle et al. (2011), the authors compare the post-course conceptual understanding of over 200 students (across 8 sections) of an algebra-based, undergraduate, introductory statistics course (Stat 101) after completing an early version of a randomization-based curriculum (an early version of Tintle et al. 2014). These students were compared to students at the same institution as well as a national sample (U.S.A.) who completed a traditional curriculum (normal theory approaches), on the 40-question, multiple choice CAOS test (delMas et al., 2007). Students showed significant improvement overall, and, in particular, with regards to their understanding of items related to tests of significance, data collection and design and simulation using the new curriculum as compared to students using the traditional curriculum at the same institution and the national sample. Furthermore, for almost all remaining items there was no significant change. One lone exception (an item on estimation of the standard deviation from histograms), which showed significantly worse performance with the new curriculum, led to a subsequent change to the curriculum. In sum, the authors argued that there was significant

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

improvement in the key areas anticipated by Cobb and others, with 'no harm done' in most other areas.

A subsequent paper re-assessed the same students (randomization and normal based course) four months after the course ended to assess retention (Tintle et al., 2012). The authors found significantly more retention of concepts related to tests of significance and study design with the randomization curriculum than the traditional curriculum, arguing that the potential improvements to students' conceptual understanding were not necessarily short-term gains, but were retained by students after the course ended better than they had before.

The promising findings of these initial papers lead to a host of subsequent questions. Perhaps two of the most important questions are:

1. As preliminary versions of randomization curricula mature, are conceptual learning gains maintained or, better yet, improved?
2. Are the findings transferable to institutions beyond the single institution described in the initial papers (Tintle et al., 2011, Tintle et al., 2012)?

In this paper we will consider these questions by presenting assessment data (a mix of CAOS and other multiple-choice questions) from the beginning and end of a full-semester implementation of a randomization curriculum. We will present data on (a) before and after implementation of such a curriculum (Tintle et al., 2014) at an additional institution and (b) assessment data at 11 institutions which used the curriculum during Fall 2013. Data on student attitudes is presented in a companion paper (Swanson et al., 2014).

METHODS

Assessment results are broken into two separate analyses.

Sample #1

In the first analysis, the conceptual understanding of statistics students at Dordt College are compared between a semester using a traditional approach textbook (Moore 2010; 94 students; spring 2011), and two semesters using the fall 2011/spring 2012 version of a randomization curriculum (current version is Tintle et al. 2014; 63 fall 2011 and 92 spring 2012; 155 total). Students completed the 40-question CAOS test during the first week of the semester and again during the last week of the semester. Students were given course credit for completing the assessment test, but not for their performance, and the test was administered electronically outside of class. One instructor was the same during all semesters, but the others differed between semesters.

Sample #2

In the second analysis, the conceptual understanding of statistics students in 17 sections of statistics, taught by 16 different instructors at 11 different institutions comprising a total sample of 454 students all using the fall 2013 of Tintle et al. (2014). Administration of the tests varied between instructors but was generally at or during the first week of for the pre-test and the week before or during finals week for the post-test. The assessment was a total of 30 questions including a mix of CAOS and other questions developed by our group.

RESULTS

Table 1 illustrates the pretest and posttest mean scores on the CAOS test for both cohorts. While significant improvement was see in both groups, the magnitude of improvement was approximately twice as large for the randomization curriculum. This difference in improvement was statistically significant (independent samples t-test; $p<0.001$) between the two cohorts, with an estimated difference of 6% (95% CI: 3.4% to 8.6%).

Table 1. Overall pre and post-course performance on CAOS

	Pretest Mean (SD)	Posttest Mean (SD)	Difference	Paired t-test p-value	Cohort effect p-value	95% CI for cohort
Randomization	44.9% (10.1%)	56.5% (11.6%)	11.6% (10.6%)	<0.001	<0.001	(3.4% to 8.6%)
Traditional	46.4% (9.3%)	52.0% (11.0%)	5.6% (9.9%)	<0.001		

Table 2 extends the analysis in Table 1, by illustrating the differences by cohort for nine of the CAOS subscales (see delMas et al. 2007 for further description).

Table 2. Pre and posttest performance by cohort on subscales of CAOS

Averages by Topic							
Subscale	Cohort	Pretest	Posttest	Difference	Paired t-test p-value	Cohort p-value	95% CI for cohort
Data Collection and Design	Randomization Consensus	34.8% 34.9%	53.1% 36.5%	18.2% 1.6%	<0.001 0.54	<0.001	(9.2%, 23.9%)
Descriptive Statistics	Randomization Consensus	55.1% 53.5%	61.1% 69.6%	6.0% 16.1%	0.015 <0.001	0.014	(-2.1%, -18.1%)
Graphical Representations	Randomization Consensus	55.8% 58.5%	64.4% 60.9%	8.6% 2.4%	<0.001 0.23	0.03	(0.6%, 11.4%)
Boxplots	Randomization Consensus	35.0% 32.4%	41.6% 34.1%	6.6% 1.6%	0.010 0.55	0.18	(-2.3%, 12.3%)
Bivariate Data	Randomization Consensus	58.1% 56.4%	60.7% 64.8%	2.6% 8.4%	0.28 0.005	0.12	(-13.3%, 1.6%)
Probability	Randomization Consensus	31.9% 32.4%	56.5% 35.2%	24.5% 2.7%	<0.001 0.52	<0.001	(10.8%, 32.7%)
Sampling Variability	Randomization Consensus	36.7% 38.7%	39.4% 43.5%	2.7% 4.8%	0.22 0.11	0.57	(-9.4%, 5.2%)
Confidence Intervals	Randomization Consensus	37.9% 42.9%	51.8% 47.8%	13.9% 4.9%	<0.001 0.12	0.026	(1.1%, 16.7%)
Tests of Significance	Randomization Consensus	46.1% 50.0%	70.0% 60.6%	23.9% 10.6%	0.000 <0.001	<0.001	(6.6%, 19.9%)

Of the nine subscales, six showed significantly different performance between the two cohorts, with five of the six subscales showing improvement (data collection and design, graphical representations, probability, confidence intervals and tests of significance). One of the six subscales (descriptive statistics) showed a significant decrease in performance.

When analyzing the fall 2013 sample of 454 students across multiple institutions, similar overall results were obtained (Pretest Mean: 44.9% (SD=10.6%), Posttest Mean: 54.0% (SD=13.1%), paired t-test, p<0.001). Furthermore (as shown in Table 3), on the seven subscales of the ISI assessment test, significant improvement was seen across each subscale except data collection and design. *Note: We are currently conducting analyses which will make the subscales in Tables 2 and 3 comparable.*

Table 3. Pre and posttest performance by cohort on subscales of ISI

Subscale	Cohort	Pretest	Posttest	Difference	Paired t-test p-value
Data Collection and Design	Randomization	61.1%	61.8%	0.7%	0.64
Descriptive Statistics	Randomization	31.0%	41.2%	10.2%	<0.001
Graphical Representations	Randomization	44.4%	52.9%	8.5%	<0.001
Probability	Randomization	28.9%	40.2%	11.3%	<0.001
Sampling Variability	Randomization	20.1%	29.0%	8.9%	<0.001
Confidence Intervals	Randomization	50.3%	60.8%	10.5%	<0.001
Tests of Significance	Randomization	55.5%	65.5%	10.0%	<0.001

CONCLUSION

Early papers showed promising results from the implementation of full-length randomization curricula. In this analysis we considered two analyses of later versions of the same curriculum. In the first analysis we saw similar results with a revised version of the curriculum at another institution. In particular, significant improvements as compared to the traditional curriculum were noted in numerous areas when using a randomization approach. In the second analysis we showed that in a more widespread implementation at multiple institution students showed improved conceptual understanding in most areas from the start of the course to the end.

While we are encouraged by the additional evidence about improvements of the randomization curriculum vs. the traditional approach, and from this preliminary evidence of the transferability of learning pre-post course learning gains to other institutions there are a number of limitations worth noting. First, because of the nature of the study design we can infer that it is, necessarily, the randomization curriculum that is causing the improvement. The pedagogical style, teacher attitude and other factors may also be contributing. However, seeing the results consistently occur at multiple institutions and over time is promising.

We continue to look at ways to improve the Tintle et al. (2014) version of the curriculum to improve student's conceptual understanding of statistics. In particular, due to the early weak performance of some students in areas related to histograms and standard deviation, we revised the curriculum to introduce these concepts differently. Further analysis is needed to fully evaluate the potential benefit of these changes. We also note that there are other areas in which both traditional and randomization curricula perform poorly. Further work is needed to better understand student performance in these areas and how curricula can be modified to improve student performance.

This analysis provides further evidence of the effectiveness of randomization and simulation approaches in improving students' conceptual understanding of concepts in introductory statistics. More in-depth analyses and studies are needed to pinpoint the aspects of such a curriculum that improve student learning, and to better understand student learning trajectories in such a course.

REFERENCES

Cobb, G.W. (2007). The introductory statistics course: a Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1).

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal 6*(2), 28-58.

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*. Wiley.

Swanson T., VanderStoep J., Chance, B., Cobb, G., Rossman, A., Roy, S., & Tintle N. L. (2014). *Student attitudes toward statistics from a randomization based curriculum*. Paper presented at the 9th International Conference on Teaching Statistics (ICOTS9).

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).