HOW IS SUCCESS IN STATISTICS MODERATED BY INSTRUCTIONAL PRACTICES USED AT THE UNIVERSITY LEVEL: INTERPRETING 30+ YEARS OF RESEARCH ON THE TOPIC

Hannah Wilson, Shawn Fitzgerald, Jian Li, and Jason Schenker
Kent State University, Kent Ohio, USA
hwilso20@kent.edu

Results from this study provide information regarding the effectiveness of various "innovative" instructional approaches employed in university level statistics courses. The average effect across all "innovative" instructional approaches, when compared to the traditional lecture approach, indicates that these strategies influence student achievement in a positive manner (d = 0.3389). However, this should be interpreted with caution as it was determined that several study characteristics moderated the relationship between instructional type and statistics achievement. Average effects of the innovative instructional practices on statistics achievement were moderated by type of publication (i.e., journal, presentation, dissertation), subject assignment to courses, (i.e., random, nonrandom) and length of implementation of innovative instructional practice. Implications for using various instructional approaches are discussed.

INTRODUCTION

In 1967, the Joint Committee of the American Statistical Association and the National Council of Teachers of Mathematics on the Curriculum in Statistics and Probability was formed to plan and coordinate improvements in the science and teaching of statistics and probability at all levels of education. Since this time the research on and innovations related to the teaching of statistics at the university level has advanced rapidly. As a result of the advances in this field, many mathematics education journals, such as the *Mathematics Teacher*, *Educational Studies in Mathematics*, and *The College Mathematics Journal*, have regularly published articles on the teaching of statistics at this level. Furthermore, journals such as *Teaching Statistics* and *The Journal of Statistics Education*, have evolved and are devoted specifically to publishing research on statistics education.

A search of several on-line research databases revealed over 500 articles and papers related to the teaching of statistics at the university level from 1965 to 2013. Most of this literature was found to be nonempirical. However, over 200 studies were empirical in nature with a large number focusing on the effectiveness of different teaching methods and teaching aids employed in the statistics classroom. Among the literature focusing on statistics instruction, it appears that there is growing support among those teaching statistics and those conducting research on the teaching of statistics for the increased use of innovative methods of instruction in the statistics classroom.

The innovative methods which have been incorporated into the statistics classroom over the years include personalized systems of instruction, programmed instruction, the use of small groups in cooperative learning environments, as well as the use of computers as both a tool for delivering instruction and as a teaching aid. In addition, statistics laboratories have been used to supplement instruction in order to give students experience with analyzing data using various statistical packages and conducting other hands-on activities related to data analysis and interpretation. Other teaching aids which have been "experimented" with in the class have included the use of writing assignments, the use of humor, the use of analogies and metaphors, and the use of exercises and assignments that are oriented toward the students' field of interest. While this description is not inclusive of all the various instructional methods and teaching aids which have been explored in the statistics classroom, it does indicate the diverse nature of what has been explored in an attempt to improve statistics instruction and education since the 1960's.

In general, the rationale for incorporating these innovative methods into the statistics classroom is that students will likely understand statistical concepts more easily; will be able to deal with more complex data; and will be more motivated to learn and ultimately be able to attain a higher level of achievement in a statistics course than they would had they been exposed to the traditional lecture method of instruction. Unfortunately, among the studies which have focused on the effectiveness of these innovative instructional methods of teaching statistics, few have

produced consistent results when compared to the traditional lecture approach. This finding is evident even when considering those studies focusing on the same method of instruction.

METHODS

The purpose of this study was to determine the overall effectiveness of various instructional interventions employed in the statistics classroom at the university level and to determine if effectiveness was related to characteristics of the research study or to outcome measures used. In this research, individual studies on the effectiveness of instructional interventions employed in the statistics classroom were combined to help determine the general effectiveness of instructional interventions employed in different situations.

Studies for inclusion in this analysis addressed some form of instructional intervention used in the statistics classroom at the university level. Only those studies which clearly defined an intervention as a condition being examined were included. Those designs that did not include a control group or a comparison group were not included for analysis. Control group designs included random assignment to treatment conditions whereas comparison group designs used some form of group assignment based on comparative samples. Those studies that did not include adequate statistics for transformation into effect sizes were excluded. Necessary statistics for these transformations included means and standard deviations or a variety of parametric test statistics. Conversions of effect sizes from parametric statistics are outlined in several texts (e.g., Glass et al., 1981). Effect sizes associated with achievement outcomes in the obtained studies were the dependent variables in this meta-analysis. The effect sizes were standardized measures of change associated with each study.

Locating Documents

Several searches were completed to identify relevant studies for this meta-analysis. First, searches were conducted using several on-line retrieval systems including ERIC, psych-INFO, and Dissertation Abstracts. In addition, the conference proceedings of the International Conference on the Teaching of Statistics (ICOTS) and the proceedings of the section on statistics education of the American Statistical Association (ASA) were also searched. In addition to these searches, the bibliographies of studies included based on the searches of the on-line databases and the conference proceedings were also searched as another source of articles or documents. All searches were conducted using specific time constraints (i.e., 1965-2013) and key descriptors. These descriptors included: TEACHING, INSTRUCTION, STATISTICS, and COLLEGE STUDENTS. The combined searches yielded 134 studies which were considered. After obtaining the studies and reviewing their contents only 70 could be included in the final analysis. Most of the studies that were excluded were either anecdotal in nature or did not include sufficient statistical data to compute an effect size.

Coding Study Characteristics

The 70 studies located for use in this synthesis described two major types of instructional interventions: technology based active learning strategies and non-technology based active learning strategies.

Ratings of Study Quality

This meta-analysis included steps to control for the effects of poorly designed or executed studies as outlined by Wortman (1994). The first step was to strictly control the studies included in the current research. Only studies that clearly stated the use of an instructional intervention and use a control or comparison group were included. Each study result was coded to specify if it was the result of a randomly assigned control group or a non-equivalent comparison group. All of these sources were then tested in the analysis to determine if bias was present according to the coded characteristics of study quality.

Analysis

Both descriptive and inferential statistical analyses were completed to address the research questions of this study. This analysis determined 1) the overall effect size for the dependent

variable, and 2) investigated the influence of study characteristics on outcome effect sizes. Practical Meta-Analysis Effect Size Calculator (Lipsey & Wilson, 2001) and the computer package of HLM 7.0 (HML: Raudenbush et al, 2004) were used to analyze the data.

RESULTS

As each study could use a different outcome variable to measure students' success in statistics, we calculated the standardized effect size (i.e., Cohen's d) for the studies. The standardized effect size was calculated by subtracting the mean of the comparison or control group from the mean of the treatment group and dividing by the standard deviation of the control group, or the pooled standard deviation. Using standardized measure of effect size converted each study's results into a common scale; as a result, all the studies became comparable. At the same time, studies generally vary in the sample size used. Studies with larger sample size were assumed to be more precise estimates of the population effect size than studies with smaller sample size (Lipsey & Wilson, 2001). Therefore, we allowed the studies with larger sample size to carry more weights in the meta-analysis by weighting the outcome variable by the inverse of the variance of the effect size. In addition, the Cohen's d effect size has been reported to be a biased estimator, especially when computed using studies having small sample sizes, and was adjusted for unbiased estimates using the formula provided by Hedges (1981).

The research questions were analyzed using the variance-known application of hierarchical linear modeling, as the sampling variances of the effect size estimates can be assumed to be known (Raudenbush & Bryk, 2002). The criterion variable for all analyses was the effect sizes obtained in the studies. A total of 85 effect sizes were obtained from the 70 individual articles. The predictor variables were dummy-coded so that the effect of each individual coding feature could be isolated. Specifically, the predictors included length of implementation (coded as "0" if the course was offered shorter than a semester and "1" if offered as a full semester class), instructor bias ("0" if different instructors were involved in the instruction and "1" if the same instructor taught the whole course), course level ("0" if the course is introductory statistics and "1" if the course is intermediate or advanced statistics), academic standing ("0" if the course was offered to undergraduate students and "1" if offered to graduate students or both), research design ("0" if the study used a quasi-experimental design and "1" if used a true experimental design), and a series of dummy coded variables for disciplines in which the study was conducted.

A two level hierarchical linear model was constructed to address the data structure in the meta-analysis (Raudenbush & Bryk, 2002). The level-1, within-site (study) model specifies the observed effect size as a function of the true effect size and sampling error. The level-2, between-sites (studies) model specifies the distribution of the true effect sizes as a function of study characteristics and random error (Raudenbush & Bryk, 2002). The analysis was conducted by using a two-step procedure. First, an unconditional model (i.e., no predictors were added in the model) was fitted to estimate the average effect size across the different studies and the variability of the true-effect sizes. Second, a conditional model in which the predictors were added was fitted to estimate if the study characteristics would help explain the variability in the true-effect sizes.

By fitting the first unconditional model, we found that the estimated grand-mean effect size was moderate with a value of 0.3866, implying that, on average, students in the treatment group scored about 0.39 standard deviation units above the students in the control or comparison group. However, this average group effect was only marginally significant (t (85) = 1.786, $p = 0.078$). The estimated variance of the effect-size parameter was 0.1764 and was statistical significant (χ^2 (85) = 473.26, $p < 0.001$), indicating that a significant amount of variability still existed in the true-effect sizes. Next, we ran the conditional model by adding predictors into the analysis; unfortunately, none of the study level characteristics were significant predictors. The treatment effects were not statistically significant although its magnitude was moderate to high with a value of 0.4773 (t (85) = 0.605, $p > 0.05$) after controlling the study characteristics. The variance of the treatment effects between studies after controlling the study characteristics was still significant with a value of 0.1654 (χ^2 (74) = 469.11, $p < 0.001$), indicating there may exist some other study level characteristics that could significantly explain the between study variations, but were not provided in the original studies, therefore were not included in the analysis.

CONCLUSION

Results from this study provide information regarding the effectiveness of various "innovative" instructional approaches employed in university level statistics courses. The average effect across all "innovative" instructional approaches, when compared to the traditional lecture approach, indicates that these strategies influence student achievement in a positive manner (d = 0.3866). However, this average treatment effect of the "innovative" instructional approaches is not statistically significant if the between-study variation is taken into consideration and it should be interpreted with caution.

The findings also indicated that the relationship between instructional type and statistics achievement could be positively moderated by the research design, instructor bias, and course level. If holding all the other variables constant, the effects for those studies which assigned subjects via random assignment produced larger effects on average (d = 0.5665) than those which used intact groups (d = 0.4773); those studies in which a same instructor taught the whole course generated larger effects on average (d = 0.5773) than those in which different instructors provided the instruction (d = 0.4773); and those studies which focused on an intermediate or advanced statistics course produced larger effects (d = 0.5933) than those which focused on an introductory statistics course (d = 0.4773). Although these moderation effects were not statistically significant, their effects had practical significance and meaningfulness as the average treatment effect became nontrivial when these study characteristics were allowed to moderate the relationship between instructional type and statistics achievement.

REFERENCES
Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.
Lipsey, M. W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
Norušis, M. J., & Inc, S. (1994). *SPSS professional statistics 6.1*: Prentice Hall.
Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.
Wortman, P. M. (1994). Judging research quality. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97-110). New York: Russell Sage Foundation.