Running head: APPLICATION OF STATISTICAL METHODS

The Application of Statistical Methods

in a Study of Technology Integration

Rod Myers

Indiana University

Y502 – Intermediate Statistics Applied to Education

Professor Kelley

November 18, 2007

The Application of Statistical Methods in a Study of Technology Integration

This paper discusses the application of statistical methods in a study of technology integration in secondary education (Boon, Burke, Fore, & Spencer, 2006). The purpose of the study was to determine whether the use of specialized software (Inspiration 6) to create cognitive organizers increased the learning of students as compared with traditional textbook instruction. The study was conducted with secondary students, nearly half of whom had been identified as learning-disabled, in two inclusive social studies classrooms in a suburban high school in a large metropolitan area of the southeastern United States.

The authors used a quasi-experimental pretest/posttest group design. In the abstract, the authors claim that the students were randomly assigned to control and treatment groups. However, it seems that the students were previously assigned to classrooms and that the selection of classes for treatment and control was random. Regardless, the selection of students was not truly random but was based on convenience. While the class sizes were comparable (treatment $n = 25$, control $n = 24$), the treatment group had over twice as many students with learning disabilities (treatment $n = 14$, control $n = 6$). This method of participant selection is problematic and is addressed below in the discussion of statistical methods and assumptions.

Both groups studied the same chapter of a tenth-grade textbook on world history over the course of four 90-minute periods in three weeks. Both groups were introduced to the chapter and then took a 35-item pretest of the content material. The students in the treatment group were provided a paper cognitive organizer and followed along as the teacher presented the content and completed the organizer. When the presentation of the chapter material was completed, the students convened in the computer lab to copy their notes into an outline template in Inspiration 6. They printed their organizer in both outline and diagram format, which they used for review.

The teacher also presented the content to students in the control group, and the students read aloud and discussed sections of the textbook. They also completed a guided reading worksheet, participated in a cooperative learning activity, and watched a video related to the chapter. At the end of the final period, both groups completed a posttest that was identical to the pretest.

Grading of pre/posttests was subjective, with answers to items being awarded from zero to two points. The researcher and a graduate student conducted reliability checks of the grading, and discrepancies were discussed and resolved to reach complete agreement. This method of obtaining reliability is questionable because both the researcher and the graduate student were familiar with the research. Independent graders who were unaware of the study should have been used instead.

Statistical Methods and Assumptions

The authors used descriptive statistics and analysis of variance (ANOVA) to compare the pre/posttest results for the two groups. There are three underlying assumptions for the analysis of variance: normality, homogeneity of variance (homoscedasticity), and independence of observations. The authors do not address the assumptions of the method, and they do not provide the results of Levene's test for equality of variance. The first two assumptions may be violated to some degree without seriously affecting the results, especially given similar sample sizes. However, because the participants were not randomly selected and received instruction within their groups, independence of observations has been violated. This makes the calculation of sums of squares within groups particularly questionable and affects the calculation of mean squares, the F statistic, and the probability of the results being statistically significant. In order to use ANOVA properly, the study should have involved many more classrooms that were independent of each other.

Results and Conclusions

The study found no significant difference between the groups on the results of the pretest, $F(1, 48) = .216$, $p = .644$. The treatment group had a mean pretest of 11.60 ($SD = 12.708$) and the control group had a mean pretest of 13.08 ($SD = 9.604$). However, the posttest scores showed significant differences, $F(1, 48) = 41.44$, $p < .01$; the treatment group had a mean posttest of 52.52 (reported as 52.54 in the text) ($SD = 13.305$) and the control group had a mean posttest of 26.84 ($SD = 14.860$). The descriptive statistics also included the minimum and maximum scores, with the posttest treatment group achieving a maximum of 81. This is curious because the instrument was described as a 35-item assessment with a maximum of two points per item, which results in a maximum of 70 possible points.

The authors concluded that the use of the cognitive organizer had a significant positive effect on the learning gains of the treatment group between pretest and posttest, and that the use of the cognitive organizer was more beneficial than the traditional textbook instruction. They noted several limitations of the study which were generally related to the specificity of the content area and grade level. They said that the relatively small sample size did not allow them to analyze separately the results for the students with learning disabilities. This is unfortunate because, as noted earlier, the treatment group had over twice as many students with learning disabilities. It is possible that the use of the cognitive organizer was more beneficial for those students and affected the overall results. Nevertheless, the authors stated that “the use of cognitive organizers has the potential to significantly improve content-area learning for both students with and without disabilities in content-area classrooms” (p. 9).

There also may have been a novelty effect from using the cognitive organizer. Students from both groups surely interacted with each other and were aware of the differences between their classes. The students in the treatment group may have tried harder because they were doing something new and thought they were expected to do better. The interaction between groups also invalidates the assumption of independence of observations.

It should also be noted that the pre/posttest was for declarative knowledge, the lowest level of Bloom's taxonomy of educational objectives in the cognitive domain. It may be that the use of the cognitive organizer improved the students' ability to recall facts but inhibited their acquisition of higher-order skills, such as the ability to apply their new knowledge, to integrate it with their existing knowledge, to analyze it and make inferences, and to synthesize it and draw new conclusions. Testing high school students' ability to memorize facts is setting expectations unreasonably low.

While the authors of this study had good intentions, they were less than rigorous in designing, implementing, and evaluating their research. In particular, they violated the assumption of independence of observations in several ways, and this makes their results questionable. In their discussion of the limitations of the study, they failed to note this and several other factors that may have influenced the results. While cognitive organizers may make a greater contribution to learning than traditional textbook instruction, this study fails to make a convincing case.

References

Boon, R. T., Burke, M. D., Fore, C., III, Spencer, V. G. (2006). The impact of cognitive organizers and technology-based practices on student success in secondary social studies classrooms. *Journal of Special Education Technology*, *21*(1), 5-15.