

What's the Difference, Still?:

A Follow-Up Review of the Quantitative Research Methodology in Distance Learning

Justus Randolph

University of Joensuu, Department of Computer Science

Paper presented at the Joensuu Learning and Instruction Symposium, October 14th and 15th, 2005

Joensuu, Finland

Abstract

A high quality review of the distance learning literature from 1992-1999 concluded that most of the research on distance learning had serious methodological flaws. This paper presents the results of a small-scale replication of that review. From three leading distance education journals, a sample of 66 articles was categorized by study type and the experimental or quasi-experimental articles were analyzed in terms of their research methodologies. The results indicate that the sample of post-1999 articles had the same methodological flaws as the sample of pre-1999 articles.

What's the Difference, Still?:

A Follow-Up Review of the Quantitative Research Methodology in Distance Learning

In April of 1999, The Institute for Higher Education Policy released an influential review of the distance learning literature entitled, *What's the Difference?: A Review of Contemporary Research on the Effectiveness on Distance Learning in Higher Education*. The report, based on a large sample of the distance learning literature, concluded that although a considerable amount of research on the effectiveness of distance learning has been conducted, "there is a relative paucity of true, original research dedicated to explaining or predicting phenomena related to distance learning" (p.2). Although many of the studies included in *What's the Difference?* suggested that distance learning compares favorably with classroom based instruction (Russell, 1999; see also Hammond, 1997; Martin & Rainey, 1993; Sounder, 1993), a closer investigation by the authors of *What's the Difference?* revealed that the quality of those studies was questionable and that the results of the body of the literature on distance learning was largely inconclusive.

What's the Difference? reported four main shortcomings in the research on distance learning:

1. Much of the research does not control for extraneous variables and therefore cannot show cause and effect.

2. Most of the studies do not use randomly selected subjects.
3. The validity and reliability of the instruments used to measure student outcomes and attitudes are questionable.
4. Many studies do not adequately control for the feelings and attitudes of the students and faculty - what the educational research refers to as "reactive effects." (pp. 3-4)

Extraneous variables, poor validity and reliability of measures, and reactive effects, alone or in combination are enough to undermine the validity of a generalized causal inference. Since the authors of *What's the Difference?* found that the majority of research on distance learning contained these shortcomings, it follows that the majority of distance learning research was also inadequate to make sound conclusions about the actual effects that distance learning has on academic achievement and student satisfaction.

Given the exponential growth of distance learning programs (see Conhaim, 2003; Imel, 2002; Salomon, 2004) and the potential consequences of imprudent policy decisions concerning distance education (see Kelly, 2002; "Pros and Cons of E-Learning," 2002), it would be logical to presume that the distance learning research community would have taken heed of the suggestions for improving the methodology reported in *What's the Difference?*. That presumption is investigated here by reviewing a small sample of the distance learning research where *What's the Difference?* left off. Specifically, the current review examines the distribution, by type of study, of English language articles that have been recently published in three leading

distance education journals and analyzes the research methodology of the quantitative experimental or quasi-experimental articles in those journals.

Method

This section reports the method used to replicate a previous review of the research on distance learning published from 1992-1999. Articles published after 1999 from a sample of journals used in *What's the Difference?* were categorized by type of study; the experimental or quasi-experimental articles were critically analyzed in terms of their research methodology.

The Sample

Of the five journals included in *What's the Difference?*, a purposive sample of three leading distance education journals, *The American Journal of Distance Education*, *Distance Education*, and *The Journal of Distance Education*, was included in the current review. All of the articles from these journals, besides book reviews, forewords, and editorials, were included in this review if they were written in English. See Table 1 for more information about the origins, number of articles, and time periods of the sample of articles used in the current review.

[Insert Table 1 here.]

Categorization of Articles

All of the articles from the sample mentioned above were divided into six categories. The categories were a) qualitative articles b) quantitative descriptive articles, c) correlational articles, d) quasi-experimental articles, e) experimental articles, and f) other types of articles.

Qualitative articles used qualitative methodologies exclusively. Quantitative descriptive articles described the characteristics of a group of students on one or more variables

and were usually measured by self-report surveys. (One-group posttest-only designs were classified as descriptive studies.) Correlational articles were defined as articles that examined the relationship between two variables without establishing causality. Experimental articles were defined as articles with randomly assigned participants that investigated the effects of distance learning on academic achievement or student satisfaction. Quasi-experimental articles were defined the same way as experimental research articles except that participant assignment was not random. The 'other' category of articles consisted of reviews of literature, meta-analyses, program descriptions, theoretical articles, project management guidelines, or fictional cases.

Critique of Articles

The studies that used quantitative experimental or quasi-experimental research designs with a form of distance education as the independent variable and at least one measure of academic achievement or student satisfaction were analyzed in terms of the shortcomings found in *What's the Difference?*. The method for evaluating the scientific control of extraneous variables was to identify the research design and its weaknesses in terms of Shadish, Cook, and Campbell's (2002) descriptions of threats to internal validity. The text was then scanned to determine if the extraneous variables inherent in the design were reportedly controlled for. The text was also reviewed to determine if participants were randomly selected and assigned, if the author(s) reported evidence about the instrument's reliability and validity, and if reactive effects, specifically novelty and the John Henry effect, were controlled for.

Results

For the quantitative experimental and quasi-experimental studies, the research design, experimental controls, selection, assignment, and reliability and validity of instruments are presented. The results also include the numbers of articles distributed into each category.

Distribution of Articles by Type

From the 3 journals sampled, 66 articles were reviewed. Of these, 18 were categorized as qualitative, 12 as quantitative descriptive, 8 as correlational, 4 as quasi-experimental, 0 as experimental, and 24 were categorized as 'other' which included reviews of literature, meta-analyses, program descriptions, theoretical articles, project management guidelines, or fictional cases. See Table 2 for the distribution of articles by study type.

[Insert Table 2 here.]

Results of the Article Critique

Since only four studies were classified as quasi-experimental and none were categorized as experimental, the results of the article critique are reported here on a study-by-study basis. Results include a description of the methodology and the threats to validity in each study.

Bisciglia and Monk-Turner's study. Bisciglia and Monk-Turner (2002) examined the effect of distance learning on reported

attitudes toward distance learning. They used a posttest-only design with a nonequivalent control group. Participants in the treatment group were offsite while control group participants were onsite. The same instructor taught both groups at the same time but the groups were at different locations. Intact groups were randomly selected from the population of local distance learning courses being conducted at the time; however, of the groups selected, only 38% of the teachers agreed to let their classes participate in the study. Students self-selected to participate either on-site or on distant sites. The instruments were self-report surveys without reliability or validity information.

Major threats to validity in the Bisciglia and Monk-Turner study were selection and the construct validity of the control condition. Although there was an attempt at randomly selecting classes, only a small percentage of teachers who were selected volunteered to participate. Students self-selected not only which class they would be in, they also selected which experimental condition they would be in. Demographic variables were taken as an attempt to measure the prior differences between both groups, yet this does not completely control for selection since there were other variables related to outcomes (i.e., prior knowledge of subject and motivation) not measured by the demographic variables. In fact, on several important

variables, (e.g., prior experience with distance education, gender, hours at work, and marital status) the control and treatment groups differed markedly.

Concerning the construct validity of the control condition, usually the comparison in distance education involves distance education programs versus traditional programs. However, in this study the comparison involved onsite distance education versus offsite distance education. Onsite distance education courses, although they are conducted face-to-face, are quite different than traditionally administered courses and, therefore, do not represent the control condition of most interest (i.e., the traditional classroom instruction.) Onsite students have to deal with many of the pedagogical disadvantages of distance learning, (e.g., waiting in an electronic queue to interact verbally) and have more problems with instructor accessibility than offsite students (Phillips & Peters, 1999). However, onsite students do not receive the same benefits related to distance learning as offsite students (e.g., not having to relocate or commute to the physical site of instruction.)

Kennepohl's study. Kennepohl (2001) examined the use of computer simulations on university-level students' performance in a chemistry lab. The investigator used a posttest only design with a nonequivalent control group. The control group conducted the usual laboratory exercises for 32 hours. The treatment group

conducted 4 to 8 hours of simulations before 24 hours of the usual laboratory exercises. No information was given about selection or assignment; however, the text implies that the groups were intact and that the experimenter decided which intact group would be the treatment group and which would be the control group. Instruments used were teacher-made lab quizzes and teacher-assignment of course and lab grades.

Major threats to validity were selection and instrumentation. Selection was problematic because it was probable that groups were not equivalent before application of the treatment. For example, one group may simply have been higher achievers than the other group. This was especially problematic if the experimenter had assigned participants to conditions based on his or her prior knowledge of group performance. Instrumentation was a problem if the researcher either knowingly or unknowingly assigned grades and scores influenced by the knowledge of which group the student was in. Reliability and validity of measures were not reported. Other threats such as attrition and reactive effects may have been possible since little description of the participants, procedure, and setting was provided.

Litchfield, Oakland, and Anderson's study. Litchfield, Oakland, and Anderson (2002) examined the effect of computer mediated learning on computer attitudes. An untreated control group design with dependent pretest and posttest samples was used with

adult dietetic students. Students were not reported to be randomly selected or assigned. The instrument was a self-report survey. No validity or reliability information was reported.

This relatively strong design used in the Litchfield et al. study helped ruled out most major threats; therefore, there were only minor plausible threats. Since a pretest and demographic data were used to compare the groups before treatment, this helped control the selection threat. While, the researchers reported the overall change between pretest and posttest for each group, they did not report initial pretest results for each group. Little information was provided about the reliability or validity of measures and about procedures pertinent to reactive or other effects.

Neuhauser's study. Neuhauser (2002) investigated the effect of computer mediated learning, with learning style as a moderating variable, on the effectiveness of learning and student satisfaction with adults studying business management. The investigator used a posttest-only design with a nonequivalent control group. Students in the experimental condition received instruction via computer-mediated learning. Students in the control group received face-to-face instruction and used e-mail to correspond about issues concerning evaluations, review, and reflection. Students were not randomly selected or assigned; however, the demographic characteristics of each group were

reported. The measurements, without reports of validity or reliability, were self-report surveys, teacher-made tests, and grades given by the teacher.

The major validity threat in the Neuhauser study was selection. Selection was probable since students self-selected into treatment conditions. Although, the demographic characteristics of each group were approximately equal, there may have been some factors related to outcomes that were not measured through demographics alone (e.g., prior knowledge of course content). It is difficult to determine to what degree reactive threats affected the study outcomes because little information was given about settings and circumstances. Attrition was addressed in the Neuhauser study by reporting the number and characteristics of students who quit attending the course in each group.

Discussion

In this section, findings from the four quasi-experimental studies and the distribution of articles by study type are discussed in terms of the criticisms found in *What's the Difference*. In short, the methodology flaws in distance learning research before the 1999 publication of *What's the Difference* are still present in distance learning research after 1999.

A Paucity of Original Quality 'Quantitative' Research, Still

In terms of quantitative designs, although descriptive and correlation research certainly is of significant value, only experimental and quasi-experimental research is appropriate for

establishing causal links between treatments and outcomes (Shadish et al., 2002). Of the 66 articles included in the current review, only 4 used quasi-experimental designs and 0 used experimental designs. Therefore, it is still appropriate to conclude that there is a paucity of quality quantitative research that investigates the link between distance learning and academic achievement or student satisfaction.

Poor Control of Extraneous Variables, Still

The posttest-only design with nonequivalent controls, which was used in 3 out of 4 studies reviewed here, leaves a host of extraneous variables uncontrolled. This design is especially open to selection and selection-interaction threats to internal validity. Although attempts were made to measure selection threats by comparing demographic data, this still is inadequate in most cases because a proxy test may not measure the factors that are most related to the outcomes. Only one study (Litchfield et al., 2002) used a design strong enough to control for most extraneous variables. Poor description of procedures and settings in these research reports, overall, do not inspire confidence that other validity threats have been controlled for.

Lack of Randomized Selection, Still

None of the studies reviewed here used random selection. This severely limits causal generalization and violates the assumptions of many statistical procedures. More troubling, however, is that none of the studies used random assignment. Although random assignment of participants cannot ensure the elimination of threats, it increases the likelihood of making correct causal assumptions. When randomized assignment is not feasible, strong designs and thoughtful control of variables can allow a researcher to make cogent arguments about general causality between independent and dependent variables.

Questionable Validity and Reliability of Instruments, Still

None of the studies analyzed here reported convincing information about the validity and reliability of instruments. Either the instruments were a self-report Likert-type survey that is subject to strong reactive effects or they were teachers who gave grades based on teacher-made tests and quizzes, among other factors. Much work must be still done on creating, researching, and reporting the validity and reliability of instruments in the distance education literature.

Inadequate Controls for Reactive Effects, Still

None of the articles directly addressed how they controlled for reactive effects such as novelty effects. Likewise, none of the articles gave enough information to determine to what degree the John Henry effect was present and how it was controlled.

Conclusion

Based on the sample reviewed here, the same shortcomings in the distance learning literature mentioned in *What's the Difference?* are still present. More research that uses strong designs which control for extraneous variables and reactive effects and that uses instruments which are proven to be valid and reliable is sorely needed in the research on distance learning. Until that point, we will just have to keep wondering, "What's the difference?"

References

- Bisciglia, M. G., & Monk-Turner, E. (2002). Differences in attitudes between on-site and distance-site students in group teleconference courses. *The American Journal of Distance Education, 16*(1), 37-52.
- Conhaim, W. W. (2003, December). Education ain't what it used to be. *Information Today, 20*, 37-38.
- Hammond, R. J. (1997, August). *A comparison of the learning experience of telecourse students in community and day sections*. Paper presented at the meeting of the Distance Learning Symposium sponsored by Utah Valley State College, Orem, UT.
- Imel, S. (2002). *E-Learning. Trends and issues alert* (Report No-40). Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education. (ERIC Document Reproduction Service No. ED469265)
- Kelly, M. F. (2002). The political implications of e-learning. *Higher Education in Europe, 27*(3), 211-216.
- Kennepohl, D. (2001). Using computer simulations to supplement teaching laboratories in chemistry for distance delivery. *The Journal of Distance Education, 16*(2). Retrieved April 21, 1976, from <http://cade.athabasca.ca/vol16.2/kennepohl.html>
- Litchfield, R. E., Oakland, M. J., & Anderson, J. A. (2002). Relationship between intern characteristics, computer attitudes, and use of online instruction in a dietic training program. *The American Journal of Distance Education, 16*(1), 23-36.
- Martin, E. D., & Rainey, L. (1993). Student achievement and attitude in a satellite-delivered high school science course. *The American Journal of Distance Education, 7*(1), 54-61.
- Neuhauser, C. (2002). Learning style and effectiveness of online and face-to-face instruction. *The American Journal of Distance Education, 16*(2), 99-113.

- Phillips, M. R., & Peters, M. J. (1999). Targeting rural students with distance learning courses: A comparative study of determinant attributes and satisfaction levels. *Journal of Education for Business*, 74(6), 351-356.
- The pros and cons of e-learning. (2002, December). *School Planning and Management*, 41(12), 9.
- Russell, T. L. (1999). *The no significant difference phenomenon*. Chapel Hill, NC: Office of Instructional Telecommunications, North Carolina State University.
- Salomon, K. (2004, February). Distance ed: The return of the DEPD. *University Business*, 7, 24-25.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.
- Souder, W. E. (1993). The effectiveness of traditional versus satellite delivery in the three management of technology Master's degree programs. *The American Journal of Distance Education*, 7(1), 37-53.
- The Institute for Higher Education Policy (1999, April). *What's the difference: A review of the contemporary research on the effectiveness of distance learning in higher education*. Retrieved April 21, 2004 from <http://www.nea.org/he/abouthe/diseddif.pdf>

Table 1

Origin, Time Period, and Quantity of Articles Included in Review

| Journal title | Volume/issue range | Year(s) | # of articles |
|---|--------------------|-----------|---------------|
| <i>The American Journal of Distance Education</i> | V. 16.1 – 16.4 | 2002 | 12 |
| <i>Distance Education</i> | V. 23.1 - 23.2 | 2002 | 14 |
| <i>The Journal of Distance Education</i> | V. 15.1 – 18.1 | 2002-2003 | 40 |

Table 2

Distribution of Types of Articles Included in the Review

| Type of article | Number of articles | Percent |
|--------------------------|--------------------|---------|
| Qualitative | 18 | 27.3 |
| Quantitative descriptive | 12 | 18.2 |
| Correlational | 8 | 12.1 |
| Quasi-experimental | 4 | 6.0 |
| Experimental | 0 | 0.0 |
| Other ^a | 24 | 36.4 |
| Total | 66 | 100.0 |

^a The 'other' category includes reviews of literature, meta-analyses, program descriptions, theoretical articles, project management guidelines, or fictional cases.