

What's the Difference, Still? A Follow up Methodological Review of the Distance Education Research

Justus J. RANDOLPH

*University of Joensuu, Department of Computer Science
P.O. Box 111, FIN-80110, Joensuu, Finland
e-mail: justus.randolph@cs.joensuu.fi*

Received: December 2005

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. A sample of 66 articles was drawn from three leading distance education journals. Those articles were categorized by study type, and the experimental or quasi-experimental articles were analyzed in terms of their research methodologies. The results indicated that the sample of post-1999 articles had the same methodological flaws as the sample of pre-1999 articles: most participants were not randomly selected, extraneous variables and reactive effects were not controlled for, and the validity and reliability of measures were not reported.

Key words: distance education, methodological review, research methodology.

1. Introduction

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* [hereafter – *What's the Difference*] (Phipps and Merisotis). That review, which was 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 and Rainey, 1993; Souder, 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 were 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 or 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 causal conclusions about the 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. Also, the research methodologies of the quantitative experimental or quasi-experimental articles in those journals are analyzed in detail.

2. Methods

This section reports the method used to replicate *What's the Difference*. In short, 66 recently published articles from a sample of journals used in *What's the Difference* were categorized by study type, and the experimental or quasi-experimental articles were critically analyzed in terms of the research methods used.

2.1. 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 chosen for the current review. These journals were chosen because they were assumed to be representative of typical research in the field of distance education. All of the articles from these journals; besides book reviews, forewords, editorials, and articles not written in English; were included in the current review. 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. None of the issues in this sample were special issues.

Table 1
Origin, time period, and quantity of articles included in the current 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 |

2.2. Categorization of Articles

The articles from the sample mentioned above were divided into six categories. The categories were (1) *qualitative articles* (2) *quantitative descriptive articles*, (3) *correlational articles*, (4) *quasi-experimental articles*, (5) *experimental articles*, and (6) *other types of articles*.

In the current review, *qualitative articles* reported on investigations that used qualitative approaches. *Quantitative descriptive articles* described the characteristics of a group of students on one or more variables. (One-group posttest-only designs were classified as descriptive studies.) *Correlational articles* examined the association between two continuous variables. *Experimental articles* investigated the effects of distance learning on academic achievement or student satisfaction and used random assignment to control and treatment conditions. *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.

The majority of categories in the current review corresponded with the categories in *What's the Difference*. What were called *qualitative articles*, *quantitative descriptive articles*, and *correlational articles* in the current review corresponded with what were called *case studies*, *descriptive articles*, and *correlational articles*, respectively, in *What's the Difference*. There were two differences though between the categories in the current review and the categories in *What's the Difference*. First, in the current review, a distinction was made between quasi-experimental and experimental research. What were called *quasi-experimental articles* and *experimental articles* in the current review would have simply been called *experimental articles* in *What's the Difference*. Second, in the current review, an *other* category was included to account for studies that did not fall into any of the categories above.

2.3. 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 then, by using Shadish, Cook, and

Campbell's (2002) description of threats to internal validity, determine what extraneous variables need to be controlled for when a particular design is used. The text was scanned to determine if the relevant extraneous variables were controlled for. The text was also reviewed to determine if participants were randomly selected and randomly 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.

3. 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 number of articles distributed into each category.

3.1. *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.' See Table 2 for the distribution of articles by study type. In order to compare proportions of article types between the current review and the previous review (i.e., *What's the Difference*), Table 3 shows the results of the current review when the *other* category is removed and when the *quasi-experimental* and *experimental* categories are collapsed into a single *experimental* category.

3.2. *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

Table 2
Distribution of types of articles included in the current 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.

Table 3

Comparison of proportions of article types between the current review and the previous review

| Type of article | Current review | Previous review |
|-----------------|----------------|-----------------|
| Qualitative | 43% | 15% |
| Descriptive | 29% | 31% |
| Correlational | 19% | 3% |
| Experimental | 9% | 51% |

basis. These include a description of the methodology and the threats to validity in each study.

3.2.1. *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; participants in the control group 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 into either on-site or distant sites. The instruments were self-report surveys without reliability or validity information.

Selection was the major threat to internal validity in the Bisciglia and Monk–Turner study. 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 into which class they would be in, they also self-selected into which experimental condition they would participate in. Demographic variables were used as an attempt to measure the pre-intervention differences between treatment and control 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.

In the Bisciglia and Monk–Turner study, the construct validity of the control condition was slightly questionable. 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., 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 and Peters, 1999). However, onsite students do not receive

some of the same benefits as offsite students do (e.g., not having to relocate or commute to the physical site of instruction.)

3.2.2. *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 did laboratory exercises for 32 hours. The treatment group did 4 to 8 hours of simulations before doing 24 hours of 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. The instruments used were teacher-made quizzes and tests.

Major threats to validity in the Kennepohl (2001) study were selection and instrumentation. Selection was problematic because it was probable that groups were not equivalent before implementation of the treatment. For example, one group, to begin with, may simply have had more high-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's scoring of quizzes and tests was influenced by knowing which group a student was in. The reliability and validity of measures were not reported. Other threats, such as attrition and reactive effects, may have been possible because little description of the participants, procedure, and setting was provided.

3.2.3. *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 rule out most major threats; therefore, there were only minor plausible threats. Since 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 or about procedures pertinent to reactive or other effects.

3.2.4. *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 computer-mediated instruction. Students in the control group received face-to-face instruction. Students were not randomly selected or assigned; however, the demographic

characteristics of each group were reported. The measures, 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 because 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.

4. 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 methodological flaws in distance learning research before the 1999 publication of *What's the Difference* are still present in distance learning research after 1999.

One surprising discrepancy, however, between the current review and *What's the Difference* is that the proportion of article types differed significantly. For example, in the current review 9% of the articles were experimental studies, but in *What's the Difference* 51% of the articles were experimental studies; see Table 3. I hypothesize that this discrepancy might have happened (a) because of sampling error in the current review's small sample, (b) because the current review's sample was not representative of the population that *What's the Difference*'s sample was representative of, or (c) because the proportion of articles types had actually changed since *What's the Difference* was published. Sampling error is a possibility because so few articles were included; it is entirely possible that the sample of journal issues in the current review are representative of distance education research in general, but those particular issues had a higher proportion of qualitative articles, just by chance. A second possibility is that the sample chosen for the current review is representative of a different population than *What's the Difference* sample is representative of. Although the majority of the journals that were included in *What's the Difference* were also included in this review, *What's the Difference* included other non-journal sources, which the authors of *What's the Difference* broadly specified as "original research." A third possibility is simply that the proportion of article types actually did change since *What's the Difference* was published. Each of these hypotheses is plausible. Replication would be needed to determine which hypothesis, or combination of hypotheses, is correct. Although there was a discrepancy, many of the main findings in *What's the Difference* were, nonetheless, supported by the results of the current review: there is still a paucity of original research, poor control of extraneous variables, a lack of randomized selection, questionable validity and reliability of instruments, and inadequate controls for reactive effects.

4.1. *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 appropriately investigates the causal link between distance learning and academic achievement or student satisfaction.

4.2. *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 for. 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 may be inadequate because demographic variables may not measure the factors that are most related to 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 had been controlled for.

4.3. *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.

4.4. *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 self-report Likert-type surveys, which are subject to strong reactive effects, or they were teacher-made tests or quizzes. Much work must be still done on creating, researching, and reporting the validity and reliability of instruments used in distance education research.

4.5. *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.

5. Study Limitations

There are several limitations of the current review that should be taken into account. First, the sample size was small. A small sample size increases the possibility that the sample selected is not representative of the population. One benefit of there having been only four experimental studies, however, is that they could be analyzed in detail as case studies, which would not have been possible had there been many experimental studies. Another limitation of the current review is that there are no interrater reliability estimates for the categorizations of articles, because only person was involved in this review. It is not known whether a second, independent reviewer would have categorized the articles in this sample in the same way as they were categorized in the current review.

6. Conclusion

Based on the sample reviewed here, the same shortcomings in the distance learning literature mentioned in *What's the Difference* were still present in the recent distance learning literature. 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., and E. Monk-Turner (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). Education ain't what it used to be. *Information Today*, **20**, December, 37–38.
- Hammond, R.J. (1997). 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.athabascau.ca/vol16.2/kennepohl.html>
- Litchfield, R.E., M.J. Oakland and J.A. Anderson (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., and L. Rainey (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., and M.J. Peters (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.
- Phipps, R., and J. Merisotis (2004). *What's the Difference: A Review of the Contemporary Research on the Effectiveness of Distance Learning in Higher Education*. The Institute for Higher Education Policy (1999, April). Retrieved April 21, 2004, from <http://www.nea.org/he/abouthe/diseddif.pdf>

- The pros and cons of e-learning (2002). *School Planning and Management*, **41**(12), 9.
- Russell, T.L. (1999). The no significant difference phenomenon. Office of Instructional Telecommunications, North Carolina State University, Chapel Hill, NC.
- Salomon, K. (2004). Distance ed: The return of the DEPD. *University Business*, **7**, February, 24–25.
- Shadish, W.R., T.D. Cook and D.T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, New York.
- 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.

J.J. Randolph is a planning officer at the Department of Computer Science, University of Joensuu, Finland. He is also a PhD candidate in Utah State University's Education Research and Evaluation Program. His research interests include evaluation and program planning, research and evaluation methodology, and technology education.

Ar padarėme išvadas? Nuotolinio mokymo tyrimų metodologinės apžvalgos tąsa

Justus J. RANDOLPH

Nuotolinio mokymosi literatūros, išleistos 1992–1999 metais, išsami apžvalga atskleidė, jog daugelis nuotolinio mokymosi tyrimų turi nemenkų metodologinių spragų. Straipsnyje supažindinama su rezultatais, gautais tam tikru mastu replikuojant minėtą apžvalgą. Iš trijų pagrindinių mokslinių žurnalų, skirtų nuotolinio mokymosi klausimams, atrinkti bei, atsižvelgiant į jų tipą, suklasifikuoti 66 straipsniai, išanalizuotos eksperimentiniuose bei kvaziekperimentiniuose straipsniuose aprašomos tyrimų metodologijos. Gautieji rezultatai atskleidė, jog ir po 1999-ųjų metų pasirodę straipsniai pasižymi tomis pačiomis metodologinėmis spragomis, kaip ir straipsniai, spausdinti iki 1999-ųjų: dauguma aprašomų eksperimentų dalyvių būdavo parenkami ne atsitiktiniu būdu, išoriniai kintamieji ir grįžtamieji ryšiai tinkamai nekontroliuojami, taip pat nebūdavo aprašomas pasirinktų matavimo vienetų pagrįstumas bei patikimumas.