

# Measuring Learning Effectiveness: A New Look at No-Significant-Difference Findings

*Ernest H. Joy II, Ph.D.*

Florida Institute of Technology  
Hampton Roads Graduate Center  
P.O. Box 4323  
Fort Eustis, VA 23604-0323  
Phone: 757-440-9005  
Fax: 757-440-9309  
E-mail: [ftechjoy@pinn.net](mailto:ftechjoy@pinn.net)

*Federico E. Garcia, Ph.D.*

Center for Naval Analyses  
4401 Ford Avenue  
Alexandria, VA 22302  
Phone: 703-824-2875  
Fax: 703-824-2949  
E-mail: [garciaf@cna.org](mailto:garciaf@cna.org)

## ABSTRACT

Researchers, instructional designers and consumers of ALNs must be cautious when interpreting results of media comparison studies. Much of the literature purports to have found no significant difference in learning effectiveness between technology-based and conventional delivery media. This research, though, is largely flawed. In this paper, we first outline the philosophical positions of the opposing sides of an intense debate in the literature as to whether delivery media alone influence learning outcomes. We then select at random several representative media comparison studies to illustrate the inadequacy of their methodologies and conclusions. More important, we derive critical design considerations for those who evaluate or conduct media comparison research. ALN practitioners should not assume that students would learn better from technology delivery systems. Rather, ALN practitioners should adhere to time-tested instructional design strategies, regardless of the medium they choose. Learning effectiveness is a function of effective pedagogical practices. Accordingly, the question for ALN practitioners ought to be: "What combination of instructional strategies and delivery media will best produce the desired learning outcome for the intended audience?"

## KEY WORDS

ALN, Instructional assessment, Computer-aided instruction

## I. INTRODUCTION

Much of the literature in the field of instructional technology purports to have found no significant difference in learning effectiveness between technology-based and conventional delivery media. In the wake of these findings, and as technology has improved, there is a growing demand for instruction delivered through electronic media, especially through asynchronous learning networks. Practitioners and consumers of asynchronous learning networks (ALNs) need to be aware that much of the research in this field is seriously flawed, rendering many of the conclusions inaccurate or open to debate. Designers of

ALN should not assume that one delivery medium is as effective as another. Such a determination should be made through properly designed and executed research. Regrettably, this proposition has eluded researchers for years.

For much of the twentieth century, educators and behavioral scientists have compared the instructional effectiveness of different delivery media. At first glance, this seems a logical pursuit given the changes in instructional media over the years, particularly in computer-mediated instruction. Although there are many media comparisons in the literature, the most common approach has been to compare traditional (teacher-mediated) learning with technology-based devices as either a substitute for or supplement to the teacher. In the United States prior to World War II, film and radio were the focus of many comparison studies [1]. After the war, the ensuing race for space between the two superpowers led to extensive government funding for research and development. About this time, researchers began to assess not just the equipment, but its capabilities for improving instructional effectiveness [2]. The focus on specific attributes of media has grown with improvements in technology. In recent years, comparisons have shifted their emphasis from television and programmed instruction to personal computers and distance education.

Prior to 1980, there were more descriptive studies than experimental studies comparing computer-delivered instruction with traditional delivery modes. According to Maddux [3], this trend shifted in the 1980s as researchers and educational software developers became interested in establishing cause-and-effect relationships between computer and non-computer delivery modes. Many turned to scientific experimentation as the method of choice. As we will explain shortly, however, many of these studies suffer from problems of internal and external validity because of specification and design error. In the 1990s, Maddux observed a shift toward improvement in the internal design of computer delivery in experimental comparison studies, which he attributed to better-controlled laboratory research. Still, the major flaw in much of the research was its limited generalizability beyond the laboratory.

Throughout these developmental stages of media comparison research, the debate as to whether media alone influence learning outcomes has continued. In the lead of those who believe that media will never influence learning has been Richard E. Clark [4], [5], [6]. He has argued (and has been supported independently by Gagne et al. [7]) that media per se do not influence learning. Rather, “learning is caused by the instructional methods embedded in the media presentation” [6, p. 26]. As cited by Clark [6, p. 23], Gavriel Salomon defined instructional method as “any way to shape information that activates, supplants, or compensates for the cognitive processes necessary for achievement or motivation.” Gagne et al. [7] also specified those events that are necessary to learning as comprising both internal and cognitive information processing events, such as temporary storage, encoding, and retrieval. He also specified external instructional events, such as gaining attention, drill and practice, and feedback, which are necessary to activate and support the internal process [7, p. 202]. Gagne et al. treat media not as an external instructional event, but rather as a “vehicle for the communications and stimulation that make up instruction” [7, p. 205].

Because Clark contends that media themselves are merely the conveyors of instructional methods and content, he goes on to conclude that they do not directly influence learning in any way. His challenge to the critics embodied his main argument: “We need to ask whether there are other media or another set of media attributes that would yield similar learning gains. If a treatment can be replaced by another treatment with similar results, the cause of the results is in some shared (and uncontrolled) properties of both treatments” [6, p. 22].

Opposing Clark is Robert Kozma of the Center for Technology and Learning. His main argument has been that media and methods are inextricably interconnected. According to Kozma, both media and methods are part of the instructional design. “Media must be designed to give us powerful new methods,

and our methods must take appropriate advantage of media’s capabilities” [8, p. 16]. Further, Kozma has maintained that learning from media can be thought of as a complementary process within which representations are constructed and procedures performed, sometimes by the learner and sometimes by the medium [8, p. 11]. By means of its technological capabilities, symbol system, and processing capabilities, Kozma argues that a particular medium “can be described in terms of its capability to present certain operations in interaction with learners who are similarly engaged” [8, p. 11].

In our view, Clark has made a stronger case than Kozma. First, Clark’s argument that media are primarily vehicles for instructional methods coincides with widely accepted learning theory. Until that theory is revised to include media as an external instructional event, or as an instructional method, the two must be treated as separate yet complementary components of an instructional process. Even more compelling is Clark’s argument that many different media attributes may accomplish the same learning goal [6].

Until a single medium becomes *necessary* for learning, which is not likely in most settings, media will continue to serve only as a conveyor of the processes contributing to learning. This is not to say that some media are not vastly superior to others as conveyors of methods and content. Technology has made great strides and is rapidly outpacing other modes of delivery in terms of cost and efficiency, as we will show later. Nevertheless, as long as there is another medium capable of conveying similar methods of instruction in any given instructional setting resulting in similar learning outcomes, Clark’s argument that the cause of the learning must be something other than the media themselves is compelling indeed. Kozma’s retort [9] is that, if two treatments yield a similar outcome, it does not mean that they resulted from the same cause. Again, Clark is not claiming that there necessarily is a single causal factor for learning to occur with different media. He is simply arguing that, as long as similar learning occurs with different media, there must be some cause other than the media themselves.

While this fundamental debate rages on, researchers continue to conduct media comparison studies. Most of these studies report no significant differences in learning effectiveness between electronic and conventional classroom delivery. One of the best known summaries of such research [10] lists several hundred studies claiming to have found such results. Because of the increasing demand for electronic delivery of instruction, and because of the controversial arguments just cited, a closer look at the issues of research design quality, results and interpretations is warranted.

## II. MEDIA COMPARISON: FLAWS IN THE EXPERIMENTAL RESEARCH

In a meta-analysis of a randomly selected 30 percent of approximately 500 experimental studies conducted in the 1970s, Clark [5] found many examples of achievement gains for technology-delivered instruction. He found that the impact of technology-delivered instruction was “overestimated due to often uncontrolled for but robust instructional treatments embedded in computer-based-instruction (CBI)” [5, p. 249]. Specifically, he noted that many of the instructional methods built into CBI, such as corrective feedback and individual pacing, were not used by teachers [5, p. 250]. In essence, Clark determined that in such confounded research, “causes and effects cannot be unambiguously identified. When causes are confounded, it is still possible to determine the size of the effect of the treatment on the outcome. However, it is not possible to determine which of many uncontrolled—yet correlated—parts of the treatment contributed to the differences” [5, p. 252]. Overall, Clark concluded that about 75 percent of the studies he examined had serious design flaws, 50 percent failed to control for time on task, and only 50 percent controlled for instructional method. There were significant differences in favor of CBI in only two of fifteen studies that controlled for instructional method [5, p. 259].

In a meta-analysis of 12 studies involving computer-aided instruction (CAI) in adult basic and secondary education, Rachal [11] found similar design flaws, including lack of specification of time on task and the size of control and experimental groups. Also, he found that the number of subjects was small, treatment periods were too short, and the studies never assessed the abilities of the individual instructors assigned to teach the control groups. He concluded that, while results in at least one well-controlled study found significant differences favoring CAI over non-CAI delivery, expectations of CAI becoming a miracle cure for adult education programs are unrealistic [11, p. 172].

In a paper dealing with problems of learner control research, Reeves [12] held that “much of the research in the field of CBI is pseudoscience because it fails to live up to the theoretical, definitional, methodological, and/or analytic demands of the paradigm upon which it was based” [12, p. 39]. He recommended that graduate students and researchers “develop an improved understanding of contemporary philosophy of science, a topic largely ignored in many research methodology and statistics courses” [12, p. 43]. In this way, researchers would be exposed to a wider range of options in conducting experimental and non-experimental designs.

In a report on effectiveness of technology in schools, Sivin-Kachala and Bialo [13] summarized the findings of 133 media comparison studies. Sponsored by the Software Publishers Association, the authors argued that “educational technology has demonstrated a significant positive effect on achievement” [13, p. 2]. As with several literature reviews, the authors did not report the adequacy of the research methods in these studies.

We examined five of the studies cited by Sivin-Kachala and Bialo and found significant design flaws in each of them, as shown in Table 1. A brief summary of the findings for each of these five studies follows:

Author(s), year	Random sampling		Size	Controls						
	Select	Assign	N	PK	Ability	Learn style	Teacher effects	Time on task	Instruct. method	Media familiarity
Reglin, 1985	No	No	48	No	No	No	No	Yes	No	No
Clements, 1991	Yes	Yes	73	Yes	No	No	No	No	No	No
Skinner, 1990	No	Yes	36	Yes	Yes	Yes	No	No	No	No
Gardner et al., 1992	No	Yes	114	No	No	No	Yes	Yes	No	No
Funkhouser, 1993	No	No	40	Yes	No	No	NA	No	Yes	No

Definition of abbreviations:

- Select: Selected randomly
- Assign: Subjects assigned to control or treatment groups randomly
- N: Number of observations
- PK: Prior knowledge

Table 1. Design Characteristics of a Random Sample of Media Comparison Studies (1985-1993).

In a study of CAI effects of math achievement on African-American students seeking admission to teacher education programs, Reglin [14] found significant effectiveness differences between CAI and traditional instruction. However, he failed to control for prior knowledge, ability, learning style, instructor effects, learner familiarity with technology, and method of instruction. Based on lack of adequate controls, there simply is no way to verify that the differences in pre- and post-tests were the result of the media used in the experimental group.

Clements [15] studied enhancement of creativity in computer environments using LOGO software versus a non-LOGO CAI user group. While he, too, found a significant difference in pre- and post-test scores favoring LOGO users, he failed to adequately explain differences in instructional methods other than to

say that one group used LOGO and the control group did not. Also, he failed to control for ability, learner style, and learner familiarity with the technology, and students in the control group had less time on task than the LOGO users.

Skinner [16] examined the effects of CBI on achievement by tracking 36 college students who rotated among three instructional delivery treatments: Text/CBI/Guided Tutorial, Text/CBI/Solo, and Text-only. Skinner also used a matched-pair procedure for assignment of subjects into two groups to control for possible confounding effects of different module difficulty levels. He controlled for ability by dividing students into levels of achievement based on previous course performance. Results indicated that test scores were significantly higher for students in CBI modules. Average performance was lower in Text-only modules. Although he controlled for individual effects across all three delivery modes, he failed to control for instructional methods because the Text-only modules lacked many of the methods employed by CBI, such as drill, practice and feedback. Also, because learners were able to do as much as they chose, time on task varied widely, particularly in the SOLO delivery mode.

Gardner, Simmons, and Simpson [17] conducted an experiment from which they determined that combining CAI with hands-on science activities significantly increased elementary students' learning outcomes. They administered the same pre- and post-test to both control and experimental groups. Although they controlled for time on task and instructor effects, they failed to control adequately for method of instruction, prior knowledge, and learner ability and familiarity with technology. As with many other media comparison studies, it was impossible to determine whether the improvement in achievement was the result of the delivery mode or some other uncontrolled factors.

Finally, among those studies reported by Sivin-Kachala and Bialo [13] as proving the positive effects of computers on learning achievement, Funkhouser [18] investigated the influence of problem-solving software on student attitudes about math. He determined that the use of problem-solving software resulted in more positive student attitudes about themselves as learners of math, and about math as a discipline [18, p. 339]. While he used a pre-test to control for prior knowledge and attitudes toward math, there were no other experimental controls, nor was there any control group (all 40 students were administered the treatment). To measure attitudinal changes, he compared test scores using the same assessment test both before and after treatment. With no control group and lack of controls for random selection and assignment of subjects, this procedure was seriously flawed from the outset. It is impossible to determine whether the changes in student attitudes were caused by the use of the same pre- and post-tests, by the instructional methods presented in the software, or by any other factor, such as familiarity with the software program.

In a meta-analysis of military training, Orlansky [19] assessed the cost and instructional effectiveness of major innovations to training, such as flight simulators, CBI and maintenance training simulators. Orlansky reviewed 30 studies conducted since 1968, which compared conventional and CBI delivery modes. He selected studies in which the control and experimental groups used nearly identical course content, allowed equivalent time on task between delivery modes, and used similar performance measures [19, p. 17]. According to Orlansky, the only difference was that one group was instructed conventionally and the other group used CBI. Conspicuously absent from his criteria were more explicit controls to avoid confounding in critical factors, such as instructional method, prior knowledge and ability. In 40 courses where CAI was compared with conventional instruction, he found student achievement the same in 24, superior in 15 and inferior in one. He concluded that there were no significant differences between delivery modes. For example, in 12 of 13 maintenance training comparisons, he found that "students trained with simulators achieved the same or better test scores than those trained with actual equipment" [19, p. 26]. This is not surprising in view of the limits that Orlansky placed on control mechanisms in the research he cited.

Based on all of the above, it seems clear that findings of no significant difference are often misinterpreted. For example, Orlansky [19] concluded that “flight simulators, CBI, and maintenance simulators are *as effective* for training as aircraft, conventional instruction, and actual equipment, respectively” [19, p. 41]. In poorly designed research, this may not be a valid conclusion.

If the researcher has not carefully controlled for the most likely factors explaining variance in student achievement, it is less likely that one will find significant difference between experimental and control groups. This is because there is no absolute difference affecting only the treatment group. Likewise, if a significant difference is found in poorly designed research, it may be the result of one or more uncontrolled variables, such as a specific method of instruction that was presented to the experimental group only. This common misinterpretation was also addressed in Clark’s seminal article on learning from media [4], in which he argued that “no significant difference results simply suggest that the changes in outcome scores (e.g., learning) did not result from any systematic differences in the treatments compared” [4, p. 447].

### III. CONCLUSION

The outlook for improving the design of media comparison studies is bleak. Even if a legitimate scientific model could be designed to properly control for each independent variable, its usefulness for predicting learning outcomes would, in all likelihood, be extremely limited. This is because the researcher would have to impose artificial controls to produce such a result. For those who choose to persist in comparing delivery media, the evaluation criteria in Table 1 provide a helpful list of variables to consider when evaluating or designing such research. Any media comparison or design that fails to control for the variables listed in Table 1 is probably suspect and may produce inconclusive results.

This paper has argued that ALN practitioners must be cautious when interpreting results of media comparison studies. ALN practitioners should not assume that students would learn better from technology delivery systems. Rather, ALN practitioners should adhere to their time-tested instructional design strategies, regardless of the medium they choose. It is widely accepted that learning effectiveness is a function of effective pedagogical practices. Accordingly, the question for researchers, instructional designers, and consumers of ALNs ought to be: “What combination of instructional strategies and delivery media will best produce the desired learning outcome for the intended audience?”

### ACKNOWLEDGMENTS

This paper was presented at the Fifth International Asynchronous Learning Networks Conference, College Park, Maryland, in October 1999. The authors would like to thank the conference chair for ALN evaluation and program results, Starr Roxanne Hiltz, for her valuable inputs and encouragement to publish this paper. The authors would also like to thank David M. Moore, John Burton, Kusum Singh, Katherine Cennamo, and Glen Holmes for their comments on an earlier version of this paper.

### REFERENCES

1. **Moore, D. M.**, Personal communications, Spring 1997.
2. **Schramm, V.**, *Big Media, Little Media: Tools and Technologies for Instruction*, Beverly Hills, CA, Sage Publications, 1977.
3. **Maddux, C. D.**, Research in Educational Computing: Problems of Internal and External Validity, *Computers in the Schools*, 11(3), 7-10, 1995.

4. **Clark, R. E.**, Reconsidering Research on Learning from Media, *Review of Education Research*, Vol. 53, No. 4, pp. 445-459, 1983.
5. **Clark, R. E.**, Evidence for Confounding in Computer Based Instruction Studies: Analyzing the Meta-Analyses, *Educational Technology Research and Development*, Vol. 33, No. 4, pp. 235-262, 1985.
6. **Clark, R. E.**, Media Will Never Influence Learning, *Educational Technology Research and Development*, Vol. 42, No. 2, pp. 21-29, 1994.
7. **Gagne, R. M., Briggs, L. J., and Wager, W. W.**, Principles of Instructional Design, 4<sup>th</sup> ed., Orlando, Harcourt Brace Jovanovich, 1992.
8. **Kozma, R. B.**, Will Media Influence Learning? Reframing the Debate, *Educational Technology Research and Development*, Vol. 42, No. 2, pp. 7-19, 1994.
9. **Kozma, R. B.**, A Reply: Media and Methods, *Educational Technology Research and Development*, Vol. 42, No. 3, pp. 11-13, 1994.
10. **Russell, T. L.**, Television's Indelible Impact on Distance Education: What We Should Have Learned From Comparative Research, *Research in Distance Education*, October 2-4, 1992.
11. **Rachal, J. D.**, Computer Assisted Instruction in Adult Basic and Secondary Education: A Review of the Experimental Literature 1984-1992, *Adult Education Quarterly*, Vol. 43, No. 3, pp 165-172, 1993.
12. **Reeves, T. C.**, Pseudoscience in Computer-Based Instruction: The Case of Learner Control Research, *Journal of Computer Based Instruction*, Vol. 20, No. 2, pp 39-46, 1993.
13. **Sivin-Kachala, J., and Bialo, E. R.**, Report on the Effectiveness of Technology in Schools, 1990-1994. New York: Interactive Educational Systems Design, Inc., 1994.
14. **Reglin, G. L.**, CAI Effects on Mathematics Achievement and Academic Self-Concept Seminar, *Journal of Education Technology Systems*, Vol. 18, No. 1, pp. 43-48, 1989.
15. **Clements, D. H.**, Enhancement of Creativity in Computer Environments, *American Education Research Journal*, Vol. 28, No. 1, pp. 173-187, 1991.
16. **Skinner, M. E.**, The Effects of Computer-Based Instruction on the Achievement of College Students as a Function of Achievement Status and Mode of Presentation, *Computers in Human Behavior*, pp. 351-360, 1990.
17. **Gardner, C. M., Simmons, P. E., and Simpson, R. D.**, The Effects of CAI and Hands-on Activities on Elementary Students' Attitudes and Weather Knowledge, *School Science and Mathematics*, Vol. 92, No. 6, pp. 334-336, 1992.
18. **Funkhouser, C. P.**, The Influence of Problem-Solving Software on Student Attitudes about Mathematics, *Journal of Research on Computing in Education*, Vol. 25, No. 3, pp. 339-346, 1993.
19. **Orlansky, J.**, The Cost Effectiveness of Military Training. Paper presented at the Symposium on Military Value and Cost-Effectiveness of Training, Brussels, Belgium, 1985.