to teach, but also to be interrupted, cut short, canceled, and overlooked. The key word, *flexibility*, is just another way of saying that the world of the inner-city school is rarely predictable, neat, and tidy – and almost never divided into 52-minute segments.

8. Expect to be in bed and asleep by 8 p.m. I remember quite clearly that, in my last life as a junior high teacher, I cheerfully taught all day, took classes at night, and partied all weekend. Perhaps teaching is only for the young and vigorous. Certainly, it's only for the very healthy. Whatever you do, don't get sick, don't make any plans for a rich social life, and don't even think about having a personal crisis.

9. Remember that rudeness is a concept of middle-class adults. After a few years in the university "refinery," it's hard not to be shocked, insulted, and hurt by junior high schoolers. They shout; they scream; they ignore you completely; they refuse your most reasonable request. My personal breakthrough came when I asked a student to sit down and she shouted, "Ain't *nobody* tells me what to do!"

I barely managed to curb my impulse to shout back, "I'm your teacher, and I damn well *will* tell you what to do!" Instead, I asked her why she was talking to me in that mean, nasty way.

She thought about my question for a few seconds and then responded, "That's just the way we talk in my family. It don't mean nothing." And so it don't.

How do I feel about my return to the inner-city classroom, now that it's over? The year was neither as dreadful as Ornstein might have predicted nor as easy as I naively expected it to be. In fact, the experience was rather like giving birth. There were excruciating moments, when I cursed the day I had made the fateful decision to embark on this enterprise. But in between there were good times. The students did learn a thing or two, and we eventually grew to enjoy one another. Like childbirth, my return to the classroom transformed me. How long this transformation will last is another question.

Last week one of my student teachers asked, "How long since you've been in the classroom?"

"Just last year," I answered proudly.

"Well, things have changed a lot since then." K

The Political Arithmetic Of Cost-Effectiveness Analysis

In the June Kappan, *Richard Niemiec, Herbert Walberg, and Madeline Blackwell critiqued an article by Henry Levin and Gail Meister. Here the original authors, joined by Mr. Glass, fire back a countervolley.*

BY HENRY M. LEVIN, GENE V GLASS, AND GAIL MEISTER

IN JUNE 1986 two of us (Levin and Meister) published an article in the *Kappan*, "Is CAI Cost-Effective?" This article was based on some of the findings of a four-year research project that was carried out, in part, with Gene Glass, then a professor at the University of Colorado.[1] Essentially, we suggested in the June article that, while the cost-effectiveness of one of the most widely used approaches to computer-assisted instruction (CAI) was superior to that of extending the school day or reducing class size, it was not as good as the cost-effectiveness associated with an exemplary program of peer tutoring. Since our data and analysis had been reviewed by numerous experts on the subject, we felt comfortable in reporting them to the *Kappan* audience.

HENRY M. LEVIN is a professor of education and economics at Stanford University, Stanford, Calif. GENE V GLASS is a professor in the policy studies program of the College of Education at Arizona State University, Tempe. GAIL MEISTER is a research assistant at Stanford Univeristy.

Therefore, it was with some surprise that we received the June issue of the *Kappan* and found that our article was followed by a critique that had been solicited by the editors without our knowledge.[2] We would have been pleased to have been active participants in a symposium on the subject, in which our article was reviewed by other experts on cost-effectiveness analysis and in which we would have provided a response. However, we were neither informed of the solicited critique, nor did the authors (Richard Niemiec, Madeline Blackwell, and Herbert Walberg) or the editors send us a prepublication copy. Worse yet, the fact that none of the authors of the critique had training, expertise, or publications in cost-effectiveness analysis led to a presentation on the subject that is beset with errors and confusion. The purpose of this article is to attempt to clarify some of the issues and to correct the misunderstandings implicit in the critique, while giving readers a more detailed understanding of the application of cost-effectiveness to educational interventions.

COST-EFFECTIVENESS AND EDUCATIONAL DECISIONS

Cost-effectiveness analysis is designed to guide decision makers in the allocation of scarce resources by spelling out the consequences of decision alternatives for costs and for educational results.[3] Cost-effectiveness analysis must satisfy at least two requirements in order to be of use to a decision maker. First, the educational interventions that are evaluated must be readily implementable. The alternatives under consideration should be interventions that

have been applied in conventional settings, that have been established for a reasonable time, and that have characteristics that make them transferable to other settings. This requirement tends to rule out interventions that were designed solely for research purposes, that were used for only a short period, or that required special conditions (such as university sponsorship) that cannot be easily replicated on an ongoing basis.

Second, the methods used to evaluate costs and effectiveness must be acceptable. Unfortunately, many assessments of effectiveness violate the most basic tenets of evaluation methodology, and assessments of costs are often unreliable because they draw on readily available data that are flawed or incomplete.[4] In our work, we reviewed hundreds of studies to find ones that met the dual standards of ease of implementation and reliability.[5] In comparing CAI with peer tutoring, we finally chose an approach for each that met both criteria. The Boise (Idaho) tutoring model, PROJECT INSTRUCT, has been used for many years and has been subject to regular evaluations.[6] The approach was so highly regarded that it was recommended by the National Joint Dissemination Review Panel of the U.S. Department of Education as an exemplary Chapter 1 project that should be disseminated nationally. Finally, the technical quality of the evaluations of the effectiveness of PROJECT INSTRUCT on reading and mathematics achievement ranked them among the best that we had seen for peer-tutoring interventions, and our ability to estimate program costs was enhanced by the availability of detailed information on the resources required to replicate PROJECT INSTRUCT.

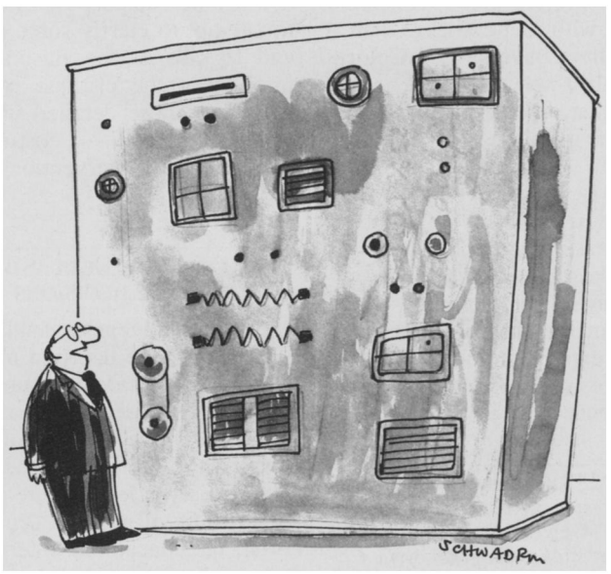
Even worse is the total lack of thinking that accompanies the use of averages of many studies to provide information on cost-effectiveness to decision makers.

From the large number of CAI interventions, we also chose one that has received national recognition and that has been implemented at hundreds of sites across the U.S.: the drill-and-practice curriculum of the Computer Curriculum Corporation (CCC). In the most recent comprehensive summary of computer-based education at the elementary level, more than half of the evaluations were based on this CCC intervention.[7]

Not only has the CCC intervention been widely implemented, but it is also the subject of what is generally acknowledged to be the best technical evaluation of the effectiveness of CAI: the four-year longitudinal study that the Educational Testing Service (ETS) conducted in the Los Angeles Unified School District (LAUSD), under the aegis of the National Institute of Education.[8] The ETS/LAUSD evaluation design surpasses in scope and sophistication other evaluations of CAI, including those evaluations of CCC approaches that have been conducted at other sites. Detailed information on the resource requirements for the intervention were obtained both through data collection at CCC sites and from CCC headquarters, and 1984 costs of equipment and software were obtained directly from CCC.[9]

In summary, each of these approaches was an exemplary and readily implemented intervention. Excellent sources of evaluation data regarding both costs and effects were also available for each.

When we analyzed the evaluation data, we found peer tutoring to be associated at the elementary level with an effect size of .97 in mathematics (a gain of almost one school year) and with an effect size of about .48 in reading (a gain of almost five months, or half a school year). CAI was associated with considerably smaller effects: .12 (or 1.2 months) in mathematics and .23 (or 2.3 months) in reading. However, the costs of peer tutoring were higher (about $212 per student for each subject) than the costs of CAI (about $119 per student for each subject). Even when the higher costs were taken into account, we found peer tutoring to be almost five times as cost-effective as CAI (per dollar of cost) for teaching mathematics and slightly more cost-effective than that for teaching reading. (For teaching math and reading, both of these approaches were considerably more cost-effective than reductions in class size or extending the school day.)



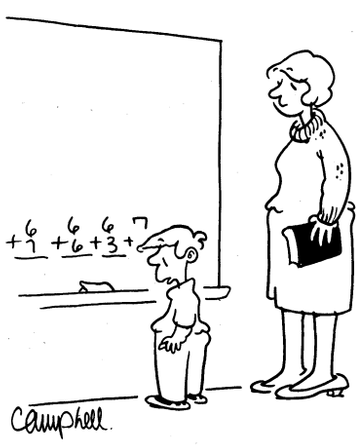
"Here's one I bet you can't answer: What's black and white and read all over?"

THE CRITIQUE

On the basis of other data, Niemiec, Blackwell, and Walberg assert that the cost-effectiveness of CAI is superior to that of peer tutoring. Their argument is

based solely on a challenge of our findings regarding effectiveness, for they do not present new cost estimates to support their claim. This was probably a wise decision, given their dearth of knowledge of and experience with costs and cost estimation. However, with regard to effectiveness they suggest that meta-analyses or statistical summaries of some 65 studies of peer tutoring and some 45 studies of CAI show average effect sizes that are considerably higher for CAI and considerably lower for peer tutoring than our estimates. Using our cost estimates, they conclude that the cost-effectiveness of CAI is superior to that of peer tutoring.

Essentially, their assertion is based on the seemingly logical assumption that the results of a large number of interventions provide a basis for decision making that is superior to that provided by the results of two carefully selected interventions that meet certain standards regarding ease of implementation and reliability of the evaluations. What they do not point out is that their averages are based on a motley group of interventions – including those that failed and those that succeeded, those that were experiments of a few weeks' duration and those that were ongoing programs in schools for a semester or more, those with unacceptable evaluation designs and those with good ones, those that used specially constructed tests to demonstrate their effectiveness and those that used nationally standardized tests for that purpose. Niemiec, Blackwell, and Walberg arrived at the average effect that they report for peer tutoring by combining interventions as diverse as those in which tutors were carefully selected, trained, supervised, and provided with tutoring materials and those in which students were just thrown together with other students and told to tutor them.



"Where's the keyboard?"

Decision makers are interested in the consequences of the real alternatives they face, not in averages of studies with unknown properties conducted over three decades.

They have given "average answers" to "average questions." But we did not ask "average questions." Rather, we inquired into the effectiveness of specific programs that were the result of serious efforts by school districts to incorporate high-quality peer tutoring or CAI into their curricula. An average of many programs – including those that failed – tells decision makers nothing about programs that succeed. What are the characteristics of the "average program," how can it be implemented, and what are its costs? The answer is that the average is not a "program" at all, but a conglomerate of many different programs.

Even worse is the total lack of thinking that accompanies the use of averages of many studies to provide information on cost-effectiveness to decision makers in education. For example, James Kulik and his colleagues have shown that there is an average effect size of .56 for CAI in projects lasting less than four weeks but an average effect size of only .20 for CAI in projects lasting more than eight weeks.[10] In other words, projects of longer than eight weeks show effect sizes that are very close to our estimate, and we would argue that these longer projects – not the shorter experimental studies – are pertinent to implementation in the schools.

Meta-analyses exaggerate the effects of CAI by attributing to it the achievement gains due to non-CAI instructional treatments that are embedded in the so-called CAI interventions.[11] Overstatements of CAI effects also stem from the use of inappropriate control groups in experimental studies, the use of "local" tests that are tailored to the interventions, and the involvement of the evaluators in the interventions.[12] Richard Clark found that 75% of the CAI evaluations in the meta-analyses that he reviewed "gave evidence of serious design flaws."[13] Some idea of the potential overestimation of CAI effects can be gained by comparing a meta-analysis of 15 CCC interventions in elementary schools[14] with the results of the ETS experiment in the Los Angeles Unified School District, which used CCC equipment and curricula. Because the meta-analysis includes evaluations of widely varying quality and duration, as we have noted above, the average effect size is likely to be overstated relative to CAI interventions that are more carefully evaluated and of longer duration. By contrast, the ETS/LAUSD study is recognized as one of the best-designed and statistically most carefully analyzed evaluations; this study provides results for one year of instruction or more in various subjects and at various levels of exposure to CAI treatments.[15] In line with our expectations, the meta-analysis of CCC interventions shows an average effect size that is more than twice as large as that of the careful ETS/LAUSD evaluation.[16]

Accordingly, we believe that a more refined comparison of specific alternatives is the appropriate basis for estimating cost-effectiveness. Decision makers are interested in the consequences of the real alternatives they face, not in averages of studies with unknown properties conducted over an interval of three decades.

Finally, we are perplexed about the derivation of the meta-analysis results for CAI. One of the authors of the critique, Herbert Walberg, has reported these results in three different places, each time providing an estimate of effect size that differs considerably from the others. For example, in 1984 he published an article for a general audience of educators that claimed that CAI has a modest effect size of .24.[17] (This is almost identical with our estimate of the effect size of CAI in reading.) In an article written with Richard Niemiec for an audience of computer

educators and published in 1985, Walberg asserted that the effectiveness of CAI is much greater, with effect sizes of .34 for tutorial approaches and of .47 for drill and practice.[18] Then, at virtually the same time that Walberg was working on the critique of our article in the *Kappan* – in which he and his co-authors claimed that "CAI is about twice as effective as peer tutoring" – he was serving as the major academic advisor to the authors of the Education Department's widely distributed publication, *What Works*, in which cross-age tutoring is endorsed and CAI is ignored. When Secretary William Bennett was asked by reporters about the absence of attention to CAI in *What Works*, he cited Walberg as the authority for the view that "there is no research consensus" on its effectiveness.[19]

The most comprehensive technical evaluation of the methodology of the meta-analysis has stated:

> In recent years a plethora of meta-analyses have emerged in social science research. The need to arrive at policy decisions affecting social institutions fostered the momentum toward summarizing research. But, as with most methodologies, abuse frequently accompanies use.[20]

One of those abuses is "political arithmetic," in which the results from large numbers of studies are invoked to bolster a political position in lieu of carefully subjecting that position to scientific scrutiny.

We believe that a disinterested evaluation of our technical reports on the cost-effectiveness of CAI would support both the logic and the substance of our conclusions. Relative to the other interventions we have examined, a major drill-and-practice approach using CAI comes out well on the basis of cost-effectiveness criteria – but not as well as peer tutoring. Other approaches using CAI may be found to be more powerful with regard to costs, and we may see improvements in the cost-effectiveness of CAI as new developments occur. We would be pleased to report such changes in the future, but they are not reflected in existing data.

1. Henry M. Levin, Gene V Glass, and Gail R. Meister, *Cost-Effectiveness of Four Educational Interventions* (Stanford, Calif.: Institute for Research on Educational Finance and Governance, Stanford University, Project Report No. 84-A11, 1984); Gene V Glass, *The Effectiveness of Four Educational Interventions* (Stanford, Calif.: Institute for Research on Educational Finance and Governance, Stanford University, Project Report No. 84-A19, 1984); and Henry M. Levin, "Cost and Cost-Effectiveness of Computer-Assisted Instruction," in Jack Culbertson and Luvern Cunningham, eds., *Microcomputers and Education* (Chicago: 85th Yearbook of the National Society for the Study of Education, University of Chicago Press, 1986), Ch. 8.

2. Richard P. Niemiec, Madeline C. Blackwell, and Herbert J. Walberg, "CAI Can Be Doubly Effective," *Phi Delta Kappan*, June 1986, pp. 750-51.

3. Henry M. Levin, *Cost-Effectiveness: A Primer* (Beverly Hills, Calif.: Sage, 1983).

4. Ibid., Ch. 3.

5. Glass, *The Effectiveness of Four Educational Interventions*.

6. A telephone conversation with Richard Niemiec revealed that he and his co-authors did not review the Boise study, despite their harsh critique of it. The nationally prominent 1983 study is available from the National Joint Dissemination Review Panel of the U.S. Department of Education or from the authors, Gerri Plumb of the Boise Independent School District and Clair Bowman of Boise State University.

7. James A. Kulik, Chen-Lin C. Kulik, and Robert L. Bangert-Drowns, "Effectiveness of Computer-Based Education in Elementary Schools," *Computers in Human Behavior*, vol. 1, 1985, pp. 59-74.

8. Marjorie Ragosta, Paul W. Holland, and Dean T. Jamison, *Computer-Assisted Instruction and Compensatory Education: The ETS/LAUSD Study* (Princeton, N.J.: Educational Testing Service, April 1982).

9. Levin, Glass, and Meister, *Cost-Effectiveness of Four Educational Interventions*.

10. James A. Kulik, Robert L. Bangert, and George W. Williams, "Effects of Computer-Based Teaching on Secondary School Students," *Journal of Educational Psychology*, vol. 75, 1983, pp. 19-26.

11. Richard E. Clark, "Evidence for Confounding in Computer-Based Instruction Studies: Analyzing the Meta-Analyses," *Educational Communication and Technology Journal*, vol. 33, 1986, pp. 249-62.

12. Kulik, Kulik, and Bangert-Drowns, "Effectiveness of Computer-Based Education. . . ."

13. Clark, p. 259.

14. Kulik, Kulik, and Bangert-Drowns, "Effectiveness of Computer-Based Education. . . ."

15. Ragosta, Holland, and Jamison, *Computer-Assisted Instruction and Compensatory Education*. . . .

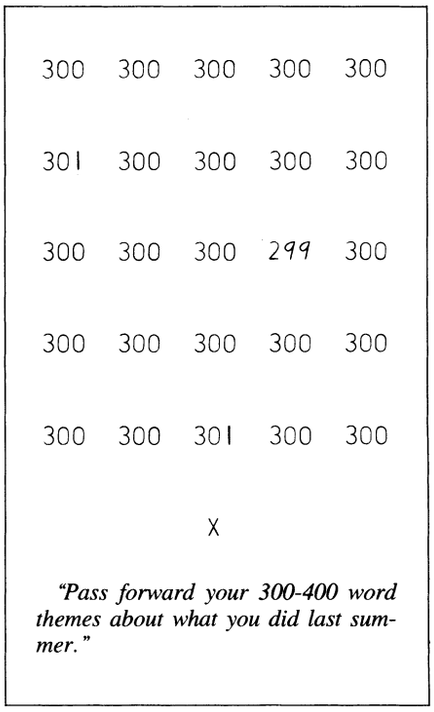
16. Kulik, Kulik, and Bangert-Drowns, p. 69.

17. Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership*, May 1984, pp. 19-27.

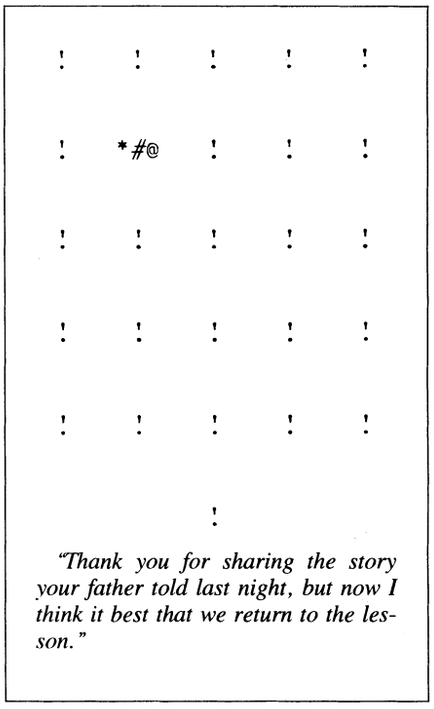
18. Richard Niemiec and Herbert J. Walberg, "Computers and Achievement in the Elementary Schools," *Journal of Educational Computing Research*, vol. 1, 1985, pp. 435-40.

19. James Hertling, "President Promotes New Report," *Education Week*, 12 March 1986, pp. 10-12.

20. Larry V. Hedges and Ingram Olkin, *Statistical Methods for Meta-Analysis* (Orlando, Fla.: Academic Press, 1985), p. xv. K



"Pass forward your 300-400 word themes about what you did last summer."



"Thank you for sharing the story your father told last night, but now I think it best that we return to the lesson."