CAI/CAL and Students' Achievement in Taiwan: A Meta-analysis

Yuen-kuang Cliff Liao
Department of Elementary Education, National Hsinchu Teachers College
521 Nan-Dah Rd. Hsinchu City, Taiwan, R.O.C
yliao@mail.nhctc.edu.tw

Abstract

A meta-analysis was performed to synthesize existing research comparing the effects of computer- assisted instruction/learning (CAI/CAL) with traditional instruction on students' cognitive achievement.

The studies were located from five sources. Totally 52 studies were collected. The quantitative data of the collected studies were transformed into Effect Size (ES). In addition, 17 studied variables were selected and each studied variable was analyzed by one-way ANOVA to relate ESs. The main findings of this study are summarized as follows:

1. The results of this study, suggest that CAI/CAL is more effective than traditional instruction on students' cognitive learning. The mean ES of 0.552 indicates that CAL/CAL was mildly better than traditional instruction on students' cognitive outcomes.
2. For the studied variables, 2 variables, statistical power, and comparison group, had significant main effects on mean ES.

Keyword: computer-assisted instruction, CAI, computer assisted learning , CAL, achievement, outcomes, meta-analysis

Computer-assisted instruction (CAI) has been employed in instruction in US for more than 40 years. Many studies were published to evaluate the effectiveness of CAI on students' learning since 1970s. Meta-analysis is a statistical process whereby the findings of several studies, focusing on a common problem or topic, are pooled in an effort to draw inferences as to the meaning of a collective body of research (). Early meta-analysis studies of CAI were published prior to the microcomputer revolution, beginning in the late 1970s. These studies reported significant improvements for students in elementary and secondary schools on computational arithmetic skills, with effect sizes ranging from .37 to .42 standard deviations [see, for example, Bums (1981) and Hartley (1978)].

Beginning in the early 1980s, a series of meta-analyses related to the effects of computers on learning were published by Kulik and his associates at the University of Michigan (e.g., Bangert-Drowns, Kulik, & Kulik, 1985; J. Kulik, Bangert-Drowns, & Williams, 1983; Kulik, Kulik, & Bangert-Drowns, 1984; Kulik, Kulik, Cohen, 1980). These studies included mainframe CAI studies as well as the emerging literature on microcomputer instruction. Significant effect sizes have been reported for learning via CAI from elementary school (J. Kulik, C-L. Kulik & Bangert-Drowns, 1985), to secondary school (Bangert-Drowns, J. Kulik & C-L. Kulik, 1985), to college (C-L. Kulik & J. Kulik, 1986), and adult learners (J. Kulik, C-L. Kulik, & Schwalb, 1986) ranging from .47 to .26. to .36 to .42, respectively.

In spite of claims regarding the potential benefits of using CAI/CAL in education, research results comparing the effects of CAI/CAL and traditional instruction in Taiwan are conflicting. For example, Chou (1998), Li (1999), Wu (2000a), Wu (2000b), all report significant gains for Web-based learning over traditional instruction. On the other side, Fang (2000), Li (2000), Wang (1999), Wu (1999), and Hsih (2000) have found no significant differences between Web-based learning and traditional instruction. Owing to the contradictory evidence provided by existing research in the area, it is important to conduct a meta-analysis to clarify the research conclusions. The purpose of this study was to synthesize the effects of CAI/CAL verse traditional instruction on students' achievement in Taiwan using meta-analysis.

Method

The research method used in this study is the meta-analytic approach which was similar to that suggested by Kulik, Kulik, & Bangert-Drowns (1985). Their approach requires a reviewer to (a) locate studies through objective and replicable searches; (b) code the studies for salient features; (c) describe outcomes on a common scale; and (d) use statistical methods to relate study features to outcomes (Kulik, Kulik, & Bangert-Drowns, 1985). Their method differs from Glass, McGaw, & Smith (1981) approach is that a single study, defined as the set of results from a single publication, is weighted equally to all other studies, so that aggregate multiple effect sizes from one study can be avoided.

The purpose of this study was to synthesize and analyze the research on effects of two instructional approaches. It is important to define these approaches to provide for proper selection of appropriate studies.

Computer Assisted Instruction (CAI) -- classes using computer-assisted instruction or computer-assisted learning software as instructional tools to teach students.

Traditional Instruction (TI) -- classes using traditional instruction to teach students.

Data Sources

The studies considered for use in this meta-analysis came from four sources and were published in Taiwan. One large group of studies came from computer searches of Chinese Periodical Index. A second group of studies came from Dissertation and Thesis Abstract System of Taiwan. A third group of studies was retrieved from Government Research Bulletin (GRB) of Taiwan. The last group of studies was retrieved by branching from bibliographies in the documents located through review and computer searches.

Fifty-three studies were located through these search procedures; 35 studies came from Dissertation and Thesis Abstract System, 6 studies were retrieved from published journals, and 12 studies were from National Science Council (NSC) research projects. However, one study reported by Tseng (1987) had an Effect Size (ES) several times higher than mean ES of other studies included in the synthesis (i.e., ES = 6.35 for Tseng's study). The study was therefore considered as outlier and excluded from this meta-analysis.

Several criteria were established for inclusion of studies in the present analysis.

1. Studies had to compare the effects of CAI vs. TI on students' achievement.
2. Studies had to provide quantitative results from both CAI and TI classes.
3. Studies had to be retrievable from university or college libraries by interlibrary loan or from GRB, or Dissertation and Thesis Abstract System of Taiwan
4. Studies were published in Taiwan.

Outcome Measures

The instructional outcome measured most often in the 52 studies was student learning, as indicated on standard or researcher-develop achievement tests at the end of the program of instruction. For statistical analysis, outcomes from a variety of different studies with a variety of different instruments had to be expressed on a common scale. The transformation used for this purpose was the one recommended by Glass et al. (1981) and modified by others (e.g., Hunter, Schmidt, and Jackson, 1982). To reduce measurements to a common scale, each outcome was coded as an Effect Size (ES), defined as the difference between the mean scores of two groups divided by the standard deviation of the control group. For those studies that did not report means and standard deviations, F values or t values were used to estimate the ES. Also, in studies which used one-group pretest-posttest design, in which a control group did not exist, an alternative approach suggested by Andrews, Guitar, and Howie (1980) was used. In their approach, the ES is estimated by comparing the post-treatment mean with the pre-treatment mean, and dividing by the pre-treatment standard deviation.

In most cases, the application of the formula given by Glass and his colleagues was quite straightforward. But in some cases, when more than one value was available for use in the formula of ES, the value which measured outcomes most correctly was selected. For example, some studies reported both differences on posttest measures and differences in pre-post gains, and some studies reported both raw-score differences between groups and covariance-adjusted differences between groups. In such cases, pre-post gains and covariance-adjusted differences were selected for estimating ES.

In addition, several subscales and subgroups were used in measuring a single outcome (e.g., those that reported separate data by gender or grade). In such cases, each comparison was weighted in inverse proportion to the number of comparisons within the study (i.e., 1/n, where n = number of comparisons in the study) so that the overweighing of ES of a study could be avoided (see, for example, Waxman, Wang, Anderson, & Walberg, 1985, p. 230).

Variables Studied

Seventeen variables were coded for each study in the present synthesis. Each of these variables was placed in one of the following set of characteristics: (a) study characteristics, (b) methodological characteristics, and (c) program characteristics.

Results

Of the 52 studies included in the present synthesis, 42 (81%) of the study-weighted ESs were positive and favored the CAI group, while 10 (19%) of them were negative and favored the TI group. The range of the study-weighted ESs was from –1.356 to 2.543. The overall grand mean for all 52 study-weighted ESs was 0.552. When this mean ES was converted to percentiles, the percentiles on students' achievement were 71 for the CAI group and 50 for the TI group. The overall grand median for all 53 study-weighted ESs was 0.378, suggesting that percentiles on students' achievement were 65 for the CAI group and 50 for the TI group. The standard deviation of 0.733 reflects the great variability of ESs across studies.

Among the 134 ESs included in the present synthesis, 114 (85%) were positive and favored the HI group, while 20 (15%) were negative and favored the NHI group. The range of the ESs was from -0.768 to 1.914.

The ESs for the 134 comparisons are displayed in a scatter diagram in Figure 1. The diagram shows that despite several large effects, most of the ESs are small to moderate in magnitude. About 85% of ESs lie between -0.5 and 0.5, while less than 10% of the ESs are greater than 0.5.

Table 1 lists the F values for the 17 variables for all study-weighted ESs in the study. Descriptive statistics for the 17 variables are presented in Table 2. Two variables (statistical power and comparison group) showed statistically significant impacts. For each of these variables, a post hoc (Fisher Protected LSD) test was performed.

Table 1

Results of ANOVAs for Coded Variables

Variables	df	*F*	*p*
Study Characteristics			
Grade Level	3,49	2.260	0.093
Subject Area	4,48	0.607	0.660
Type of Publication	2,49	0.207	0.814
Year of Publication	3,48	1.359	0.266
Methodological Characteristics			
Instructor Bias	2,49	1.175	0.317
Instrumentation	3,48	0.917	0.440
Reliability of Measure	2,49	0.509	0.604
Sample Size	2,49	0.467	0.630
Selection Bias	2,49	0.225	0.799
Statistical Power_N_15_	1,50	6.247	0.016*
Statistics	2,49	1.229	0.302
Type of Research Design	2,49	2.564	0.087
Program Characteristics			
Comparison Group	2,49	4.903	0.011*
Duration of Treatment	4,47	1.047	0.393
Implementation of Innovation	4,49	1.062	0.386
Type of Instruction for Treatment	1,52	0.008	0.929
Type of CAI	3,48	1.444	0.242

$^*p < .05$

Table 2

Means and Standard Deviations of Study-weighted ESs for Coded Variables

Variables	N	%	ES	SD
Study Characteristics				
Grade Level				
1st - 6 th	20[a]	37.7	0.414	0.479
7th - 9th	15	28.3	0.847	0.853
10th – 12 th	12	22.6	0.231	0.697
College	6	11.3	0.823	0.936
Subject Area				
Math	12[a]	22.6	0.291	0.331
Language/Social Study	8	15.1	0.664	0.817
Science	14	26.4	0.493	0.622
Computer	4	7.5	0.765	1.076
Others	15	28.3	0.661	0.915
Type of Publication				
Journal Article	6	11.5	0.453	0.522
Dissertation/thesis	35	67.3	0.531	0.786
NSC Research Project	11	21.2	0.671	0.691
Year of Publication				
1983 - 1987	10	19.2	0.752	0.763
1989 - 1993	5	9.6	0.584	0.719
1994 - 1999	17	32.7	0.720	0.825
2000 – 2003	20	38.5	0.301	0.608
Methodological Characteristics				
Instructor Bias				
Same	20	38.5	0.586	0.686
Different	17	32.7	0.713	0.784
Unspecified	15	28.8	0.323	0.724
Instrumentation				
Local	41	78.8	0.566	0.716
Standardized	4	7.7	0.721	1.215
Mixed	3	5.8	0.839	0.500
Unspecified	4	7.7	0.024	0.378

Table 5 Cont.

Variables	N	%	ES	SD
Reliability of Measure				
Actual reliability figure	27	51.9	0.525	0.591
Adequate Indication	9	17.3	0.772	0.870
Unspecified or inadequate	16	30.8	0.473	0.883
Sample Size				
1 – 50	15	28.8	0.690	1.076

51 – 100	23	44.2	0.454	0.528
101 – 432	14	26.9	0.564	0.591
Selection Bias				
Adequately minimized	10	19.2	0.507	0.406
Probably threat	27	51.9	0.508	0.688
Unspecified or inadequate	15	28.8	0.661	0.97
Statistical Power_N_15_				
Adequately minimized	48	92.3	0.482	0.639
Probably threat	4	7.7	1.389	1.311
Statistics				
Mean & standard deviation	41	78.8	0.471	0.734
t-value	4	7.7	0.922	0.660
F-value	7	13.5	0.816	0.729
Type of Research Design				
One group repeated measure	6	11.5	1.157	0.862
Pretest-posttest control group	11	21.2	0.559	0.715
Nonequivalent control group	35	67.3	0.446	0.685
Program Characteristics				
Comparison Group				
Traditional instruction	44	84.6	0.433	0.651
No comparison group	7	13.5	1.302	0.876
Others	1	1.9	0.518	0.000
Duration of Treatment				
Less than 2 hours	12	23.1	0.436	0.603
2 to 4 hours	13	25.9	0.502	0.612
4 to 18 hours	5	9.6	1.182	0.928
26 to 162 hours	6	11.5	0.494	0.961
Unspecified	16	30.8	0.504	0.756

Table 5 Cont.

Variables	N	%	ES	SD
Implementation of Innovation				
Replacement for usual instruction	36[a]	66.7	0.525	0.757
Supplement to instruction	18	33.3	0.544	0.682
Type of Instruction for Treatment				
Large group	13	24.1	0.385	0.427
Small group (less than 5 persons in a group)	6	11.1	0.957	0.763
Individual	21	38.9	0.557	0.686
Mixed	2	3.7	-0.116	0.382
Unspecified	12	22.2	0.540	0.999
Type of CAI				
Traditional	20	38.5	0.523	0.685
Multimedia	25	48.1	0.512	0.760
Intelligent	2	3.8	1.591	1.112
Web-based	5	9.6	0.450	0.529

[a]Some studies reported more than one comparison group.

The post hoc test for statistical power, ($F(1,50) = 6.247$, $P<.05$), showed that the mean comparison of studies in which the statistical power was coded as probably threat was significantly higher than the studies coded as adequately minimized.

In addition, the post hoc test for comparison group, ($F(2, 49) =4.903$, $p<.05$), showed that the mean comparison of studies with no comparison group (i.e., one group repeated measure) was significantly higher than studies in which the comparison groups were TI.

Discussion

The results of this meta-analysis indicate that CAI has moderately positive effects on students' achievement over the traditional instruction in Taiwan. An effect is said to be medium when ES = 0.5 and large when ES = 0.8 (Cohen, 1977). The effectiveness of CAI is also confirmed by 18% of positive study-weighted ES values and 85% of positive ESs overall. The moderateness of the effect must be keep in mind, however; the overall study-weighted mean ES of 0.552 only indicates 21 percentile scores higher than the TI group. The percentile scores for the overall grand mean and median were 71 and 65, respectively. The difference of 6 percentile points between them was possibly attributed to the large overall grand standard deviation (0.733).

The analysis of studied variables suggests some interesting trends in the accumulated research base and is discussed in the following sections.

<u>Study characteristics</u>

For the grade-level variable, there was no significant difference of mean ES. However, the smallness of the ES associated with high school subjects (10^{th} – 12^{th} graders) is probably due to the fact that these students have to study very hard for a nationwide college-entrance-examination in Taiwan and using CAI may be not a good approach for this purpose. It is also possible that different instructional approaches were used for these students as compare to other students. More studies need to be conducted to clarify this variable.

CAI studies conducted to measure students' achievement tend to focus on specific subject areas. Studies included in the present meta-analysis were spread in a wide range of subject areas. About 50% studies examined the effects of CAI for teaching math or science. Another 15% of studies concentrated on the teaching of language, writing, or social study. Although no significant differences on ES were found among subject areas, the various subjects examined seem to suggest that CAI has the potential to implement in many different subject areas. However, there was one subject areas (math) that showed the smallest ES (0.291), suggesting the effects of CAI may vary for different subject areas.

Source of studies in a meta-analysis is always an important factor to be examined. The fact that approximately 67% of studies were located from dissertation/thesis and only 6% of studies were from journals was typical in Taiwan. It is probably due to the limitation of pages a paper can be published in some journals, a large number of educational journal articles in Taiwan do not report detail quantitative data, which restrain them from being included in a meta-analysis. The small difference of ES among journal articles, dissertation/thesis, and NSC research project is surprising; usually the larger ES associated with published articles is typical in meta-analysis (Glass, et al., 1981, p. 227).

The year-of-publication variable in the meta-analysis allows an assessment of the effect of CAI over time. About 71% of studies located were published after 1994 suggesting that CAI studies have just become more popular for the past ten year in Taiwan; it is expected that more studies will be published soon. It is also surprising to learn that CAI studies published between 2000 - 2003 have the lowest mean ES (0.301). A possible explanation is the novelty effects of technology.

<u>Methodological characteristics</u>

After reviewing several meta-analysis of media research, Clark (1983) suggested that the positive effects of media seemed to be the uncontrolled effects of instructional method or content differences between treatments that were compared; he concluded that effects more or less disappeared when the same instructor delivered all treatments. Although no significant differences on ES were found, the results of the present meta-analysis show that studies using different instructors for treatments had higher ES than studies using the same instructors (including “no instructor”.) The finding suggests that the positive effects of CAI over TI could be the uncontrolled effects of instructional method noted by Clark.

For instrumentation, about 80% of studies used researcher-develop instruments, and only 8% used standardized instruments. This is possibly because CAI is a new field in educational research in Taiwan and there are not many published instruments available. In addition, the higher mean ES was associated with studies used standardized instruments, suggesting that developing more standardized instruments for evaluating the effects of CAI in Taiwan is necessary.

The sample size for a study may significantly affect the statistical power of the study; in general, the larger the sample size, the better the statistical power. For the present meta-analysis, the largest mean ES was associated with studies with small sample size (1-50). This seems to suggest that the effects of CAI on students' achievement may work better for small sample size in Taiwan. However, the differences of mean ES among small, medium and large sample sizes were small. The results may be considered tentative.

For statistical power, the mean ES of studies coded as probably threat (N<15) was significantly higher than the mean ES of studies coded as adequately minimized (n>15). In general, when the sample size is less than 15 in an experimental study, the result is considered less reliable. However, there were only 4 studies coded as probably threat, again, the result should not be over generalized.

For type of research design, studies that employed repeated measure had higher mean ES than other types of designs. In general, repeated measure design is considered as methodologically weaker than other designs. This result along with the results of analyses of sample size and statistical power suggests that studies which employ weaker research designs may obtain higher ES.

Program characteristics

For comparison group, the mean comparison of studies with no comparison group (i.e., one group repeated measure) was significantly higher than studies in which the comparison groups were TI. This result suggests that CAI becomes more effective when there is no instruction for the comparison group. In one sense, this is reasonable, because students in TI classes have more or less some learning activities, but in "no instruction" classes, there is no learning activity for students. Nevertheless, the results suggest that CAI is more effective than TI on students' achievement in Taiwan.

Duration of treatment is usually a critical variable in meta-analysis. Clark (1983), after reviewed several meta-analyses of CAI, suggested that the effects of new media to instruction were due to a novelty effect, because the ES was reduced when treatment lasted for longer period of time. Liao and Bright (1991) also reported this novelty effect in their meta-analysis of programming on students' cognitive abilities. Although no significant difference was found for this variable, the results of this synthesis do not quite support the previous viewpoint of novelty effect. The largest mean ES was associated with studies lasting 4 – 8 hours, while the mean ESs for studies lasting neither longer nor shorter than 4 -8 hours were not quite different. There may be some unknown effects related to the duration of treatment that influence students' outcomes from CAI in Taiwan and US that result in the distinct outcomes from their perspective meta-analyses. More cross-nation comparative studies need to be done to clarify this issue.

It is usually difficult for teachers to decide whether use a new innovation as a replacement for usual instruction or as a supplement to instruction. In most cases, the decision depends on which approach can provide more effective outcomes. The differences of mean ES between studies coded as replacement and supplement was trivia, suggesting that CAI in Taiwan used as a replacement for usual instruction or as a supplement to instruction is equally effective on students' achievement.

Of the 52 studies included, 21 (39%) studies employed individual instruction for the CAI classes, 6 (12%) studies were for small group instruction, and only 13 (24%) studies were for large group instruction. Although no significant difference was found for this variable, the differences of mean ES among studies used individual (ES = 0.557), small group (ES = 0.957), and large group (ES = 0.385) instruction were noticeable. The results suggest that CAI may be more effective if it is implemented in individual or small group settings in Taiwan.

Four types of CAI were coded for the 52 studies included, among them 20 (39%) and 25 (48%) studies were coded as traditional and multimedia CAI, respectively. Only 7 (14%) studies were coded as intelligent or web-based CAI. This result clearly displays the process of development of CAI in Taiwan: from traditional CAI to web-based CAI. The mean ESs for studies coded as traditional and multimedia were almost identical, suggesting they were equally effective on students' achievement. The highest ES associated with intelligent CAI was obvious, though, there were only 2 studies coded, the results may be considered tentative.

Conclusion

The results from this study suggest that the effects of CAI/CAL in instruction are positive over traditional instruction in Taiwan. Although many educators devote tremendous efforts with great expectation that technology will dramatically increase students' academic achievement, the results of this study provide to classroom teachers an accumulated research-based evidence for positive outcomes of using technology in instruction. Left unanswered is the question of which factors truly contribute to the positive outcomes. Studies of this question will require further clarification of the exact relationship between CAI and learning. This meta-analysis points out only that improvements of students' academic achievement are possible. That information by itself is useful.

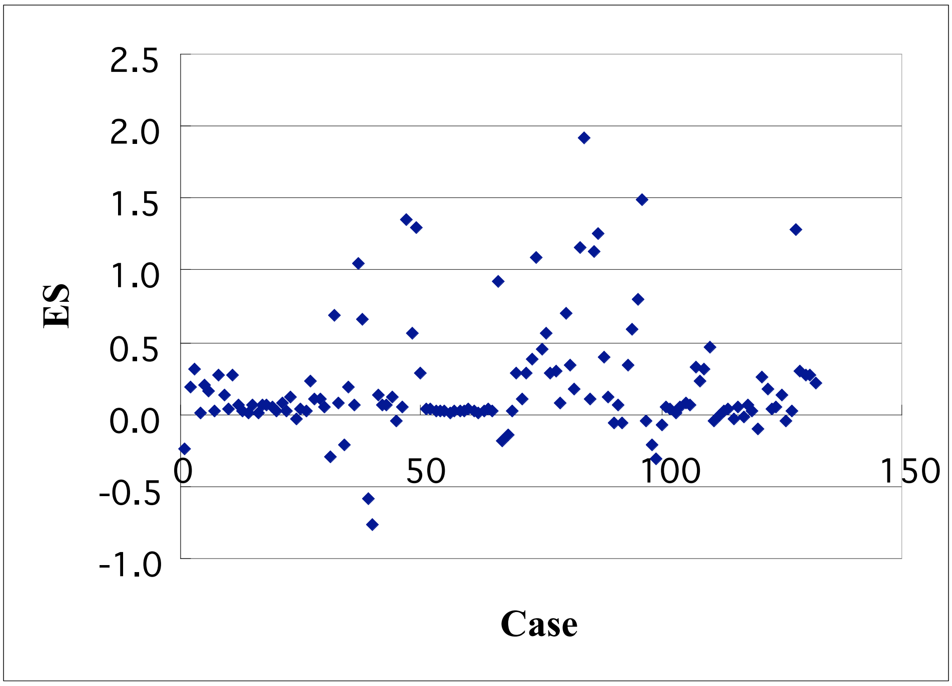

Figure 1. The Scatter Diagram of Effect Size