Marzano Research Laboratory
Powered by Solution Tree

Report on Action Research

An Analysis of the Effects of Selected Instructional Strategies on Student Achievement at Terre Haute South Vigo High School

Prepared by Marzano Research Laboratory

for

Vigo County School Corporation
Terre Haute, IN
2009-2010

May, 2010

MARZANO RESEARCH LABORATORY
www.MarzanoResearch.com
Phone: 812-336-7700 ● Fax: 866-308-3135

OUR MISSION
To provide the best research, the most useful actions, and the highest level of services to educators.

OUR VISION
To continuously develop tools that translate high quality educational research into practical applications educators can put to immediate use.

OUR GOAL
To be the place educators go for the latest information and data, synthesized into clear, concise resources that facilitate immediate action.

REPORT AUTHOR

Mark W. Haystead

For citation purposes, please refer to this document as:

Haystead, M.W. (2010). *Report on action research: An analysis of the effects of selected instructional strategies on student achievement at Terre Haute South Vigo High School*. Englewood, CO: Marzano Research Laboratory.

BUSINESS DEVELOPMENT OFFICE
555 N Morton Street
Bloomington, IN 47404
Phone 888.849.0851
Fax 812.336.7790

RESEARCH CENTER
9000 E Nichols Ave Ste 210
Centennial, CO 80112
Phone 303.766.9199
Fax 303.694.1778

www.MarzanoResearch.com

Table of Contents

Executive Summary

This report describes the findings of an analysis of a series of action research projects conducted by Vigo County School Corporation at Terre Haute South Vigo High School (hereinafter referred to as South Vigo). During the 2009-2010 school year, 20 teachers at South Vigo participated in independent action research studies regarding the extent to which selected instructional strategies enhanced the learning of students.

The following strategies were utilized in the independent action research studies (for a discussion of the research and theory regarding some of these strategies, see Marzano, 2007):

- Graphic organizers – involves providing a visual display of something being discussed or considered, e.g., using a Venn diagram to compare two items.
- Interactive games – involves use of academic content in game-like situations.
- Partial vocabulary – involves use of one or more aspects of a six step process to teaching vocabulary which may include: teacher explanation, student explanation, student graphic or pictographic representation, review using comparison activities, student discussion of vocabulary terms, and use of games. (For additional information on the six step process see Marzano, 2004, pp. 91-103.)
- Voting technology – involves interactive clicker technology to collect data regarding student knowledge during class.

Because students were not randomly assigned to experimental and control groups, all independent action research studies employed a quasi-experimental design, referred to as a pretest-posttest non-equivalent groups design. The pretest scores were used as a covariate in an analysis of covariance (ANCOVA) to partially control for differing levels of background knowledge and skill.

The following questions were considered through a random-effects meta-analysis of the independent action research studies:

1. What effect does the utilization of instructional strategies have on students' achievement regarding the subject matter content taught by their teachers?
2. Does the effect of instructional strategies differ between academic content areas?
3. Does the effect of instructional strategies differ from strategy to strategy?

The weighted average effect size estimate (Cohen's d) was not statistically significant ($d = .19$, $N = 20$, $p > .05$). When corrected for attenuation, the percentile gain associated with the use of the selected instructional strategies was 9 ($d = .23$, $N = 20$, $p > .05$). The reported average effect

sizes were positive for language arts ($d = .20$, $N = 6$, $p > .05$), mathematics ($d = .79$, $N = 2$, $p < .05$), and special education ($d = .86$, $N = 3$, $p < .01$). A comparison of the independent studies revealed a mixture of positive and negative effects for science and social studies. Finally, the reported average effect sizes were positive for graphic organizers ($d = .13$, $N = 13$, $p > .05$), interactive games ($d = .30$, $N = 7$, $p > .05$), partial vocabulary ($d = .19$, $N = 20$, $p > .05$), and voting technology ($d = .06$, $N = 3$, $p > .05$).

Introduction

This report describes the findings of an analysis of a series of action research projects conducted by Vigo County School Corporation at Terre Haute South Vigo High School (hereinafter referred to as South Vigo). During the 2009-2010 school year, 20 teachers at South Vigo participated in independent action research studies regarding the extent to which selected instructional strategies enhanced the learning of students.

The following questions were considered through a random-effects meta-analysis of the independent action research studies (see Technical Note 1):

1. What effect does the utilization of instructional strategies have on students' achievement regarding the subject matter content taught by their teachers?
2. Does the effect of instructional strategies differ between academic content areas?
3. Does the effect of instructional strategies differ from strategy to strategy?

Action Research Design

One aspect of this action research project worth mentioning is that teachers used a quasi-experimental research design in their independent action research studies. Teachers acted as their own control and were asked to collect pretest and posttest scores for two groups of students—one experimental (i.e., the group where a specific instructional strategy was used) and one control (i.e., the group where the strategy was not used). This approach was used to mitigate potential differences in teaching experience, style, etc. as possible explanations for differences between groups. Unlike experimental designs, quasi-experimental designs lack random assignment. Quasi-experimental designs are often used when it is not possible or practical to randomly assign subjects to experimental and control groups.

The quasi-experimental design used in the independent action research studies is referred to as a pretest-posttest non-equivalent groups design. The name of this design serves as a reminder that the two groups are considered to be statistically non-equivalent. In other words, the two groups may be different prior to the study and are likely to be less similar than two groups assigned through random lottery. However, an assumption was made that both groups in the independent action research studies share a similar demographic (e.g., ethnicity, socio-economic status, etc.) which would allow for a fair comparison between them.

The pretest scores were used as a covariate in an analysis of covariance (ANCOVA) to statistically equate the students and partially control for differing levels of background knowledge and skill. In basic terms, the adjustment translates the posttest scores into those that would be expected if students in both groups started with the same scores on the pretest. In effect, it is a way of controlling for students' differences in what they know about a topic prior to the beginning of instruction on the topic. ANCOVA is often used when random assignment is not feasible (see Technical Note 2). Although ANCOVA was used to make students statistically equal in terms of prior academic knowledge, arguments about causal relationships are not as strong as they would be when group membership is assigned through a random lottery.

Teachers were instructed to teach a short unit (or set of related lessons) on a topic of their choice to both groups of students—experimental and control. Detailed directions provided to teachers can be found in Appendix A. Briefly though, instructional activities in both groups were to be as similar as possible except for the fact that the instructional strategy was used in one group only (i.e., the experimental group).

The Use of Meta-Analysis

Meta-analytic techniques (see Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Cooper, 2009) were employed to aggregate the findings from the independent action research studies using the statistical software package Comprehensive Meta-Analysis (CMA), v2.2.046 (Borenstein, Hedges, Higgins, & Rothstein, 2005). Meta-analytic techniques are often used to combine the results of independent studies on a common topic.

For example, assume that multiple studies were conducted in various sites on the effects of a specific instructional technique on student achievement. The studies were different in terms of the academic content areas that were addressed (e.g., science, mathematics, etc.). Consequently, different assessments of student achievement were used to reflect the different academic content areas. This is the classic scenario requiring the use of meta-analytic techniques—independent studies on a common topic (e.g., a common instructional technique) but with different dependent measures.

To combine studies that used different dependent measures, the results of each study are translated into an effect size. While there are many types of effect sizes, the one used in this report is the standardized mean difference (Cohen's *d*). In very general terms, a standardized mean difference is the difference in the average score of the experimental group and the control group expressed in standard deviation units. Thus, an effect size of 1.00 would indicate that the average score in the experimental group is one standard deviation **higher** than the average score in the control group. Conversely, an effect size of -1.00 would indicate that the average score in the experimental group is one standard deviation **lower** than the average score in the control group.

The present meta-analysis is analogous to this situation. A common class of interventions was used in all experimental classes (i.e., use of selected instructional strategies), but the independent studies employed teacher-designed assessments of student academic achievement across various grade levels and academic content areas requiring different dependent measures.

Meta-analytic findings are typically reported in two ways:

1. Weighted average effect sizes computed from estimates of effect size for each independent study.
2. Weighted average effect sizes computed from estimates of effect size for each independent study that have been corrected for attenuation due to lack of reliability in the dependent measure (i.e., teacher-designed assessments of student academic achievement).

Technical Note 3 explains the method used to correct for attenuation and an interpretation of such corrections. Briefly though, when a dependent measure is not perfectly reliable it will tend to affect the strength of observed relationships between independent and dependent variables. An independent variable is a factor which is assumed or hypothesized to have an effect on some outcome often referred to as the dependent variable. A dependent variable is an outcome believed to be influenced by one or more independent variables. For this meta-analysis of the independent studies, the dependent variable was students' knowledge of academic content addressed during a unit of instruction and the independent variable of interest was the use of the selected instructional strategy (e.g., graphic organizers).

Estimates of effect size should always be corrected for attenuation (i.e., decrease in effect size) due to unreliability of the dependent measure (for a detailed discussion of attenuation see Hunter & Schmidt, 2004). In basic terms, every assessment is imprecise to some extent. This imprecision lowers the estimate of the true effect size. Throughout this report, observed and corrected effect sizes are displayed for comparison. When this is the case, the discussion of findings is limited to the corrected results only.

Data Analysis and Findings

As mentioned previously, in this meta-analysis one dependent variable was considered: students' knowledge of academic content addressed during a unit of instruction. The independent variable of interest was the experimental condition—whether students were exposed to an instructional strategy or not.

Data from each independent study was first analyzed with the general linear model using the statistical software package, PASW® Statistics, v17.0.2 (SPSS, 2009). One independent variable (experimental condition) was entered into the equation using a fixed-effects model (see Technical Note 4 for a discussion of fixed effects). The posttest scores were entered as the dependent variable with the pretest scores being used as a covariate. Stated differently, a fixed-effect ANCOVA was executed for each independent study. The ANCOVA findings were used to compute observed and corrected effect sizes (Cohen's *d*) for each independent study (see Technical Note 5 for a discussion regarding the formula used to compute the effect size). CMA was then used to aggregate the findings from the independent studies using the observed and corrected effect sizes for the experimental condition (i.e., use of selected instructional strategies).

Again, the following questions were considered for this report:

1. What effect does the utilization of instructional strategies have on students' achievement regarding the subject matter content taught by their teachers?
2. Does the effect of instructional strategies differ between academic content areas?
3. Does the effect of instructional strategies differ from strategy to strategy?

Findings for each question are discussed separately.

Question 1: What effect does the utilization of instructional strategies have on students' achievement regarding the subject matter content taught by their teachers?

Figure 1 presents the ANCOVA findings for each independent action research study. The column labeled "Target Strategy" contains the instructional strategy chosen by the teacher for each study. Unless otherwise indicated, the experimental group was taught using the strategy and the control group was taught without using the strategy. The columns labeled "Adjusted Mean" contain the posttest mean adjusted for differences in the pretest scores for the control and experimental groups respectively (number of students reported in parentheses). The column labeled "ES" contains the calculated effect size (Cohen's *d*) for each study, the column labeled "Sig." contains the *p*-value (2-tailed) for each study, and the column labeled "% Gain" contains the percentile gain (or loss) associated with the effect size for each study. (For a discussion of effect size and associated percentile gain see Technical Note 5.)

Figure 1: Findings for Individual Teachers

Study	Grade	Content Area	Target Strategy	Adjusted Mean (Control)	Adjusted Mean (Experimental)	ES	Sig. (2-tailed)	% Gain
1	10-12	Science	VOCAB CPS REVIEW GAME	70.15 (n=15)	67.75 (n=21)	-.18	.6090	-7
2	10	Social Studies	VOCAB GRAPHIC ORGANIZER	77.90 (n=26)	78.67 (n=19)	.07	.8257	3
3	9-12	Careers	VOCAB QUIA GAME	66.47 (n=21)	75.37 (n=27)	1.01**	.0016	34
4	10-11	Science	VOCAB GRAPHIC ORGANIZER	50.71 (n=7)	72.75 (n=8)	1.46*	.0264	43
5	10	Language Arts	VOCAB GRAPHIC ORGANIZER	66.45 (n=21)	65.90 (n=15)	-.03	.9238	-1
6	11-12	Science	VOCAB GRAPHIC ORGANIZER	78.06 (n=22)	70.78 (n=29)	-.49	.1011	-19
7	10	Language Arts (Sp Ed)	VOCAB CPS REVIEW GAME	69.03 (n=16)	69.21 (n=17)	.01	.9781	0
8		Language Arts	VOCAB CPS REVIEW GAME	74.24 (n=25)	78.43 (n=19)	.32	.3147	13
9	10	Social Studies	VOCAB NOTEBOOK	76.76 (n=25)	70.69 (n=26)	-.47	.1115	-18

Study	Grade	Content Area	Target Strategy	Adjusted Mean (Control)	Adjusted Mean (Experimental)	ES	Sig. (2-tailed)	% Gain
10	11	Language Arts	VOCAB GRAPHIC ORGANIZER	45.59 (n=29)	46.31 (n=32)	.18	.4875	7
11	9-12	Math	VOCAB GRAPHIC ORGANIZER	63.58 (n=23)	73.02 (n=34)	.58*	.0428	22
12	10-12	Science	VOCAB GRAPHIC ORGANIZER	73.50 (n=35)	75.61 (n=37)	.20	.4033	8
13	9-12	FACS	VOCAB FLASH CARDS - GAME	75.61 (n=19)	78.65 (n=19)	.29	.4000	11
14	9-12	Math (Sp Ed)	VOCAB GRAPHIC ORGANIZER	34.79 (n=17)	52.57 (n=22)	1.04**	.0038	35
15	9	Science	VOCAB QUIA GAME	63.28 (n=23)	66.70 (n=35)	.33	.2313	13
16	11	Language Arts (Sp Ed)	VOCAB GRAPHIC ORGANIZER	61.53 (n=10)	81.07 (n=10)	1.91**	.0011	47
17	9-12	Science	VOCAB GRAPHIC ORGANIZER	86.19 (n=15)	81.83 (n=17)	-.35	.3528	-14
18	10	Language Arts	VOCAB NOTEBOOK	59.58 (n=15)	53.21 (n=20)	-.56	.1259	-21
19	9-12	Spanish	VOCAB FLASH CARDS - GAME	77.94 (n=16)	81.25 (n=20)	.28	.4293	11
20	10-11	Science	VOCAB GRAPHIC ORGANIZER	64.47 (n=11)	56.34 (n=26)	-.62	.1067	-23

* $p < .05$; ** $p < .01$.

Figure 1 presents the findings for 20 independent action research studies. When considering the information displayed in the figure, it should be noted that the data for each study was checked for obvious coding errors, negative gains, and other potential outliers. Typically, when negative gains are excluded one has a better sense of the uniform effects of the experimental condition (i.e., use of the selected instructional strategy). In other words, assuming that students learn more about academic content during a unit of instruction, it would not make sense for a student to know less about the academic content at the end of the unit. “Learning theory and common sense tell us that a student might start a grading period with little or no knowledge regarding a topic but end the grading period with a great deal of knowledge” (Marzano, 2006, pp. 96-97). As such, students who scored higher on the pretest were excluded from analysis.

The last three columns are of considerable interest. Again, the column labeled “ES” contains the computed effect size (Cohen’s *d*) for each independent action research study. This value is the difference between the adjusted means of the experimental and control groups expressed in standard deviation units. The column labeled “Sig.” contains the *p*-value (2-tailed) for each study. In social science research and evaluation it is common practice to consider any contrast with a *p*-value less than .05 as “statistically significant” (see Murphy & Myors, 2004). The column labeled “% Gain” contains the percentile gain (or loss) in achievement associated with the experimental condition (i.e., use of the selected instructional strategy). The values in this column were determined by consulting a normal curve table for the area for each reported effect size.

To understand the interpretation of the standardized mean difference effect size, consider the results reported for Study 3 in the third row of Figure 1. The adjusted mean for the control group is 66.47 (n = 21) and the adjusted mean for the experimental group is 75.37 (n = 27). The percentile gain for this study is 34 ($d = 1.01$). This means that the average score in the experimental group is 1.01 standard deviations or 34 percentile points **greater than** the average score in the control group. It should be noted that in some cases the reported percentile gain is negative. This occurs when the adjusted mean for the experimental group is **less than** the adjusted mean for the control group. For example, the percentile “gain” reported in the first row is negative 7 (-7, $d = -.18$). This means that the average score in the control group is .18 standard deviations or 7 percentile points greater than the average score in the experimental group. (For a discussion regarding interpretation of effect size, see Lipsey & Wilson, 2001; Cooper, 2009.)

Figure 1 indicates that the comparison between experimental and control groups can be considered statistically significant for five independent action research studies at the .05 level ($p <. 05$). Of those five studies, three can be considered statistically significant at the .01 level ($p < .01$). In other words, 15 out of 20 studies (or 75%) would not be considered statistically significant. This is quite common in educational research where many individual studies might be deemed non-significant simply because they do not have enough subjects in the experimental and control groups (for a detailed discussion see Hedges & Olkin, 1985).

Figure 2 shows the overall average effect size estimate (Cohen's *d*) for a meta-analysis of the 20 independent action research studies using a random-effects model of error (see Technical Note 1 for discussion of fixed-effect vs. random-effects meta-analysis). The column labeled "N" identifies the number of studies included in the group, the column labeled "$\overline{ES}$" reports the weighted average effect size (Cohen's *d*) for the studies, the column labeled "SE" contains the standard error for the reported weighted average effect size, the column labeled "95% CI" identifies the 95 percent confidence interval (lower limit and upper limit) for the reported weighted average effect size, the column labeled "Sig." reports the *p*-value (2-tailed) for the reported weighted average effect size, the column labeled "% Gain" contains the percentile gain (or loss) associated with the reported weighted average effect size, and the column labeled "Fail-Safe N" identifies the number of missing studies that would be required to reduce the weighted average effect size to .01 using Orwin's formula (for a discussion of sampling bias and the fail-safe N, see Lipsey & Wilson, 2001, pp. 165-166).

Figure 2. Overall Random Effects

	N	$\overline{ES}$	**SE**	**95% CI**		**Sig. (2-tailed)**	**% Gain**	**Fail-Safe N**
				LL	**UL**			
Overall	20	.19 (.23)	.12 (.14)	-.05 (-.05)	.44 (.50)	.1164 (.1061)	8 (9)	360 (440)

Note: Corrected findings are presented in parentheses.

When the results of the 20 independent action research studies are corrected for attenuation and combined, the estimated overall percentile gain is 9 (d = .23). This means that on the average, the instructional strategies used in the independent action research studies represent a gain of 9 percentile points over what would be expected if teachers did not use the instructional strategies (for a discussion of how effect sizes are combined and an overall significance level is computed see Lipsey & Wilson, 2001).

Consider the fail-safe N reported in parentheses, 440. This means that 440 additional independent studies with an effect size of .00 (i.e., no difference between groups) would be needed to reduce the weighted average effect size from .23 to .01. The percentile gain associated with an effect size of .01 is 0.

The column labeled "95% CI" contains the 95 percent confidence interval for the reported weighted average effect size. Again, the effect size reported in Figure 2 is a weighted average of all the effect sizes from the 20 independent action research studies (see Figure 1). As such, it is considered an estimate of the true effect size of the experimental condition (i.e., use of selected instructional strategies). This interval includes the range of effect sizes in which one can be 95 percent certain the true effect size falls. For example, consider the confidence interval reported in parentheses. There is a

95 percent certainty that the true effect size for the meta-analysis of the 20 independent action research studies is between the values of -.16 and .38. When the confidence interval does not include .00, the weighted average effect size estimate is determined to be statistically significant at the .05 level ($p < .05$). In other words, $d = .00$ would not be considered a reasonable assumption. (For a detailed discussion of the meaning of statistical significance, see Harlow, Muliak, & Steiger, 1997.)

Another way to examine the general effect of the instructional strategies is to consider the distribution of observed effect sizes and percentile gain associated with the effect sizes. Figures 3 and 4 present the distribution of observed effect sizes and associated percentile gain as generated by PASW® Statistics.

Figure 3. Distribution of Observed Effect Sizes

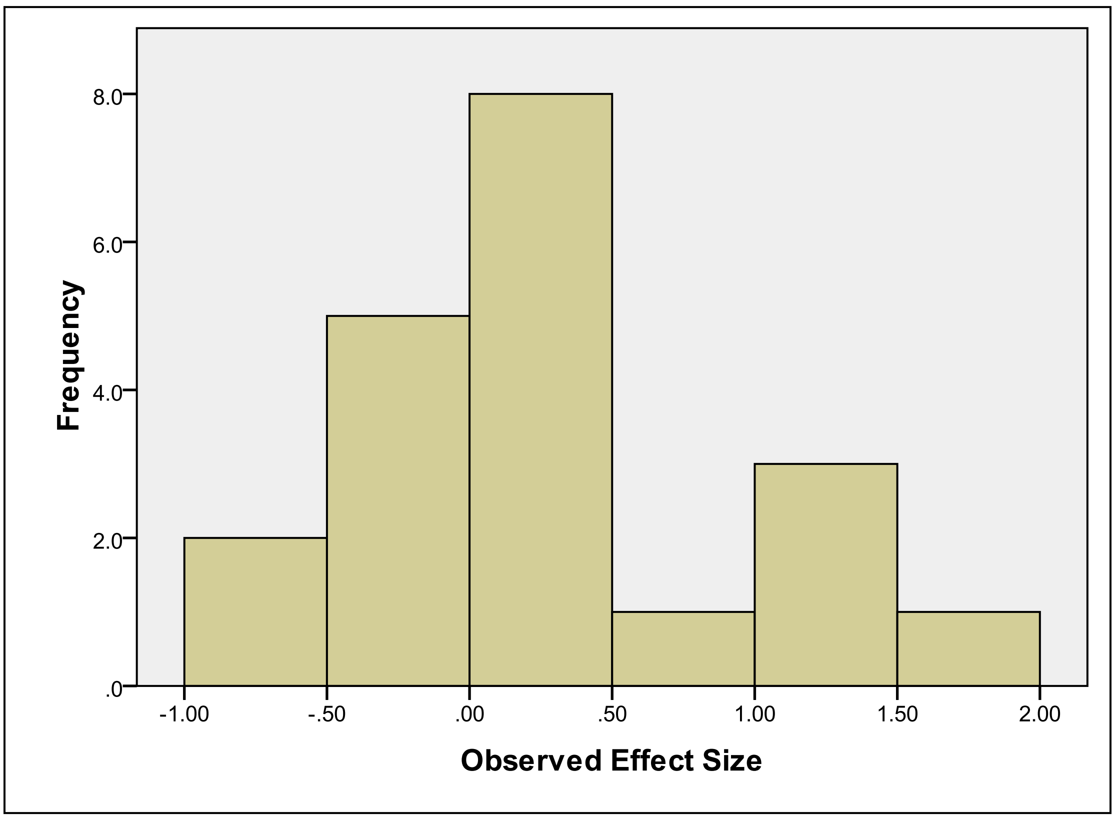


Figure 3 displays the distribution of "groups" of observed effect sizes across the 20 independent action research studies reported in Figure 1 ($M = .25$, $SD = .68$, $Min = -.62$, $Max = 1.91$). Seven studies exhibited a negative effect (see first two columns), nine studies exhibited an effect size between .00 and 1.00 (see third and fourth columns), and four studies exhibited an effect size between 1.00 and 2.00 (see last two columns). So, 13 out of 20 studies (or 65%) have a positive effect size.

Figure 4. Distribution of Percentile Gains

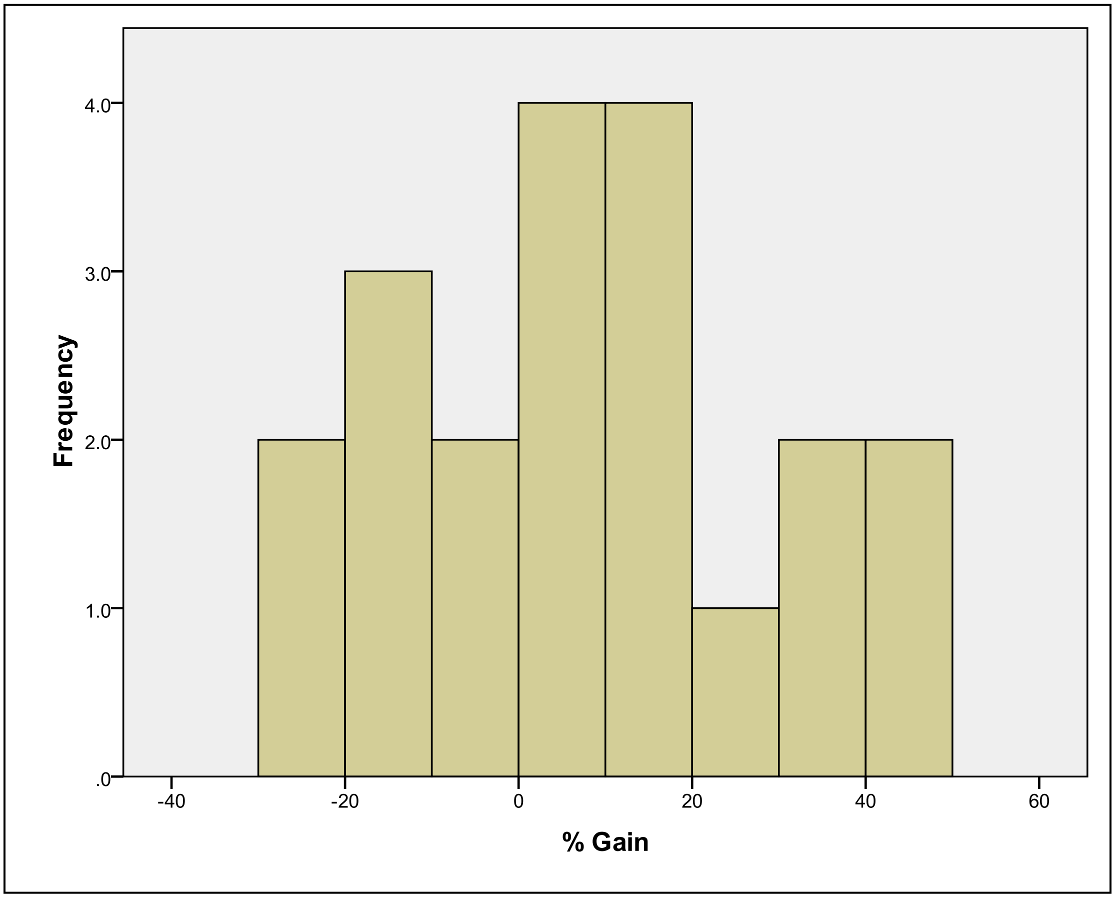


Figure 4 displays the distribution of “groups” of percentile gains across the 20 independent action research studies reported in Figure 1 (*M* = 7.20, *SD* = 21.14, *Min* = -23, *Max* = 47). Seven studies exhibited a percentile gain that is negative (see first three columns), eight studies exhibited a percentile gain between 0 and 20 (including 0, see fourth and fifth columns), three studies exhibited a percentile gain between 20 and 40 (see sixth and seventh columns), and two studies exhibited a percentile gain greater than 40 (see last column).

Question 2: Does the effect of instructional strategies differ between academic content areas?

For this question, the findings for the 20 independent action research studies were aggregated based upon the reported academic content area. Figure 5 reports the overall random effects for four content areas: language arts, mathematics, science, and social studies.

Figure 5. Random Effects for Academic Content Area

Content Area	N	$\overline{ES}$	SE	95% CI		Sig. (2-tailed)	% Gain
				LL	UL		
Language Arts[a] (Study: 5,7,8,10,16,18)	6	.20 (.24)	.22 (.25)	-.23 (-.25)	.63 (.72)	.3628 (.3410)	8 (10)
Mathematics[a] (Study: 11,14)	2	.79* (.92)*	.36 (.41)	.08 (.10)	1.50 (1.73)	.0293 (.0269)	29 (32)
Science (Study: 1,4,6,12,15,17,20)	7	-.03 (-.03)	.20 (.23)	-.42 (-.47)	.37 (.42)	.8946 (.9071)	-1 (-1)
Social Studies (Study: 2,9)	2	-.21 (-.23)	.36 (.41)	-.90 (-1.03)	.49 (.56)	.5639 (.5625)	-8 (-9)

Note: See discussion of Figure 2 for a description of column headings. Corrected findings are presented in parentheses. a) Special Education classes included. * $p < .05$.

Figure 5 depicts the findings for a random-effects meta-analysis for language arts, mathematics, science, and social studies. The weighted average effect size reported in parentheses is statistically significant at the .05 level ($p < .05$) for mathematics ($d = .92$, $N = 2$). The percentile gain is positive for language arts and mathematics. It is worth noting that the weighted effect sizes for language arts and mathematics include special education classes (i.e., classes taught by teachers in the special education department). Figure 6 displays the findings for a random-effects meta-analysis for language arts, mathematics, and special education.

Figure 6. Random Effects for Academic Content Area

Content Area	N	$\overline{ES}$	SE	95% CI		Sig. (2-tailed)	% Gain
				LL	UL		
Language Arts (Study: 5,8,10,18)	4	-.00 (-.01)	.26 (.30)	-.51 (-.59)	.50 (.57)	.9903 (.9772)	0 (0)
Mathematics (Study: 11)	1	.58 (.66)	.49 (.57)	-.39 (-.46)	1.54 (1.78)	.2435 (.2449)	22 (25)
Special Education (Study: 7,14,16)	3	.86** (1.00)**	.33 (.37)	.21 (.26)	1.51 (1.73)	.0095 (.0077)	31 (34)

Note: See discussion of Figure 2 for a description of column headings. Corrected findings are presented in parentheses. ** $p < .01$.

A comparison between Figures 5 and 6 indicates that two independent studies in special education involved language arts and one independent study in special education involved mathematics. The weighted average effect size reported in parentheses is statistically significant at the .01 level ($p < .01$) for special education ($d = 1.00$, $N = 3$). The percentile gain is positive for mathematics and special education. The weighted average effect size reported in parentheses for language arts changed from .24 to -.01 ($N = 4$, $p > .05$) when the special education classes were filtered out. Similarly, the effect size for mathematics changed from .92 to .66 ($N = 1$, $p > .05$).

Figures 7 and 8 graphically depict the average percentile gains reported in parentheses from Figures 5 and 6.

Figure 7. Average Percentile Gains for Academic Content Area (Corrected)

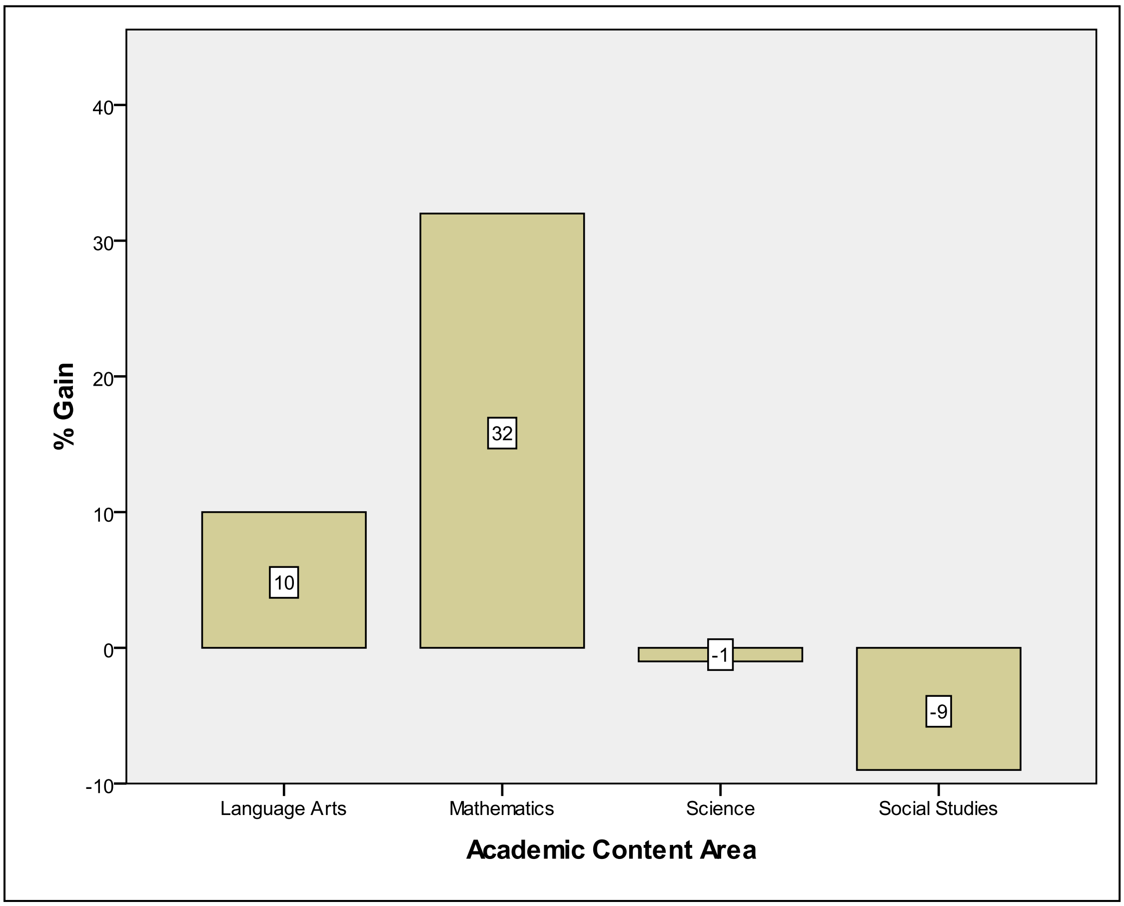
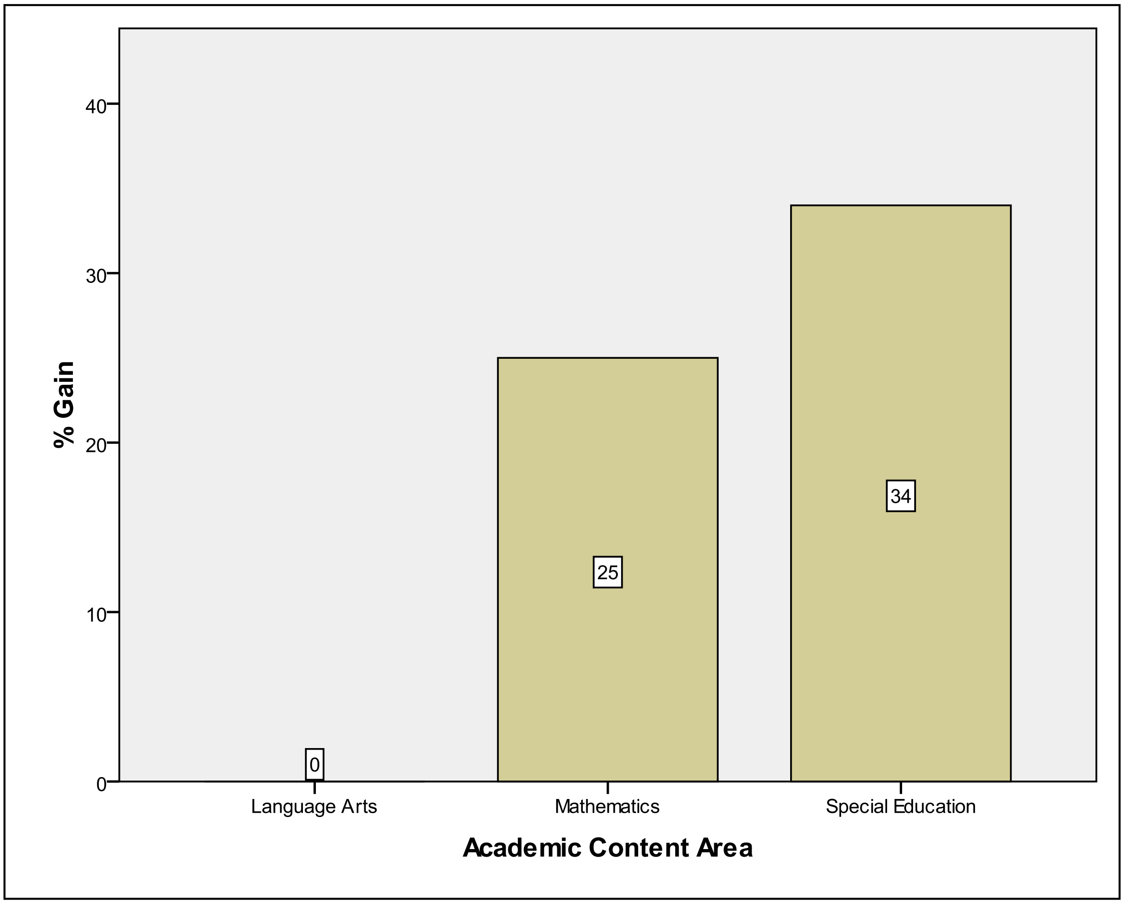


Figure 8. Average Percentile Gains for Academic Content Area (Corrected)



While a weighted average effect size can help quantify the estimate of the true effect size for each academic content area it is often helpful to display findings in different ways. Doing so allows further discussion of any patterns that may be apparent.

Figure 9 depicts the number of independent studies in language arts, mathematics, science, social studies, and special education with a positive gain, negative gain, and no gain.

Figure 9. Positive and Negative Gains for Academic Content Area

Content Area	% Gain		
	Positive Gain	**Negative Gain**	**No Gain**
Language Arts (Study: 5,7,8,10,16,18)	3	2	1
Mathematics (Study: 11,14)	2	--	--
Science (Study: 1,4,6,12,15,17,20)	3	4	--
Social Studies (Study: 2,9)	1	1	--
Special Education (Study: 7,14,16)	2	--	1

Figure 9 indicates that the number of positive and negative gains is close to an even split for language arts, science, and social studies. Again, the independent studies in special education are also counted in language arts and mathematics.

Figures 10 through 14 graphically depict the percentile gains from Figure 1 for language arts, mathematics, science, social studies, and special education.

Figure 10. Percentile Gains for Language Arts

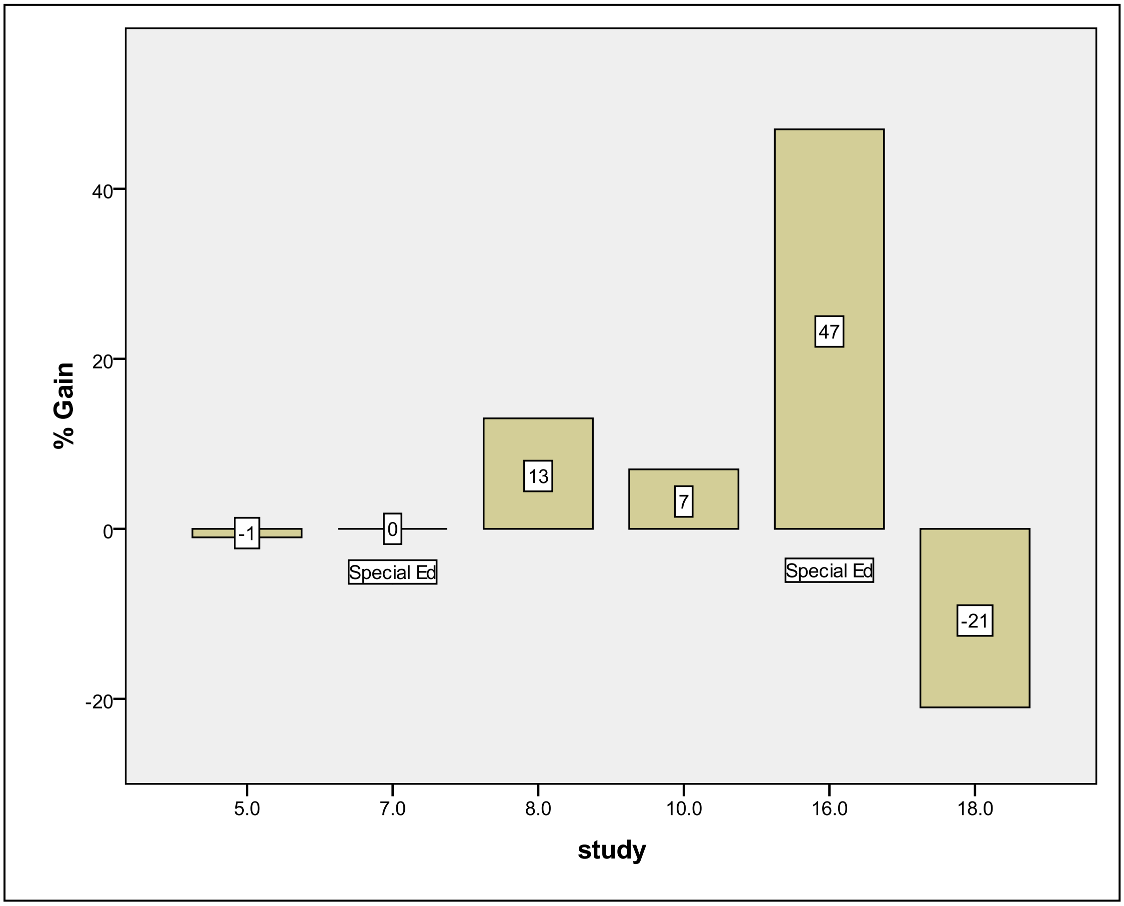


Figure 10 shows that in language arts, Study 16 exhibited a positive gain greater than 40 percentile-points and Studies 8 and 10 exhibited a positive gain around 10 percentile-points. Additionally, Studies 5 and 7 exhibited almost no gain and Study 18 exhibited a negative gain greater than 20 percentile-points.

Figure 11. Percentile Gains for Mathematics

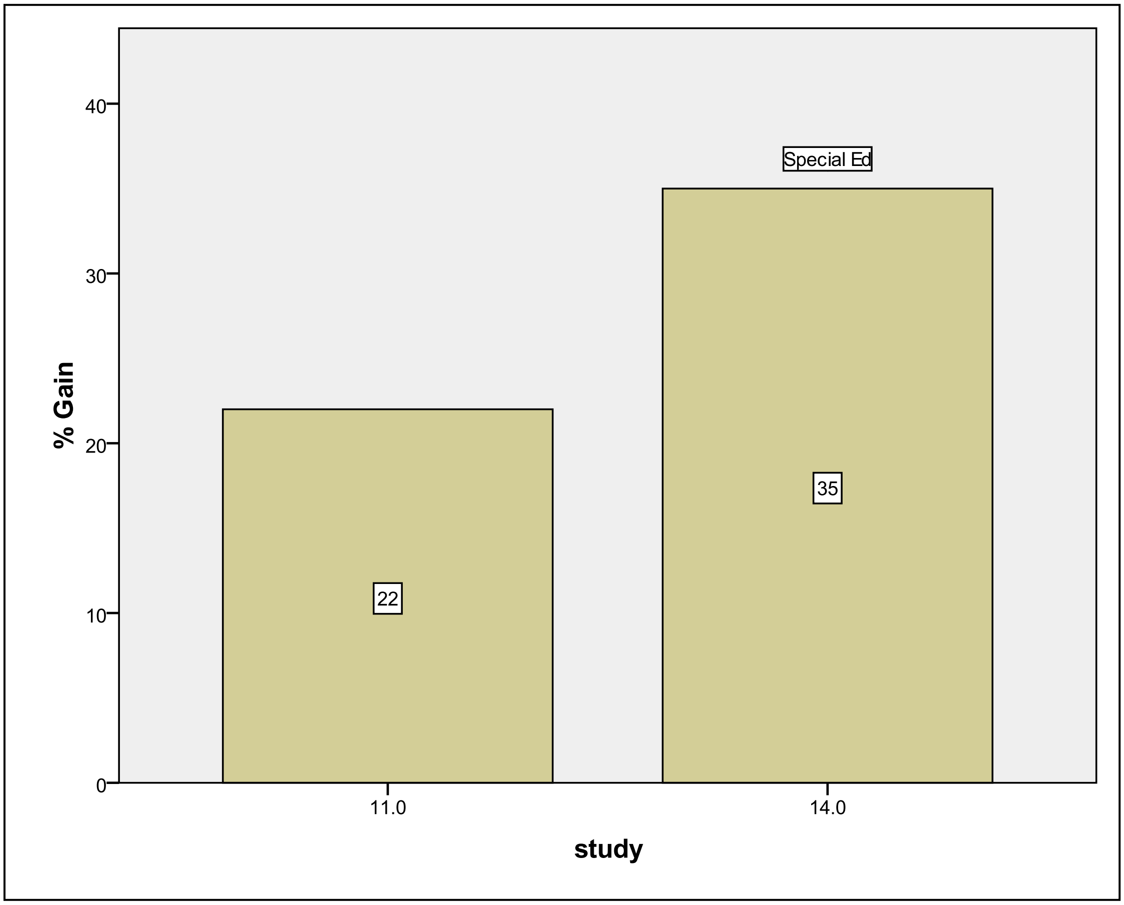


Figure 11 shows that in mathematics, Studies 11 and 14 exhibited a positive gain greater than 20 percentile-points.

Figure 12. Percentile Gains for Science

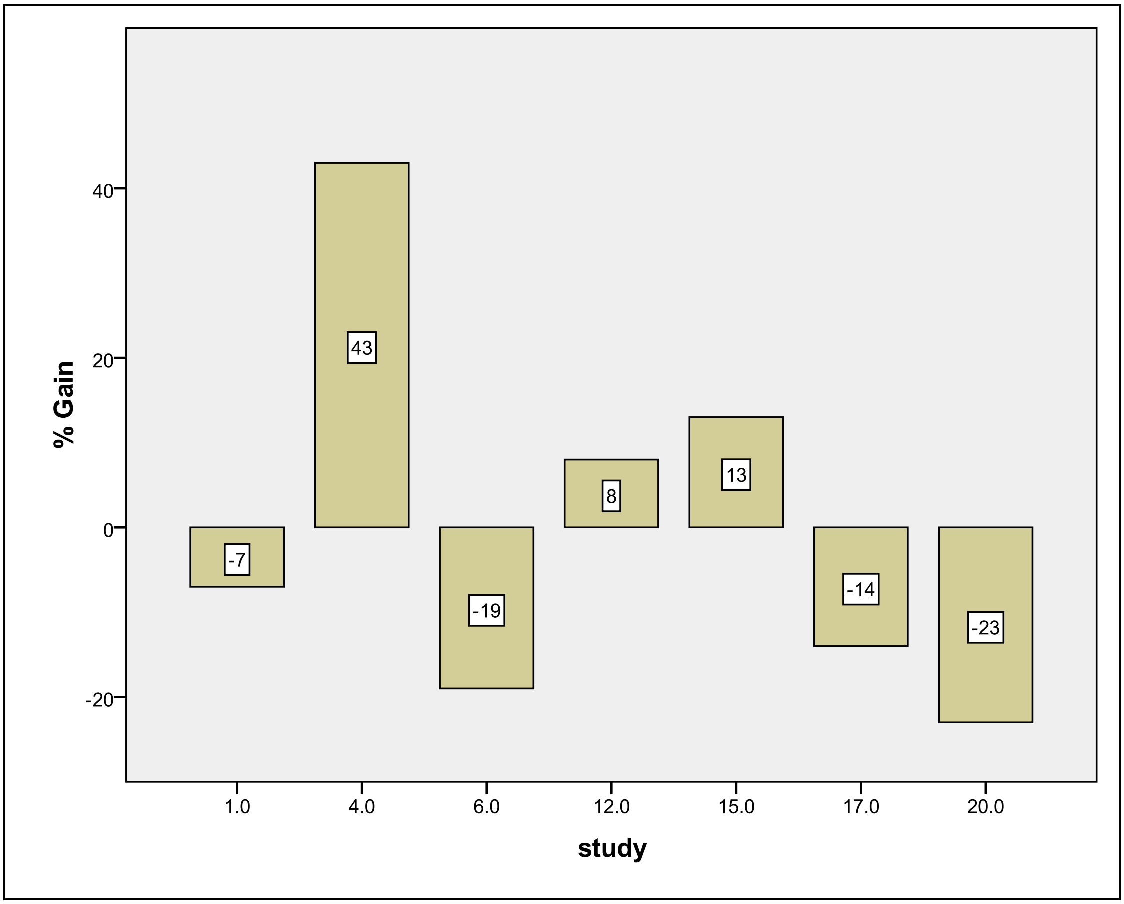


Figure 12 shows that in science, Study 4 exhibited a positive gain greater than 40 percentile-points and Studies 12 and 15 exhibited a positive gain around 10 percentile-points. Four studies exhibited negative gains ranging from -7 to -23 (Studies 1, 6, 17, and 20).

Figure 13. Percentile Gains for Social Studies

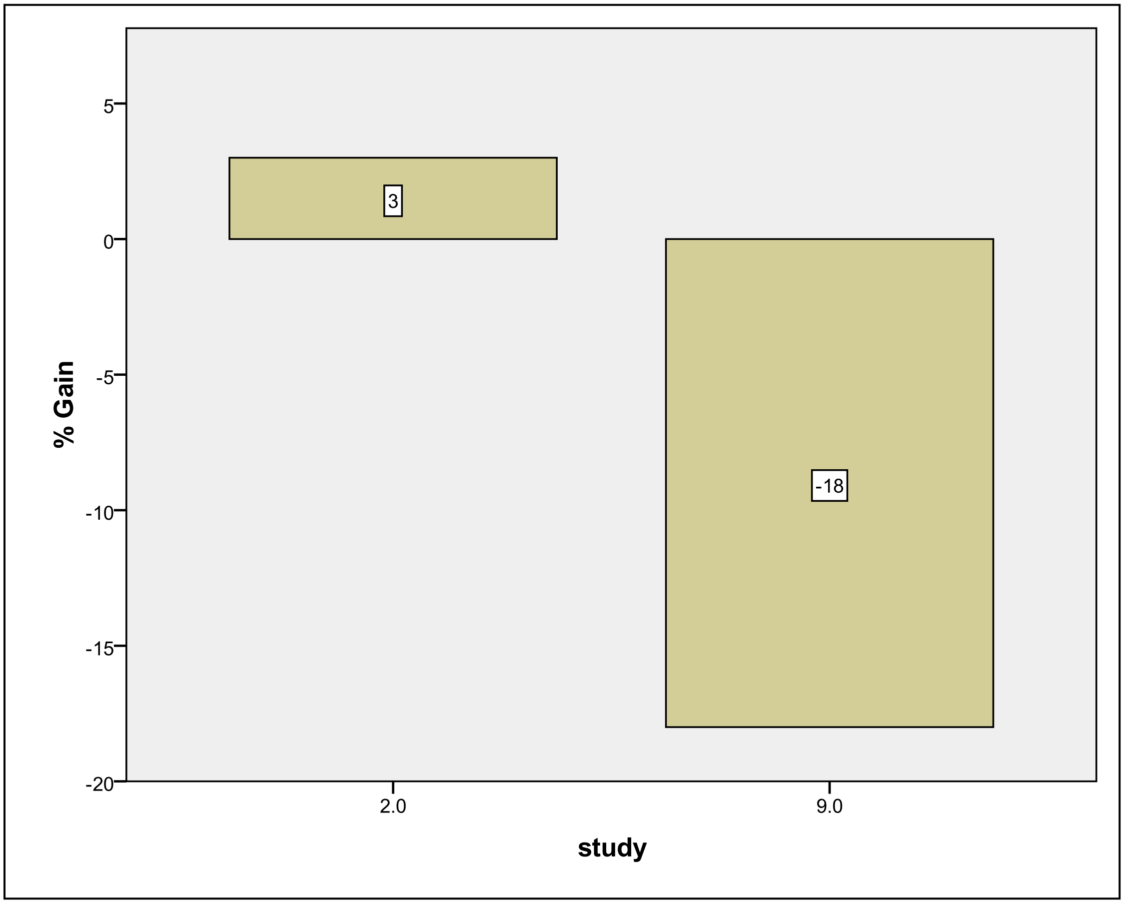


Figure 13 shows that in social studies, Study 2 exhibited a slight positive gain. However, this was offset by Study 9 which exhibited a negative gain greater than 15 percentile-points.

Figure 14. Percentile Gains for Special Education

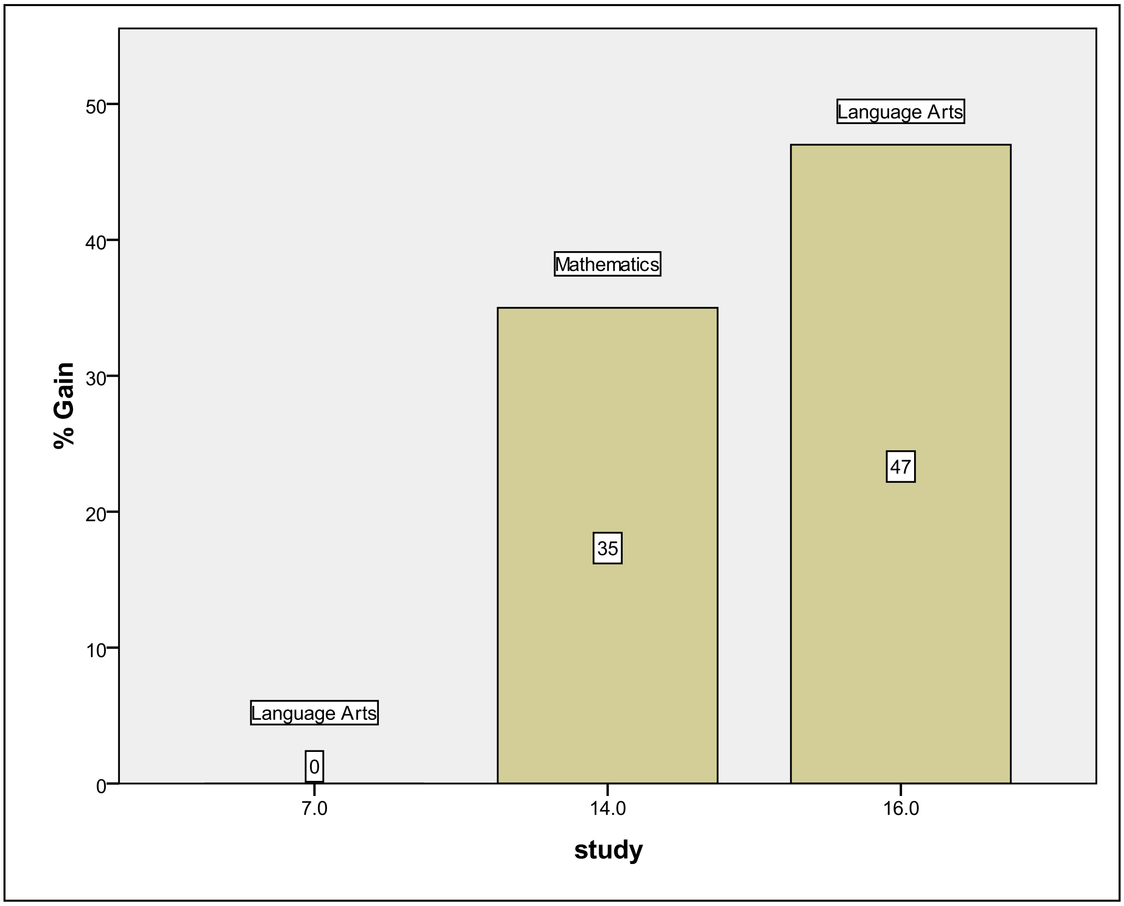


Figure 14 shows that in special education, Studies 14 and 16 exhibited a positive gain greater than 30 percentile-points, while Study 7 exhibited no gain.

Taken at face value, Figures 10 through 14 indicate that the instructional strategies were quite effective for one teacher in language arts (Study 16 [special education]), two teachers in mathematics (Studies 11 and 14 [special education]), and one teacher in science (Study 4) with positive gains greater than 20 percentile-points. The strategies were somewhat effective for two teachers in language arts (Studies 8 and 10) and two teachers in science (Studies 12 and 15) with positive gains between 5 and 15 percentile-points. The strategy was minimally effective for one teacher in social studies (Study 2) with a positive gain between 0 and 5. For the remaining teachers the instructional strategies either produced no effect or were less effective during the unit of instruction considered in the independent action research studies.

These findings should be considered with caution due to the small number of studies in the meta-analysis. For example, Haystead and Marzano (2009) reported a percentile gain of 16 (d = .42, p < .0001) in their meta-analysis of 329 independent action research studies. Their findings suggest that more often than not, the instructional strategies are effective in the classroom. If more studies are conducted in these content areas, the findings reported here would most likely change.

Question 3: Does the effect of instructional strategies differ from strategy to strategy?

Question 1 considered the general effects of instructional strategies across all of the 20 independent action research studies. Question 2 reported aggregate findings based on academic content area. Another useful way to aggregate the findings is by the reported target instructional strategy. Figure 15 displays the findings for the four strategies listed below (for a discussion of the research and theory regarding some of these strategies, see Marzano, 2007).

- Graphic organizers – involves providing a visual display of something being discussed or considered, e.g., using a Venn diagram to compare two items.
- Interactive games – involves use of academic content in game-like situations.
- Partial vocabulary – involves use of one or more aspects of a six step process to teaching vocabulary which may include: teacher explanation, student explanation, student graphic or pictographic representation, review using comparison activities, student discussion of vocabulary terms, and use of games. (For additional information on the six step process see Marzano, 2004, pp. 91-103.)
- Voting technology – involves interactive clicker technology to collect data regarding student knowledge during class.

Figure 15. Random Effects for Specific Instructional Strategies

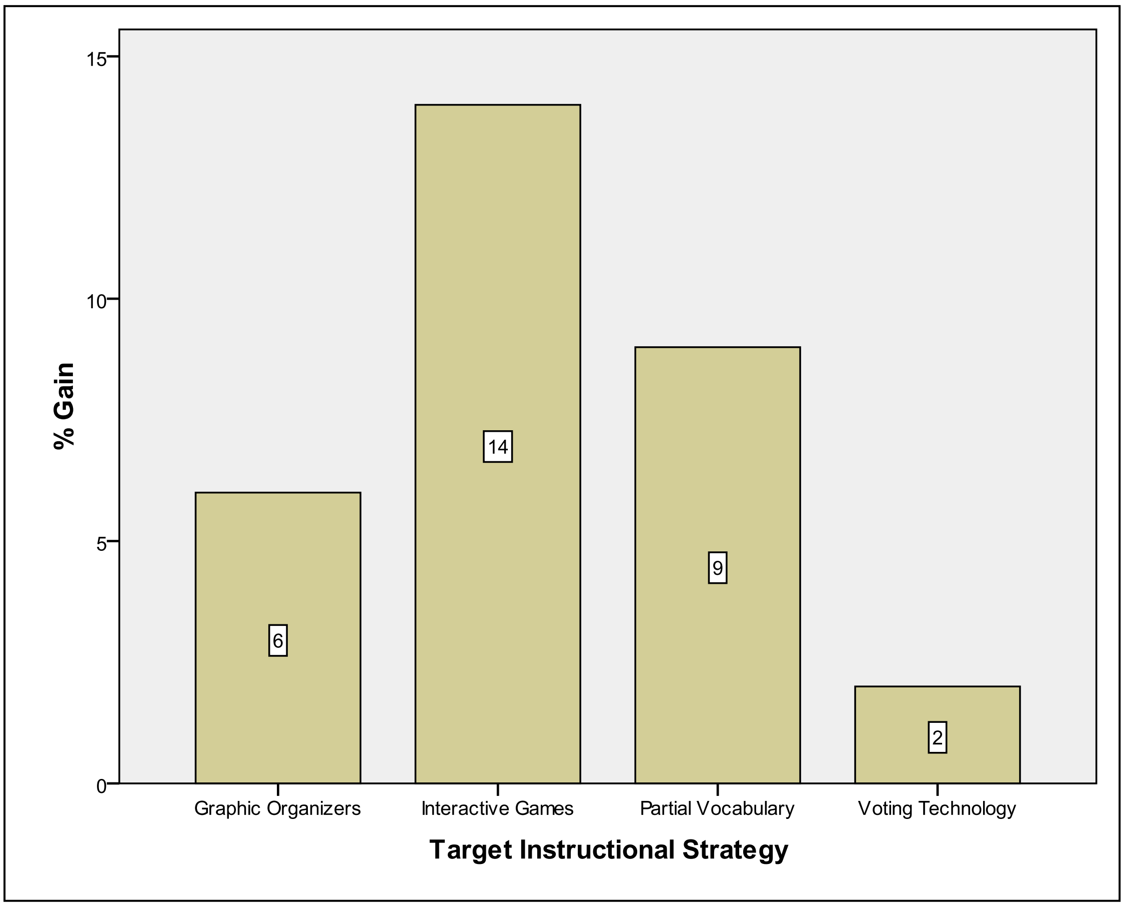
Target Strategy	N	$\overline{ES}$	SE	95% CI		Sig. (2-tailed)	% Gain
				LL	UL		
Graphic Organizers (Study: 2,4,5,6,9,10 11,12,14,16,17,18,20)	13	.13 (.16)	.15 (.17)	-.17 (-.18)	.43 (.50)	.3896 (.3616)	5 (6)
Interactive Games (Study: 1,3,7,8,13,15,19)	7	.30 (.35)	.20 (.23)	-.09 (-.10)	.70 (.80)	.1338 (.1319)	12 (14)
Partial Vocabulary (all studies)	20	.19 (.23)	.12 (.14)	-.05 (-.05)	.44 (.50)	.1123 (.1022)	8 (9)
Voting Technology (Study: 1,7,8)	3	.06 (.06)	.31 (.36)	-.56 (-.63)	.67 (.76)	.8563 (.8579)	2 (2)

Note: See discussion of Figure 2 for a description of column headings. Corrected findings are presented in parentheses.

Figure 15 lists the findings for a random-effects meta-analysis for the four instructional strategies. Some of the independent action research studies were included in the meta-analysis for more than one strategy. This occurred when teachers reported using a strategy which incorporates elements from more than one of the strategies listed above. The weighted average effect sizes reported in parentheses are not statistically significant ($p > .05$). The associated percentile gain is positive for all four instructional strategies.

Figure 16 graphically depicts the average percentile gains reported in parentheses from Figure 15.

Figure 16. Average Percentile Gains for Specific Instructional Strategies (Corrected)



These findings are lower than those reported by Haystead and Marzano (2009). However, this does not mean that these strategies are less effective. Rather, it may simply be a function of the number of independent studies. If more studies are conducted on the use of these instructional strategies, the findings reported here would most likely change.

Again, it is often useful to display numeric findings in different ways for comparison. Figure 17 lists the number of independent studies for each instructional strategy with a positive gain, negative gain, and no gain. Figures 18 through 21 graphically depict the percentile gains from Figure 1 for graphic organizers, interactive games, partial vocabulary, and voting technology.

Figure 17. Positive and Negative Gains for Specific Instructional Strategies

Content Area	**% Gain**		
	Positive Gain	**Negative Gain**	**No Gain**
Graphic Organizers (Study: 2,4,5,6,9,10 11,12,14,16,17,18,20)	7	6	--
Interactive Games (Study: 1,3,7,8,13,15,19)	5	1	1
Partial Vocabulary (all studies)	12	7	1
Voting Technology (Study: 1,7,8)	1	1	1

Figure 17 indicates that 5 out of 7 (or 71%) of the independent studies utilizing interactive games exhibited a positive gain. All of the independent studies involved vocabulary. 12 out of the 20 studies (or 60%) exhibited a positive gain. The number of positive and negative gains is close to an even split for graphic organizers and voting technology.

Figure 18. Percentile Gains for Graphic Organizers

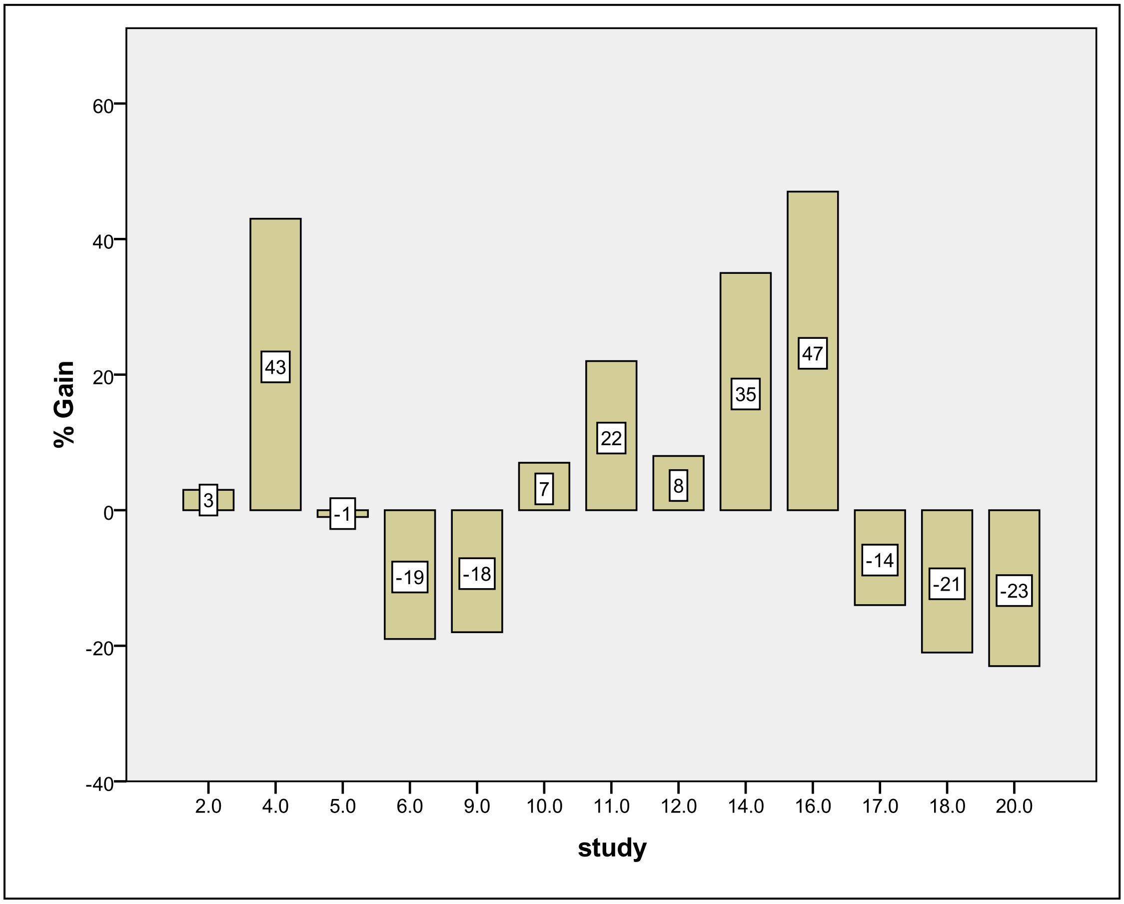


Figure 18 shows that for graphic organizers, the percentile gains ranged from -23 to 47. Studies 4, 11, 14, and 16 exhibited a positive gain greater than 20 percentile-points and Studies 6, 9, 18, and 20 exhibited a negative gain around 20 percentile-points.

Figure 19. Percentile Gains for Interactive Games

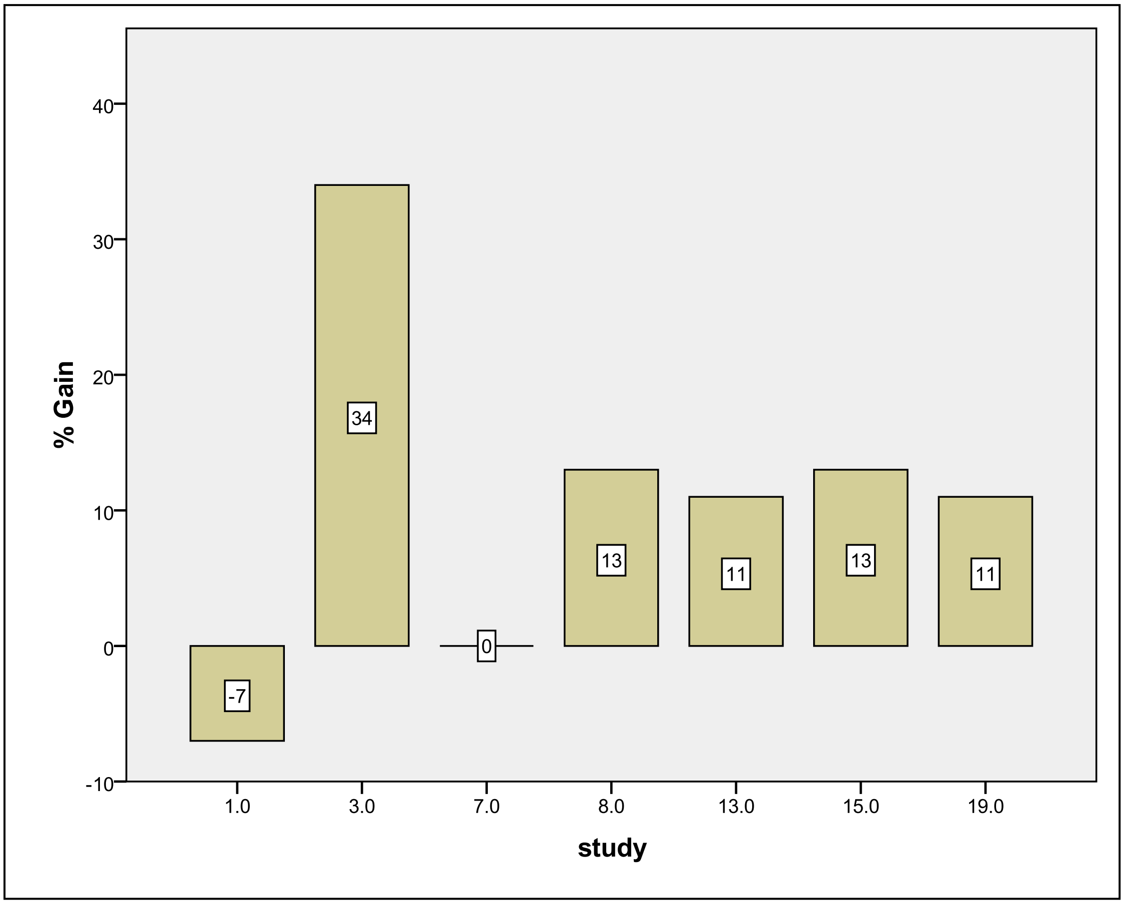


Figure 19 shows that for interactive games, the percentile gains ranged from -7 to 34. Study 3 exhibited a positive gain greater than 30 percentile-points and Studies 8, 13, 15, and 19 exhibited a positive gain around 10 percentile-points.

Figure 20. Percentile Gains for Partial Vocabulary

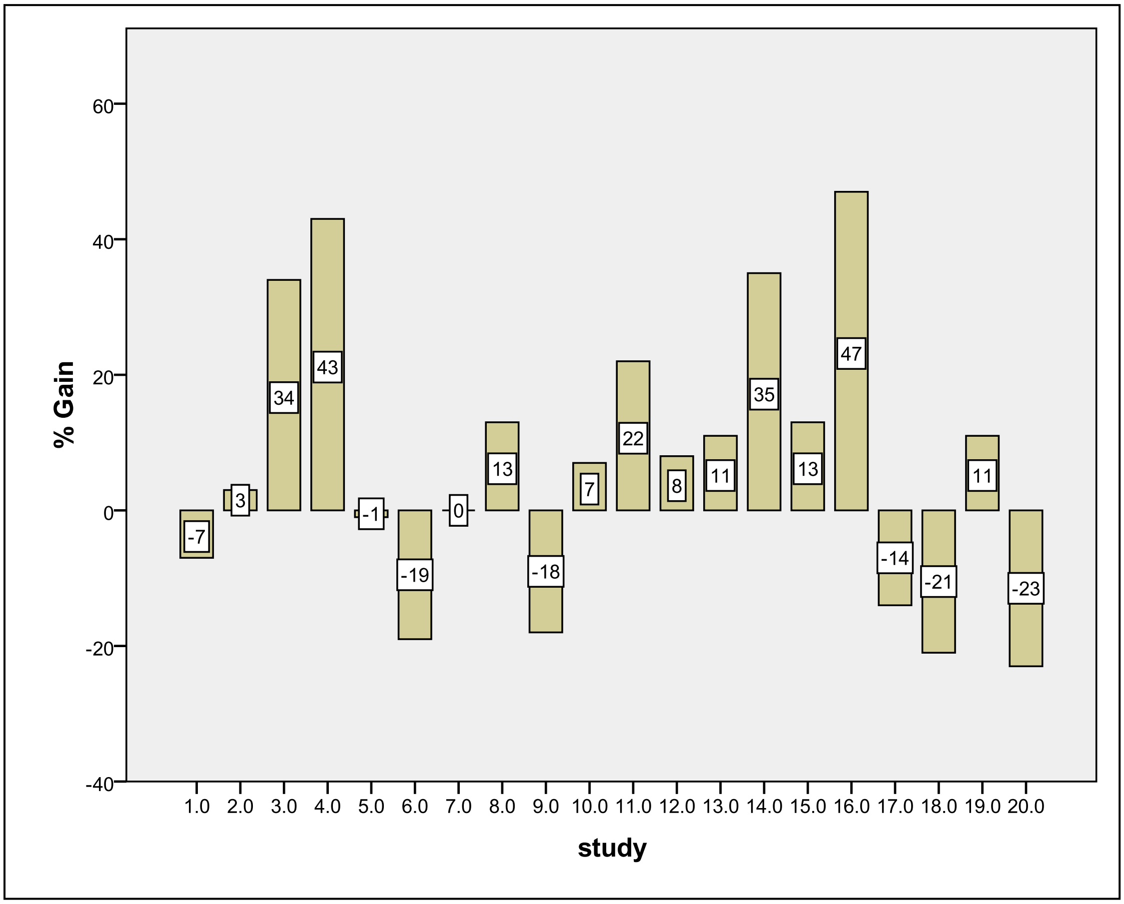


Figure 20 shows the percentile gains for all 20 studies. The positive percentile gains ranged from 3 to 47. When compared with Figures 10 through 14, it is apparent that two of the studies with negative gains were from a language arts classroom (Studies 5 and 18), four were from science classrooms (Studies 1, 6, 17, and 20), and one was from a social studies classroom (Study 9).

Figure 21. Percentile Gains for Voting Technology

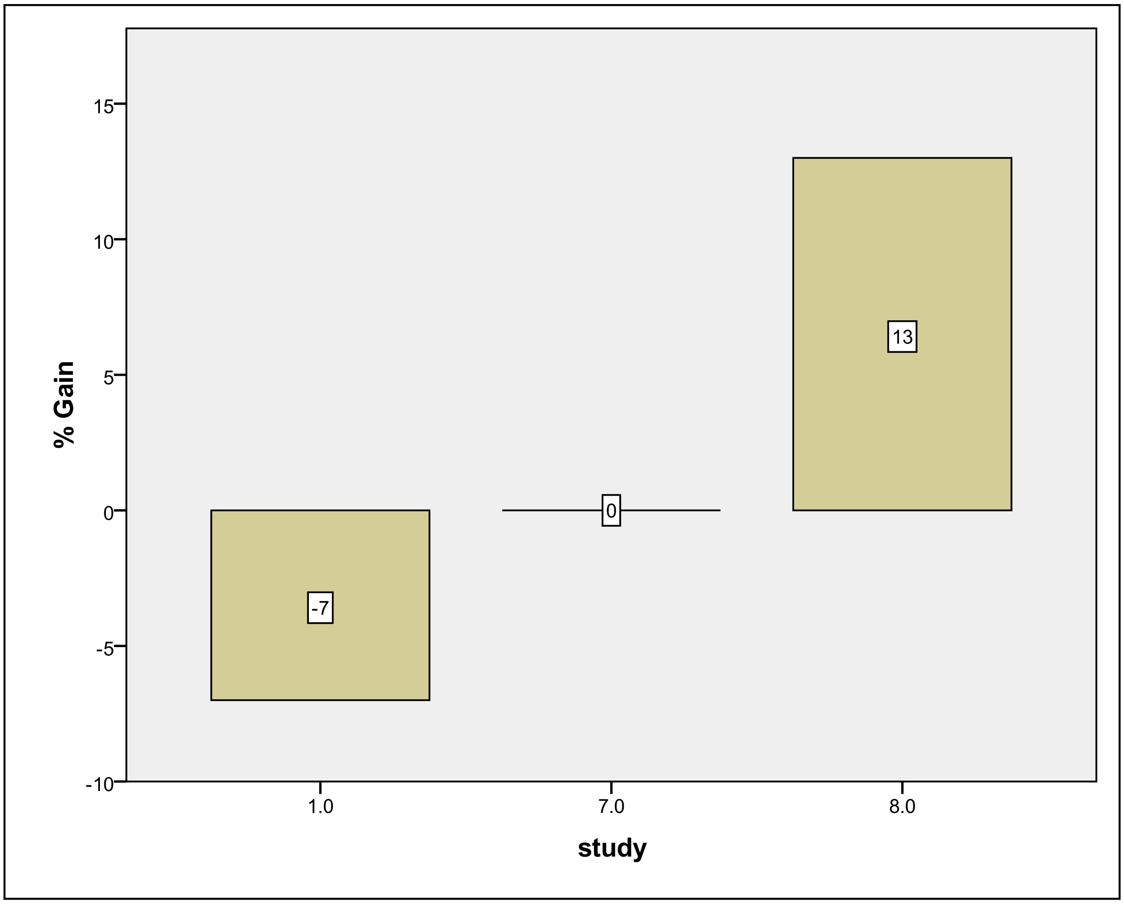


Figure 21 shows that for voting technology the percentile gains ranged from -7 to 13. Study 8 exhibited a positive gain, Study 1 exhibited a negative gain, and Study 7 exhibited no gain.

Interpretation

There are a number of ways to interpret an effect size (Cohen's *d*). One interpretation is the amount of overlap between the experimental and control groups. Consider again that an effect size of 1.00 can be interpreted as the average score in the experimental group being one standard deviation higher than the average score in the control group. Consulting a table of the normal curve (i.e., normal distribution) the associated percentile gain for an effect size of 1.00 is 34. This means that the score of the average student in the experimental group (50^{th} percentile) exceeds the scores of 84 percent of the control group. In other words, only 16 percent of the control group would be expected to have scores that exceed the score of the average student in the experimental group.

Figure 22 depicts the overall findings for the 20 independent action research studies based on this interpretation. It shows the percentage of control group students who scored lower than the average student in the experimental group (50^{th} percentile). When corrected for attenuation, the average student in the experimental group (i.e., the group that used an instructional strategy) scored higher than 59% of the students in the control group (i.e., the group that did not use an instructional strategy).

Figure 22. Amount of Overlap between Experimental and Control Groups

	$\overline{ES}$	**Percentage of Control Group Scoring Lower than Experimental Average (50^{th} Percentile)**
Overall	.19 (.23)	58% (59%)

Note: Corrected findings are presented in parentheses.

Another interpretation is to consider the hypothetical change in rank for a class with 100 students. Figure 23 displays the corrected findings based on this interpretation.

Figure 23. Hypothetical Change in a Student's Class Rank

Class Rank Without Instructional Strategies	Average Student's Class Rank With Instructional Strategies
1	1
5	5
10	10
15	15
20	20
25	25
30	30
35	35
40	**41**
45	45
50	50
55	55
60	60
65	65
70	70
75	75
80	80
85	85
90	90
95	95
100	100

Figure 23 shows the hypothetical change in class rank of the average student in the control group (50th percentile) based on the overall corrected effect size (d = .23). If that student were the only student to receive instruction using these strategies, his or her class rank would be expected to increase from 50th to 41st.

Summary

The following questions were considered through a random-effects meta-analysis of the independent action research studies:

1. What effect does the utilization of instructional strategies have on students' achievement regarding the subject matter content taught by their teachers?
2. Does the effect of instructional strategies differ between academic content areas?
3. Does the effect of instructional strategies differ from strategy to strategy?

The weighted average effect size estimate (Cohen's *d*) was not statistically significant ($d = .19$, $N = 20$, $p > .05$). When corrected for attenuation, the percentile gain associated with the use of the selected instructional strategies was 9 ($d = .23$, $N = 20$, $p > .05$). When applying Cohen's (1977, 1988) widely used convention for appraising effect sizes, the weighted average effect size estimate might be considered a small effect in the social sciences. The reported average effect sizes were positive for language arts ($d = .20$, $N = 6$, $p > .05$), mathematics ($d = .79$, $N = 2$, $p < .05$), and special education ($d = .86$, $N = 3$, $p < .01$). A comparison of the independent studies revealed a mixture of positive and negative effects for science and social studies. Finally, the reported average effect sizes were positive for graphic organizers ($d = .13$, $N = 13$, $p > .05$), interactive games ($d = .30$, $N = 7$, $p > .05$), partial vocabulary ($d = .19$, $N = 20$, $p > .05$), and voting technology ($d = .06$, $N = 3$, $p > .05$).

Technical Notes

Technical Note 1: Within the context of meta-analysis, independent studies can be analyzed using a fixed-effect or random-effects model. Fixed-effect models are based on an assumption of one true treatment effect common to every study. In other words, fixed-effect models assume that exactly the same effect size underlies all studies in the meta-analysis. Additionally, the fixed-effect model assumes that differences in observed effects are due to sampling error alone. Random-effects models do not assume the existence of a common treatment effect. In contrast, random-effects models allow for the possibility that the effect size varies from study to study and attempt to estimate that variance. Stated differently, random-effects models make an assumption that the true treatment effect sizes in the individual studies may differ from one another. Random-effects models are often used to estimate this variance. (For a more thorough discussion regarding models used in meta-analysis, see Lipsey & Wilson, 2001; Hunter & Schmidt, 2004; Cooper, 2009; Borenstein, Hedges, Higgins, & Rothstein, 2009.)

Technical Note 2: Conceptually, analysis of covariance (ANCOVA) can be thought of as using a covariate (i.e., pretest score) to predict students' performance on the posttest and then using the residual scores for each student as the dependent measure in the analysis. In other words, students' posttest scores are predicted using the scores received on the pretest. The difference between the predicted posttest scores and the observed posttest scores are then computed for each student that took both pretest and posttest. This difference is referred to as the residual score for each student. It represents the part of each student's posttest score that cannot be predicted from the pretest score for that student. Theoretically, use of residual scores based on pretest predictions is an attempt to equate all students on the dependent measure prior to execution of the intervention—in this case the use of the target instructional strategy (e.g., vocabulary).

Technical Note 3: The meta-analytic findings in this report are typically reported in two ways—observed and corrected. The corrected findings have been corrected for attenuation due to a lack of reliability in the dependent measure (i.e., teacher designed assessments of student academic achievement). Hunter and Schmidt detail the rationale and importance of correcting for 11 attenuation artifacts—one of which is random error associated with measurement of the dependent variable (2004, pp. 301-313). They explain:

> . . . error of measurement in the dependent variable reduces the effect size estimate. If the reliability of measurement is low, the reduction can be quite sizable. Failure to correct for the attenuation due to error of measurement yields an erroneous effect size estimate. Furthermore, because the error is systematic, a bare-bones meta-analysis on uncorrected effect sizes will produce an incorrect estimate of the true effect size. The extent of the reduction in the mean

> effect size is determined by the mean level of reliability across the studies. Variation in reliability across studies causes variation in the observed effect size above and beyond that produced by sampling error. . . . A bare-bones meta-analysis will not correct for either the systematic reduction in the mean effect size or the systematic increase in the variance of effect sizes. Thus, even meta-analysis will produce correct values for the distribution of effect sizes only if there is a correction for the attenuation due to error of measurement. (p. 302)

For ease of discussion we consider correcting for attenuation due to unreliability in the dependent measure using the population correlation instead of the population standardized mean difference effect size. The reader should note that the example provided regarding correcting correlations is analogous to correcting a standardized mean difference. To illustrate correcting for attenuation due to unreliability in the dependent measure, assume that the population correlation between the target instructional strategy (e.g., nonlinguistic representations) and student academic achievement is .50. A given study attempts to estimate that correlation but employs a measure of the dependent variable (i.e., a teacher designed assessment of student academic achievement) that has a reliability of .81—considered a typical reliability for a test of general cognitive ability. According to attenuation theory, the correlation would be reduced by the square root of the reliability (i.e., the attenuation factor). In other words, the population correlation is multiplied by the attenuation factor ($\sqrt{.81}$ = .90), thus reducing the correlation by 10 percent. Therefore, the observed correlation will be .45 (.50 x .90) even if there is no attenuation due to the other ten artifacts listed by Hunter and Schmidt (2004, p. 35). When the measure of the dependent variable has a lower reliability, .36 for example, the correlation is reduced by 40 percent ($\sqrt{.36}$ = .60) to .30 (.50 x .60). In order to make a correction for attenuation, the correlation is divided by the attenuation factor (i.e., the square root of the reliability).

For the purposes of this report, an estimate of reliability was used. Osborne (2003) found that the average reliability reported in psychology journals is .83. Lou and colleagues (1996) report a typical reliability of .85 for standardized achievement tests and a reliability of .75 for unstandardized achievement tests. Because the dependent measure in the independent action research studies involved teacher-designed assessments of student academic achievement, .75 was used as the reliability to correct for attenuation using the following formula:

$$d_c = \frac{d_o}{a}$$

In the formula, d_c is the corrected effect size, d_o is the observed effect size, and a is the attenuation factor. Using this formula, each effect size reported in Figure 1 was corrected for attenuation to produce the corrected meta-analytic findings considered in this report.

Technical Note 4: Independent variables can be analyzed as fixed effects or as random effects. In the context of ANOVA/ANCOVA, fixed effects are factors with levels that are deliberately arranged by the researcher. In the case of the original analysis of the 20 independent studies, the experimental condition (i.e., use of selected instructional strategies) was analyzed as a fixed effect. In contrast, random effects are factors with levels that are not deliberately arranged. Instead, random effects are factors which are randomly sampled from a population of possible samples. Generally speaking, when independent variables are analyzed as random effects, the intent is to generalize results beyond the boundaries of the independent variables employed in the study. For example, if a researcher were interested in the effect that the quality of school leadership has on academic proficiency, the researcher could select a random sample of schools in order to estimate the amount of variance in student academic achievement attributable to differences between types of school leaders. Thus, using the sample, the researcher can make generalizations regarding the influence of school leadership on academic achievement as a whole. Additional research could attempt to replicate the findings by selecting a different random sample of schools for comparison. When fixed effects are employed one typically does not generalize beyond the boundaries of the independent variables in the study. However, additional research could still attempt to generalize the findings by replicating every facet of the study. For example, multiple independent studies could be used to determine the influence of a specific instructional technique in the classroom.

Technical Note 5: In Figure 1, the column labeled "ES" contains the computed effect size for each study calculated as Cohen's *δ* using the following formula:

$$d = \frac{r}{\sqrt{(1 - r^2)(p(1 - p))}}$$

where *p* is the proportion of the total population in one of the two groups (i.e., the experimental group). Partial eta squared (η_p^2) as calculated by PASW® Statistics (SPSS, 2009) was used to determine partial eta (η_p) as an estimate for *r* (the effect size correlation) by taking its square root. Partial eta squared (η_p^2) describes the proportion of total variation attributable to the factor being considered (i.e., the target instructional strategy). This formula is used to compute the standardized mean difference effect size from an effect size correlation (e.g., the point-biserial correlation coefficient) when the experimental and control group populations are not equal (see Lipsey & Wilson, 2001, pp. 62-63). Again, partial eta (η_p) was used as an estimate for the point-biserial correlation coefficient in the formula.

The generic term *effect size* applies to a variety of indices (e.g., *d, r, R, PV*) that can be used to demonstrate the effect of an independent variable (e.g., use of the selected instructional strategy) on a dependent variable (e.g., student academic achievement). As used in this report, effect size refers to the standardized mean difference effect size. This index, first popularized by Glass (1976) and Cohen (1977), is the difference between experimental and control group means divided by an estimate of the population standard deviation.

$$\textit{standardized mean difference effect size} = \frac{\textit{mean of experimental group} - \textit{mean of control group}}{\textit{estimate of population standard deviation}}$$

Consider the following illustration of the use of effect size. Assume that the achievement mean of a group of students in a class that used a target instructional strategy (e.g., graphic organizers) is 90 on a standardized test and the mean of a group of students in a class that did not use the instructional strategy is 80. Assuming the population standard deviation is 10, the effect size would be as follows:

$$ES = \frac{90 - 80}{10} = 1.0$$

This effect size leads to the following interpretation: The mean of the experimental group is one standard deviation larger than the mean of the control group. One could infer from this that the use of graphic organizers raises achievement test scores by one standard deviation. Therefore, the effect size expresses the differences between means in standardized or "Z score" form, which gives rise to another index frequently used in research regarding education—percentile gain.

Percentile gain is the expected gain (or loss) associated with the effect size expressed in percentile points of the average student in the experimental group compared to the average student in the control group. By way of illustration, consider the same example. An effect size of 1.0 can be interpreted as the average score in the experimental group being about 34 percentile points greater than the average score in the control group. Again, the effect size translates the difference between group means into Z score form. Distribution theory dictates that a Z score of 1.0 is at the 84.13 percentile point of the standard normal distribution. To determine the percentile gain, the effect size is transformed into percentile points above or below the 50th percentile point on the unit normal distribution (e.g., 84% - 50% = 34%).

Appendix A: Instructions for Action Research

Thank you for considering engaging in an action research study regarding the effectiveness and utility of specific instructional strategies in your classroom. To be involved in a study you must be willing to do a few things. First you should select a specific instructional strategy and use this strategy in a unit that you teach. For example, you might decide to use nonlinguistic representations during the unit, or you might decide to have students engage in comparison or classification tasks as forms of identifying similarities and differences. Next, you must administer and score a pre-test and post-test for the unit. For example, if you teach mathematics, you might select a four week unit on linear equations to employ nonlinguistic representations. At the beginning of the unit, you would administer a pre-test on linear equations. At the end of the unit you would administer a post-test which might be identical to the pre-test, or it might be different. The important point is that you have a pre-test and a post-test score for each student on the topic of linear equations. Ideally the pre-test and post-test are comprehensive in nature. Additionally, you must deliver the same unit to another group of students. This, of course, means that you are teaching the same unit to two different groups of students. You would administer the same pre-test and post-test to this other group of students; however, you would not use the selected instructional strategy. In this case, you would not use nonlinguistic representations with this second class. Finally, you would score the pre-test and post-test for both groups of students and record their scores on the attached forms. You don't have to identify students by name (in fact, it is preferable that you don't). The unit can be as short or long as you wish, but it must be completed and the data submitted by ___________________.

When you have completed the study please fill out the forms below and submit them to the action research team leader for your school. That individual is ________________________.

Note that the first form asks you to provide general information about your school, the instructional strategy you used and so on. It also asks you to provide a personal ID number as opposed to your name. This is because the results of the action research projects will be reported in an anonymous fashion. Only you will know which results apply to your students.

Thank you for considering involvement in an action research project.

School ___________________________

Personal ID number _____________________________

Subject and grade level taught _________________________

Topic addressed during the unit ________________________________

Instructional strategy you used _______________________

General comments about the project:

References

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2.2.046) [computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, England: John Wiley & Sons.

Cohen, J. (1977). *Statistical power for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. (2009). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Glass, G. V. (1976). Primary, secondary, and meta-analyses of research. *Educational Researcher, 5,* 3-8.

Harlow, L. L., Muliak, A. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Haystead, M. W., & Marzano, R. J. (2009). *Meta-analytic synthesis of studies conducted at Marzano Research Laboratory on instructional strategies*. Englewood, CO: Marzano Research Laboratory.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press, Inc.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: SAGE Publications.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66* (4), 423-458.

Marzano, R. J. (2004). *Building background knowledge for academic achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2006). *Classroom assessment and grading that work.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Erlbaum.

Osborne, J. W. (2003). Effect sizes and disattenuation of correlation and regression coefficients: Lessons from educational psychology. *Practical Assessment, Research and Evaluation, 8* (11). Retrieved March 20, 2009, from http://PAREonline.net/getvn.asp?v=8&n=11.

SPSS. (2009). *PASW® Statistics* (Version 17.0.2) [computer software]. Chicago, IL: SPSS.