CHAPTER 4

The Results

This chapter presents the results of the meta-analysis of leadership variables empirically related to student outcomes. As stated in Chapter 1, the meta-analysis seeks to answer the following four questions:

1. What empirical evidence is there to confirm or support the notion that there exists a relationship between leadership and school effectiveness?
2. As identified in the quantitative analysis of available research, what is the strength of the relationship between leadership and school effectiveness?
3. To what extent is there a relationship between selected dimensions of leadership and school effectiveness?
4. How do these dimensions of leadership compare in their relationship to school effectiveness?

In presenting the results, the table of categorical model characteristics is presented and discussed. Also, the composite effect size for the studies in the meta-analysis, the confidence intervals, the deviations, and the expected reduction in heterogeneity if the study is excluded are presented in Appendix B.

A major interest of this study is to determine the aspects or dimensions of leadership that facilitate school effectiveness. However, because observed ns by dimensions often differed from the overall study N, it became necessary to create sub-samples of effect sizes by dimension. This approach results in a truer measure of the magnitude of the influence of the principal's leadership. Therefore, in presenting the leadership behaviors by dimension, the effect sizes are

based on the ns for the sub samples. In addition to this, effect sizes classified by construct are presented in table format followed by brief interpretive comments.

Before considering the findings reported in the research on the effects of leadership on student outcomes, it is necessary to provide the meta-analytic context for the studies. Therefore, in addition to the model characteristics of the studies, a summary table is presented.

Characteristics of Studies

Table 3 gives a summary of many of the encoded study characteristics. As indicated by the reported frequencies of the characteristics, most of the studies located an interest in effective leadership at the level of the elementary school, and did not categorize the schools according to their locations. Also, most of the studies either used mixed groups or did not indicate the socio-economic background of the students. However, while one study, Heck (1992), did compare the leadership behaviors of the principals in effective low socio-economic schools with those of non-effective low socioeconomic schools, for most of the studies, the SES was used as a covariate to moderate the effects on student outcomes. As indicated in Chapter 2, for most studies, effectiveness was determined by the mean school performance on an achievement test.

Twenty-three studies examined the effects of leadership based on whether the approach adopted was instructional or transformational. Ten addressed leadership styles as measured by instruments such as the LBDQ developed by the researchers at Ohio State; LEAD developed by Hersey and Blanchard (1969), and the LBA 11 by Blanchard, Hambleton, Zigarmi, and Forsyth (1982) and five looked at the overall impact of the principal's leadership. The meta-analysis spanned the years 1985-2000, the median of publication date for journal and peer-reviewed paper presentations was 1992, and for unpublished dissertations, 1990. Approximately 53% of the

studies were from published articles, and nearly 100% of the unpublished dissertations used established instruments as a measure of school leadership and school effectiveness (Appendix A). As expected, the majority of the studies were conducted in the United States.

Table 3: Summary of study characteristics

Variable and Class	Frequency	Variable and Class	Frequency
Median publication date	1992	Mean observations:	9
Publication Form:		School type:	
Journal article	20	Elementary	22
Dissertation	14	Middle	2
Eric document	4	Secondary	8
		Mixed	8
Independent Variable:			
Leadership approaches	23	Location:	
Leadership styles	10	Urban	4
Leadership impact	5	Suburban	2
		Rural	
Outcome Measure:		Mixed	32
Academic achievement	22		
Affect	9	Analysis Used	
Overall effectiveness	8	Means & SD	8
Country of study:		Inferential statistics	4
U.S.A.	25	Correlations	24
Canada	6	p-value	2
Australia	3	Type of measure used:	
Other	4	Author developed	15
		Published	23

This meta-analysis did not look at the moderating effects of gender or race. This is principally because (1), the main focus of this study is in approaches, styles and specific leadership behaviors; and (2), many of the studies did not address these issues. Of the few that did look at the moderating effect of gender and /or race, only one reported statistically significant differences in leadership influence with regard to gender, and none found differences in leader effectiveness based on race. Moreover, none of the published articles addressed gender and race

as moderating the effectiveness of the principal. The lack of interest by the journals could be interpreted as a reflection of the knowledge in the field that these two variables do not significantly moderate leader effectiveness. Also, it would be difficult to rationalize leadership as a construct if its effectiveness was dependent on the gender or race of the leader.

With regard to the samples used in the studies, they all comprised principals and/or teachers, and a measure of student performance as the determinant of effectiveness, however as can be seen in Appendix A, most of the studies aggregated the scores to the school level, making the school the unit of analysis.

Distribution of Effect Sizes

From the 38 studies, 339 effect sizes or gs were calculated. These effect sizes, "the difference between the means in standard deviation units" (Hattie, 1998) were converted into *d*s correcting for bias, and comprise the data-points used in conducting the meta-analysis. The 339 data-points give an average of nine, and a median of five data-points per study.

```
Frequency          Stem &  Leaf

     2.00          -0 .  7
     4.00          -0 .  4&
     8.00          -0 .  223
    13.00          -0 .  000001
    48.00           0 .  000000000000111111111111
    64.00           0 .  2222222222222333333333333333333
    72.00           0 .  444444444444444444555555555555555555
    36.00           0 .  666666666677777777
    28.00           0 .  8888999999999
    31.00           1 .  000000000111111
    10.00           1 .  22233
     8.00           1 .  4455
    11.00           1 .  66667              mean      = 0.5691
     1.00           1 .  &                  95% CI    = 0.5141 / 0.6241
     3.00 Extremes      (>=2.1)             median    = 0.4851
```

Figure 7: Stem-and-leaf diagram of the effect sizes calculated from the 38 studies.
Stem width: 1.000000. Each leaf: 2 case(s) & denotes fractional leaves.

The frequency of these data-points is shown in the stem-and-leaf diagram of the effect sizes presented in Figure 7. As can be seen, the distribution is positively skewed (Appendix B). There are 20 effects greater than 1.3, and 14 above 1.5 standard deviation units. An inspection of the studies that produced these effects did not reveal any discernible pattern. The 14 effects greater than 1.5 standard deviation units came from 11 studies, publication range 1988 – 1999; median publication date 1994.

Among the studies producing these outlier effects were Bamburg and Andrews (1991); Dzyacky (1988); Heck, (1992); Heck, Larsen, and Marcoulides (1990); and Silins (1994a). The Silin's study is a peer-reviewed paper of research conducted in Australia, and the effects were calculated from the correlation matrix. The Dzyacky study is a dissertation and the analysis was also based on the correlation between leadership behaviors and school effects. Heck, Heck et al, and Bamburg and Andrews may be considered the most similar; they are journal articles and compared the leadership behaviors of the principals of high achieving schools to those of low achieving schools. Despite this fact, with 53% of the studies being journal publications, and 37% of the researches utilizing means and SD, or the *t*- test or *F*- test in the analytic procedures, no significance could be attributed to the occurrence beyond that of chance.

The influence of the extreme values was evident, in that the means across all effect sizes within the 11 studies was substantially higher than the overall mean of the distribution. The overall mean was 0.5691 and the mean for the 11 studies was 0.8531.

Analysis of the Effect Size

In a meta-analysis, each study contributes an effect size to the analysis. From these, following the procedure established by Hedges and Olkin (1985), a mean effect size was computed with each effect size weighted by the reciprocal of its variance. Table 4 shows *d*, the

corrected effect size for each study, the 95% confidence interval, the deviation of d_i from the composite effect size d_+, excluding d_i, and the amount the homogeneity statistic Q would be reduced by should d_i be excluded. For this sample of studies, the mean weighted effect size; d_+ is 0.6646, indicating a positive relationship between leadership and student outcomes.

Table 4: Composite effect size of studies in the meta-analysis

Study	*d*	95% CI	Dev	Homo.
1-Chapman	+1.0188	+0.18 / +1.86	+0.356	-0.684
2-Blank	+0.2590	-0.23 / +0.75	-0.412	-2.653
3-Shoch	+0.3513	-0.12 / +0.83	-0.319	-1.710
4-Cheng1991	+1.2475	+0.51 / +1.98	+0.587	-2.435
5-Ericksen	+0.3279	-0.06 / +0.72	-0.346	-2.928
6-Dzyacky	+1.2487	+0.58 / +1.92	+0.589	-2.953
7-Hajnal, Walker, & Sackney	+1.3969	+0.94 / +1.85	+0.747	-10.219
8-Garner	+1.3085	+0.91 / +1.71	+0.660	-10.241
9-Bista, & Glasman	+0.1245	-0.16 / +0.41	-0.567	-14.376
10-Bulach, & Lunnenburg	+1.5849	+0.50 / +2.67	+0.923	-2.749
11-Hoy, Hannum, & Tschannen-Moran	+1.0895	+0.64 / +1.54	+0.433	-3.446
12-Rock	+0.5621	+0.13 / +1.00	-0.105	-0.219
13-Krug	+0.3781	-0.06 / +0.82	-0.292	-1.667
14-Montgomery	+0.6831	+0.34 / +1.02	+0.019	-0.012
15-Standley	+1.6323	+1.12 / +2.14	+0.983	-14.088

Study	*d*	95% CI	Dev	Homo.
16-Freeman	+0.5299	-0.39 / +1.44	-0.135	-0.084
17-Leithwood & Jantzi 1998	+0.6528	+0.27 / +1.04	-0.012	-0.004
18-Snyder & Ebmeier	+0.4251	-0.30 / +1.15	-0.241	-0.424
19-Hoy, Tarter & Bliss	+0.2842	-0.23 / +0.80	-0.386	-2.108
20-Bamburg, & Andrews	+1.0162	+0.08 / +1.95	+0.353	-0.550
21-Cheng1994	+0.3085	+0.00 / +0.62	-0.372	-5.360
22-van de Grift	+0.0375	-0.20 / +0.27	-0.675	-29.078
23-Silin, Mulford & Zarins	+0.7380	+0.32 / +1.15	+0.075	-0.124
24-Jackson	-0.2373	-0.65 / +0.18	-0.923	-18.810
25-Durr	+0.5124	-0.28 / +1.31	-0.153	-0.141
26-Silins1994 A	+1.3522	+1.10 / +1.61	+0.732	-29.802
27-Heck	+0.7801	+0.14 / +1.42	+0.117	-0.125
28-Leithwood, & Jantzi 1999	+0.4492	+0.04 / +0.86	-0.221	-1.089
29-Floyd	+0.3421	-0.13 / +0.82	-0.328	-1.799
30-Philbin	+0.3930	+0.12 / +0.66	-0.287	-4.172

Study	*d*	95% CI	Dev	Homo.
31-Buzzi	+1.6629	+1.44 / +1.88	+1.086	-85.161
32-Silins 1992	+0.3512	+0.10 / +0.60	-0.335	-6.616
33-Finklea	+0.3388	-0.11 / +0.78	-0.332	-2.108
34-Silins1994 B	+0.4933	+0.24 / +0.74	-0.183	-1.946
35-Fishman	+0.4285	-0.18 / +1.04	-0.239	-0.578
36-Heck, Larsen, & Marcoulides	+0.6721	+0.39 / +0.96	+0.008	-0.003
37-Williams	+1.0567	+0.72 / +1.39	+0.406	-5.430
38-Brice	+0.9011	+0.01 / +1.79	+0.238	-0.272
Overall:	+0.6646	+0.60 / +0.73	+0.399	-7.004

Note: $Q(37) = 252.035$; $\underline{p} = 0.000$

The 95% confidence interval for the mean was CI_{95} = +0.60 / +0.73, implying that leadership had a significant positive impact on student outcomes at the school level. Calculation of the homogeneity statistic *Q*, resulted in the rejection of the hypothesis of homogeneity, $Q(37) = 252.035$; $\underline{p} < 0.0001$, thus indicating that the effect sizes are inconsistent and the overall mean may not be the most reliable value for the effect sizes in this sample.

The analysis of the effect sizes revealed that study # 31 Buzzi (1990) had the largest mean absolute deviation from *d*, dev. = 1.086 and its exclusion would have resulted in *Q* being

reduced by –85.161. The study was excluded from further analysis and the variance was recomputed. Study # 32, Silins (1994a) was now flagged as the next largest outlier, *Q* remaining significant at p >.001.

As evidenced in studies by Falchikov and Goldfinch (2000); Hattie, Biggs and Purdie (1996); Johnson and Eagly (1989), a significant *Q* value is not uncommon in the social sciences where constructs often defy precise definitions. As a matter of fact, Hedges, as cited in Johnson and Eagly (1989), found in conducting meta-analysis on psychological topics, "the removal of up to 20% of the outliers in a group of heterogeneous effect sizes usually resulted in a high degree of homogeneity." (p. 30). Eagly, Karau and Makhijani (1995) in their meta-analysis of gender and the effectiveness of leaders removed up to12 outlier effect sizes before the variance was reduced sufficiently to produce a homogeneous set of effect sizes, and therefore, a non-significant *Q* statistic.

An alternative to identifying and removing outlier effect sizes is the use of categorical models which divide the studies into classes based on study characteristics. Therefore as suggested by Hedges and Becker (1986) in a procedure analogous to the ANOVA, the study attributes were used to establish sub-groups to determine whether the establishment of these categorical classes would account for the variance in the effect sizes.

The Categorical Models

Table 5 reports the tests of the categorical models. As explained in Chapter 3, following testing the homogeneity of the effect sizes, especially when *Q* is significant, categorical models are fitted to effect sizes. In presenting the relations between the various study characteristics and the categorical models, the table is divided into sub-categories and reports the between classes effects, Q_B, for each categorical model, the homogeneity of the within classes effects, $Q_{w,}$ the

mean weighted effect size for each class, calculated by weighting each effect size by the reciprocal of its variance, $d_{+,}$ and the CI for each mean. An observation of Table 5 shows clearly that almost all the Q_w estimates are significant thus indicating wide variation of effect sizes within categories. Despite this fact, it is necessary to look at the between classes effects to determine whether classes differed significantly from each other.

It is generally accepted that journals report marginally more effective studies than monographs, while dissertations tend to report interventions that are not effective (Hattie, Biggs, & Purdie. 1996). This finding is not replicated in this analysis. The Q_B statistic is significant, and analogous to the ANOVA, indicates that at least one of the weighted effect sizes differ significantly from the others. The mean for dissertation, $d_+ = 0.82$ is almost twice as large as the mean for the journal articles.

Following the established procedures (Hedges & Becker, 1986; Hedges & Olkin, 1985), post-hoc comparisons among the mean weighted effect sizes for the publication category showed that the means for the journal studies differed significantly from both the dissertation and peer reviewed papers, $c^2(1) = 32.11$, and $c^2(1) = 28.56$ respectively, $\underline{p} > .0001$. It must be noted however, that the Q_w is significant for the class, and even more so within the dissertation sub-group. Therefore a great deal of caution must be applied in the interpretation of these results.

The results for the category independent variables also indicated that there were significant differences among the weighted means of the variables that comprise the category, $Q_B(2) = 24.09$; $\underline{p} < .0001$. The post-hoc contrasts indicated that the weighted mean of studies that focused on the overall leadership effectiveness differed from the other two classes. As is evidenced by the higher value for the mean weighted effect size, $d_+ = 1.14$, and the non-significant variance statistic, $Q_w = 1.66$, the studies that examined the relationship of the overall

leadership, or leadership impact on the student outcomes consistently reported higher means than those studies that located an interest in the relationship of approaches and leadership styles to school effectiveness.

Table 5: Categorical models of study characteristic

Variable and class	Q_B	d_+	95% CI		Q_w
			Lower	Upper	
Publication type	45.76**				
Journal		0.43	0.034	0.53	52.76**
Dissertation		0.82	0.73	0.92	132.84**
ERIC Document		0.97	0.79	1.14	20.68**
Ind. Variable	24.09**				
Leadership approaches		0.66	0.59	0.74	161.01**
Leadership style		0.51	0.38	0.63	62.26 **
Leader effectiveness		1.14	0.92	1.36	1.66
Outcome Variable	78.92**				
Academics		0.52	0.43	0.60	115.27**
Affect		0.54	0.42	0.66	26.73 **
School effects		1.21	1.07	1.34	48.58**
School type	17.05*				
Elementary		0.75	0.67	0.83	184.25**
Middle		0.36	0.06	0.67	18.02*
Secondary		0.44	0.28	0.60	8.55
Mixed		0.59	0.44	0.74	28.04**
Location	8.42				
Urban		0.50	0.31	0.69	19.28*
Suburban		0.35	0.06	0.64	0.03
Mixed		0.70	0.63	0.77	224.31**
Analysis	2.82				
Means & SD		0.68	0.51	0.86	50.2**
T & F-values		0.64	0.61	1.11	8.13

Correlations		0.64	0.57	0.72	231.35**
P-values		0.68	0.45	0.90	13.92*
Variable and class	Q_B	d_+	95% CI		Q_w
			Lower - Upper		
Measure used	20.96**				
Author developed		0.53	0.45	0.62	86.86**
Published		0.82	0.73	0.92	144.21**
Approaches	23.07**				
Instructional		0.74	0.64	0.84	121.03**
Transformational		0.62	0.51	0.72	42.75**
Inclusive		1.21	0.92	1.51	1.12
Leadership style		0.48	0.34	0.62	60.06**
Dimensions	22.95**				
Instruct. Organization		0.33	0.24	0.51	14.81
Climate		0.67	0.54	0.80	42.20**
Defining mission/goals		0.75	0.60	0.90	34.75**
Consideration		0.32	0.13	0.50	14.82
Inspiration		0.59	0.42	0.76	11.99

* p <.01, ** p < .001

All statistics for the dependent variable category were significant. As discussed previously, the emphasis is on academic achievement as a measure of school effectiveness, however, it is necessary to recognize that there are other goals of schooling which are also very important (Leithwood & Jantzi, 1999). It is possible that these goals, although related to student achievement (Leithwood & Jantzi; Cheng, 1991) may be distinct constructs. Therefore, despite the significant within class variability which could mask the between classes difference, the very significant $Q_B = 78.92$, the highest among all the categorical model tests, may be a confirmation that the various outcome measures are distinct, and representative of different constructs.

According to Leithwood (1993), and Silins (1992) in their criticisms of the early conceptualization of instructional leadership cited the present reform oriented role of the principal, and the marked difference in context between the elementary and secondary school. As is observed in Table 5, there is a significant difference between the weighted means of the elementary and secondary classes; $d_+ = 0.75$ and 0.44 respectively. Also, whereas the effect sizes for the elementary class varied widely, the secondary class was homogenous as a group, suggesting that the studies that examined leader effectiveness at that level may not have been influenced by moderating variables. The post-hoc contrasts simply confirmed the significant difference between the elementary and secondary classes within this category, $c^2(1) = 9.42$; $\underline{p} = 0.024$. Despite this difference which may suggest that leadership behaviors differ depending on school level, $d_+ = 0.44$, $CI_{95} = 0.28 / 0.60$ represents a significant relationship between the effectiveness of the leadership and the effectiveness of the school.

The category location had only two studies that looked at the suburban areas and four with an interest in leadership in the urban school. Consequently, with 84% of the studies using a mixed population in their sample, it is not surprising that the post-hoc contrasts revealed no statistically significant differences among the three classes.

The model for the type of analytic procedure was not significant, $Q_B(3) = 2.82$, indicating that there were no significant mean differences among the classes. Consequently, there was no need to conduct any contrasts. Also, it is important to note two of the classes met the assumption of homogeneity indicating the findings across the various studies were reasonably consistent.

The category, type of measure used comprised two classes. Recall that almost all the dissertations used a published measurement instrument. As such, it is not surprising that the

weighted mean for published measures is higher than the measures developed for use in the specific study.

The class leadership approach was partitioned into instructional, transformational, and inclusive, a catchall word for research that combined leadership models, or presented indicators that could not be fitted into a specific model. Similar to the category independent variable, overall leadership recorded positive effect sizes in above one standard deviation (d_+ = 1.14 and 1.21 respectively). Also, as with leader effectiveness, the studies that examined the effects of leader influence are homogenous, Q_w = 1.12, thus it can be said that overall leadership has a positive effect on school effectiveness as measured by the various student outcomes.

Of specific interest is the relationship between the dimensions or aspects of leadership and school effectiveness. The table reports significant between class variance (Q_w = 22.95; $\underline{p} < .001$). This indicates that at least one weighted mean is significantly different from the other. The post-hoc comparisons identified defining the school mission as being significantly different from the others. The difference between instructional organization and defining the mission was: $c^2(1) = 13.57$, $\underline{p} < .001$, and defining the mission and consideration, $c^2(1) = 12.72$, $\underline{p} = 0.013$. Instructional organization and climate were barely significant at the .05 level, $c^2(1) = 9.51$.

It is important to note that the confidence interval around the mean of each class, or behavior cluster, does not contain zero, and all are positive in direction. Therefore, despite the need for caution due to the significant within variance estimates, it can be said that the various categories all report a significant positive effect on student outcomes.

Effects by Dimensions of Leadership

Whereas some of the studies addressed school leadership with regard to one criterion and consequently contributed one overall effect to the analysis, others did not. For example, Heck (1992) looked at elementary schools, and secondary schools. Similarly, the *d* values for some of the classes within categories are not always based on the same number of observations, and as a result, do not reflect equivalent numbers of data points. An example of this is Blank (1987). Of the 15 effect sizes calculated, 10 estimated the effects on student achievement in reading and math, and 5 looked at student attendance. In reality therefore, the Heck study produced two effect sizes and so, in conducting the analysis on type of school, separate *d* values were used for elementary and secondary schools. Similarly, in examining the five leadership approaches to determine the degree to which each contributes to the effectiveness of the school, composite mean weighted effect sizes were computed using mean gs based on the number of effect sizes by class.

The following tables show the mean weighted effect size, the confidence interval of the mean and the homogeneity statistic for the studies that produced the effect sizes for each leadership dimension. Studies that examined the effects of overall leadership are not included in this analysis.

As can be seen in Table 6, 12 studies with a focus on the instructional leadership of the principal, measured leader effectiveness on the dimension instructional organization. The studies were homogenous as indicated by a non-significant Q value, and overall produced a mean weighted effect size significantly different from zero; $d = 0.66$, $CI_{95} = 0.34 / 0.62$. As is evidenced in the table, even where the degree of the relationship was not different from what could be found by chance, all the effect sizes were positive, indicating a positive relationship between effective instructional organization and school outcomes.

Table 6: Effect size of instructional organization

Study	N	*d*	95% Lower	95% Upper
Blank 1987	32	0.28	-0.41	+0.98
Schoch	70	0.45	-0.03	+0.92
Dzyacky	30	0.93	+0.29	+1.58
Krug	81	0.35	-0.09	+0.79
Snyder & Ebmeier	30	0.50	-0.23	+1.23
Hoy, Tarter, & Bliss	58	0.44	-0.08	+0.97
Cheng 1994	164*	0.28	-0.02	0.59
Durr	25	0.45	-0.35	+1.24
Heck: elementary	23	1.10	+0.19	+2.02
Heck: secondary	17	0.43	-0.55	+1.40
Silins 1994 a	291*	1.25	+1.00	1.50
Fishman	42	0.38	-0.23	+0.99
Heck, Larsen & Marcoulides	198*	0.75	+0.47	1.04

Composite *d*: = 0.66	SD: =0.18	CI_{95} = 0.34 / 0.62
Q(11) = 7.12	p = 0.79	N = 781

* Individual responses

Table 7 reports the results of the analysis on the climate dimension of leadership behaviors. The first computation of the composite effect size identified Silins 1994a as an outlier.

Following the procedure suggested by Hedges (1987) in Johnson and Eagly (1989), these effect sizes were excluded from the analysis and the d recomputed. In an effort to achieve homogeneity, two other studies, Dzyacky (1988) and Heck, Larsen and Marcoulides (1990) were removed in order of decreasing reduction of homogeneity. Reanalysis of the effect sizes produced the overall effect size; $d = 0.29$; $Q = 8.9$; $\underline{p} = 0.18$. This effect size although significant, $CI_{95} = +0.13 / 0.45$; does not reflect as strong a relationship with school effectiveness as the dimension instructional organization.

Table 7: Effect size of the climate dimension of leadership behaviors

Study	N	*d*	95%	
			Lower	Upper
Blank 1987	32	0.13	-0.56	+0.83
Schoch	70	0.18	-0.29	+0.65
Krug	81	0.29	-0.15	+0.72
Snyder & Ebmeier	30	0.53	-0.20	+1.26
Hoy, Tarter, & Bliss	58	1.01	+0.47	+1.56
Cheng 1994	164*	0.29	-0.02	+0.60
Bista & Glasman	188*	0.12	-0.17	+0.40

Composite *d*: = 0.29	SD: =0.23	$CI_{95} = 0.13 / 0.45$
$Q(6) = 8.92$	$\underline{p} = 0.18$	N = 623

* Individual responses

Originally nine studies examined the effects of the cluster of leadership behaviors related to defining the school's mission, on the effectiveness of the school. Again the study by Silins (1994a) registered significant deviation in variance and was removed from the analysis. The remaining eight studies were homogenous, $Q(7) = 6.87$; $\underline{p} = 0.44$, and produced a significant

composite effect size, $d = 0.22$; $CI_{95} = 0.07 / 0.37$. With the exception of the study by Bista and Glasman (1998), all the studies indicated a positive relationship between leadership and school effectiveness.

Table 8: Effect size of the defining the mission dimension of instructional leadership

Study	N	*d*	95%	
			Lower	Upper
Blank 1987	32	0.36	-0.34	+1.06
Schoch	70	0.43	-0.04	+0.90
Krug	81	0.50	+0.05	+0.94
Snyder & Ebmeier	30	0.25	-0.47	+0.97
Cheng 1994	164	0.35	+0.04	+0.66
Jackson	91*	0.12	-0.29	+0.53
Fishman	42	0.23	-0.37	+0.84
Bista & Glasman	188*	-0.06	-0.34	+0.23

Composite *d*: = 0.22 SD: =0.17 $CI_{95} = 0.07 / 0.37$

Q(7) = 6.87 p = 0.44 N = 698

* Individual responses

The dimension consideration was measured by the effects sizes from seven studies. Of the seven, the study by Dzyacky (1988) was identified as an outlier and to attain homogeneity, was omitted from further analysis. The six studies produced a mean weighted effect size of $d = 0.36$; $CI_{95} = 0.21 / 0.51$ (Table 13).

As can be seen in Table 10, three studies were retained for the computation of the composite effect size of the dimension inspiration. Excluded was the study by Bista and Glasman

(1998) which had the effect of reducing the variance by 13.11 rendering a non-significant *Q* statistic, *Q* (2) = 4.92; p = 0.09.

Table 9: Effect size of the consideration dimension of leadership behaviors

Study	N	*d*	95% Lower	95% Upper
Bista & Glasman	188*	0.91	-0.19	+0.38
Dzyacky	30	0.92	+0.28	+1.57
Durr	25	0.91	+0.08	+1.73
Silins 1992	256	0.43	+0.18	+0.68
Cheng 1994	164	0.29	-0.02	+0.60
Fishman	42	0.68	+0.06	+1.30

Composite *d*: = 0.36 SD: = 0.34 CI_{95} = 0.21 / 0.51

Q(5) = 9.56 p : = 0.09 N = 716

* Individual responses

Table 10: Effect size of the inspiration dimension of leadership behaviors

Study	N	*d*	95% Lower	95% Upper
Dzyacky	30	1.10	+0.45	+1.76
Cheng 1994	164	0.29	-0.01	+0.60
Silins 1992	256	0.36	+0.12	+0.61

Composite d: = 0.40 SD: =0.63 CI_{95} = 0.21 / 0.58

Q(2) = 4.92 p = 0.09 N = 461

It must be noted that the studies that were flagged outliers, were among the 11 studies identified through the analysis of the distribution of effect sizes as having effect size values in

the extreme range. However, comparative analysis of the per-dimension mean effect sizes suggests that the cluster of leadership behaviors with emphasis on instructional organization has the strongest effect on the school effectiveness. As stated previously, this statement has to be tempered by extreme caution, given that for some dimensions the effects sizes were derived from too few studies, and even then, only the studies that would have met the criteria for selection into the meta-analysis.

Observing the Trend

As stated in Chapter 3, classifying the studies according to Hallinger and Heck's (1990) models' approach facilitates in the observation of trends that may exist, and may be reflected in the data. Looking at Table 11, a pattern becomes discernible. Of the seven studies in Model A, five found non-significant or an inconclusive relationship between the leadership of the principal and the level of school effectiveness. The date range of the publications was 1986 – 1994, and the median publication date was 1990. As evidenced by the Hedge's corrected d, of the two studies that reported significant effect sizes, the study by Bulach, Lunenburg, and McCallon (1994) found the stronger relationship, $d = 1.58$

Model A1 is a little more sophisticated in design. Of the meta-analytic sample, 10 studies met the criteria for classification into the model. Seven of the studies were dissertations, the ratio of significant findings to non-significance was 1:1, and the median publication date and range were 1987, range 1985 – 1998.

As researchers adopted more sophisticated designs for their studies, they were more able to detect whether or not a relationship existed, and if so, the strength of the relationship. According to Hallinger and Heck (1990), there is a progression from inquiry focused on the extent to which leadership behaviors directly effect responses in student behavior, is in response

to the growing body of literature that suggests that the effects of leadership may not be as direct as first envisaged, but impact upon a number of variables that mediate the effects.

Table 11: Classification by model design

Author	Date	Sc.	N	Effects	g	d	95% C.I lower - upper	
Model A- Direct effects								
Bulach, Lunenburg & McCallon	1994	1	17	3	1.67	1.58	0.50	2.67
Durr	1986	1	25	14	0.53	0.51	-0.28	1.31
Erickson	1987	1	102*	8	0.33	0.33	-0.06	0.72
Krug	1992	4	81	60	0.38	0.37	-0.24	1.01
Montgomery	1987	1	15	13	0.69	0.68	0.34	1.02
Schoch	1992	1	70	10	0.36	0.35	-0.12	0.82
Van de Grift	1990	1	275*	8	0.04	0.04	-0.20	0.27
Model A1-Direct effects with antecedents								
Bista & Glasman	1998	4	188*	4	0.12	0.12	-0.16	0.41
Blank	1987	3	32	15	0.54	0.52	-0.34	1.37
Chapman	1998	1	46	4	1.05	1.02	0.18	1.86
Cheng	1994	1	164	40	0.31	0.31	0.00	0.62
Finklea	1997	3	79	4	0.34	0.34	-0.11	0.78
Fishman	1986	1	42	30	0.44	0.43	0.18	1.04
Freeman	1987	1	19	1	0.55	0.53	-0.39	4.45
Garner	1989	1	117*	1	1.32	1.31	0.91	1.71
Jackson	1986	2	91*	3	-0.24	-0.24	-0.65	0.18

Standley	1985	1	79*	2	1.63	1.62	1.11	2.12

Model B- Mediated effects

Author	Date	Sc.	N	Effects	g	d	95% C.I lower - upper	
Bamburg & Andrews	1991	1	20	19	0.66	0.65	0.27	1.04
Buzzi	1990	1	423*	1	1.67	1.66	1.44	1.88
Cheng	1991	3	64	1	1.28	1.25	0.51	1.98
Dzyacky	1988	4	41	7	1.27	1.25	0.58	1.92
Floyd	1999	1	69	2	0.35	0.34	-0.13	0.82
Hajnal, Walker & Sackney	1998	4	93	2	1.41	1.40	0.94	1.85
Hoy, Hannum, & Tschannen-Moran	1998	2	86	6	1.10	1.09	0.64	1.54
Hoy, Tarter & Bliss	1990	3	58	9	0.29	0.28	-0.23	0.80
Philbin	1997	3	218*	2	0.39	0.39	0.01	0.77
Rock	1988	1	85	3	0.57	0.56	0.13	1.00
Silins	1994a	1	291	2	1.36	1.35	1.10	1.61
Silins	1994b	1	256	2	0.49	0.49	0.24	0.74
Silins	1992	1	256	10	0.35	0.35	0.10	0.59
Williams	1995	1	202*	4	1.06	1.06	0.72	1.39
Model B1- Mediated effects with antecedents								
Brice	1992	4	91	1	0.94	0.90	0.01	1.79
Heck	1992	1, 3	40	16	0.80	0.78	0.13	1.43

Heck, Larsen & Marcoulides	1990	3	198	44	0.67	0.67	0.39	0.96

Author	Date	Sc.	N	Effects	g	d	95%CI	
							lower -	upper
Hoy, Hannum & Tschannen-Moran	1998	2	86s	6	1.10	1.09	0.77	1.41
Leithwood & Jantzi	1999	1	94s	2	0.45	0.45	0.04	0.86
Leithwood & Jantzi	1998	4	110s	6	0.66	0.65	0.27	1.04
Silins, Mulford & Zarins	1999	3	96	5	0.74	0.74	0.32	1.15
Snyder & Ebmeier	1992	4	30s	5	0.25	0.25	-0.47	0.97

Sc: school 1: elementary 2: middle 3: secondary/high 4: all levels

The Model B design incorporates the above position. As can be seen by the 95% confidence interval of the mean, of the 14 studies that comprise the category, two did not report significant findings. The trend continues with Model B1, by the direction of the *d*, and the non-inclusion of zero in the confidence interval, all the studies report a relationship between leadership and student outcomes that is significantly different from zero.

The reported estimated strengths of the relationships ranged from just under 0.5 standard deviation (Leithwood & Jantzi, 1999; d =0.45) to approaching one standard deviation (Brice, 1992, d = 0.90). It is important to note that with the exception of Dzyacky (1988) and Buzzi (1988), the studies that utilized the Model B designs were all post 1990.

Adding to the Analysis

Table 12 is a summary of studies that were not incorporated into the meta-analysis because of the absence of statistical information essential for conducting the meta-analysis.

These studies are all journal articles, and as can be seen, of the 13 studies presented, 11 were published post 1992, median publication date 1994.

These studies all employ sophisticated analytic procedures and all utilize the Model B or Model B1 designs discussed in Chapter 3 and referred to above. Through the use of a most unsophisticated procedure, a simple count of the findings, it is evident that the school leadership, embodied in the principal, does influence the level of school effectiveness: an effectiveness that is measured by the percentage of students meeting the criteria established by the school or state on a number of outcome variables.

Of the 13 studies presented, nine (69%) found that the principal as leader, contributed significantly to the level of school effectiveness. The two studies that reported mixed or weak findings were Goldring and Pasternack (1994) and van de Grift (1999). Goldring and Pasternack looked at principals' of elementary community schools in Israel use of strategies to coordinate school activities and the relationship to school effectiveness. Van de Grift (1999) whose study was conducted in the Netherlands found that the importance of the variables that impacted student performance differed from the findings in the USA. One study, Leitner (1994) found that the relationship between leadership and school effectiveness was not different from that being found by chance.

The Silins (1994 a) study was coded not applicable because it utilized the same data and employed almost the same methodology as Silins (1992) and Silins (1994 b) used in the meta-analysis.

Addressing the Research Questions

Research questions asked, what empirical evidence is there to confirm or support the notion that there exists a relationship between leadership and school effectiveness, and research

question 2 sought to determine the strength of the relationship? The results of the analysis of the data support the notion that there is a relationship between leadership and school effectiveness. Although it is necessary to proceed with caution because the studies are heterogeneous, Q(37) = 252.04; p <.001 (Table 4), when all the studies are taken together, the composite effect size of leadership on school effectiveness is d = 0.66; CI_{95} = 0.60 / 0.73 (Table 8). However, the strength of the relationship between leadership and school effectiveness varies by dimension, ranging from < .05 to > .05 standard deviation units.

Question 3 asked to what extent is there a relationship between selected dimensions of leadership and school effectiveness? As the results show, leadership behaviors that focused on defining the school mission yielded an overall effect size, d = 0.22; CI_{95} = 0.07 / 0.37 (Table 12), whereas the strength of the association between the instructional organization dimension of leadership and school effectiveness was stronger, d = 0.66; CI_{95} = 0.34 / 0.62 (Table 10). In answer to question 4, how do the dimensions of leadership compare in their relationship to school effectiveness, the strongest dimension is related to the instructional leadership behaviors that focus on the instructional organization of the school. However, the other two dimensions of instructional leadership yield positive but low effect sizes of 0.22 for school mission, and 0.29 for school climate, whereas the transformational dimensions of consideration and inspiration yield effect size of 0.36 and 0.40 respectively.

From the above reported results of the meta-analysis of the available studies, it can be concluded that leadership significantly influences the level of school effectiveness, some leadership behaviors having greater impact than others.

Table 12: Characteristics of studies not meeting the criteria for meta-analysis

Study Author	Sample	Dept Var	Qual	Can Correl	DDA	MAN OVA	M. Reg	SEM	DVA	IRT	EPF	Effect
Brewer, 1993	P & S	Ach									X	Yes
Eberts & Stone, 1988	P & S	Ach									X	Yes
Geisel, Sleegers & Van den Berg, 1999	T	Ch'ng Pract	X				X					Yes
Goldring & Pasternack, 1994	P	Sch Effe			X							Mixed
Hallinger, Bickman & Davis, 1996	P & T	Ach						X				Yes
Heck, 1993	T	Ach			X	X						Yes
Heck & Marcoulides, 1993	T	Ach						X				Yes
Leithwood, Jantzi, Silins & Dart, 1993	T	Stud Out C		X			X					Yes
Leitner, 1994	P & T	Ach					X					No

Study Author	Sample	Dept Var	Qual	Can Correl	DDA	MAN OVA	M. Reg	SEM	DVA	IRT	EPF	Effect
Ogawa & Hart, 1985	P & Sch											Yes
Sheppard, 1996	T	Sch Char					X					Yes
Silins, 1994	T	Sch Effect		X				X				N.A.
Van de Grift & Houtveen, 1999	T	Ach									X	Mixed (weak)

P: principal T: teachers Sch: school Stud: students Ach: achievement Char: characteristics

Ch'ng pract: changed practice DVA: decomposition of variance analysis EPF: education production function

DDA: descriptive discriminant analysis IRT: item response theory