Chapter 4: Presentation of Data Analysis

Introduction

In the midst of a dramatically changing educational paradigm, from universal access and opportunity to universal proficiency (Houston, 2004b), our nation has begun to view school leadership as a catalyst for student achievement. In this context the researcher needed to inquire into the nature of leadership in the districts and into what that implied in terms of proficiency.

This chapter not only reports what the researcher found as the demographic reality related to the research questions but also applies the Strength of Relationship Scale to the hypotheses posed in this study. Chapter 5 reports the results of analyses surrounding each of the three hypotheses. The introductory section restates the three hypotheses and their concomitant analysis objectives and summarizes the main results. Variables are defined, and the sample is described. Bivariate correlations between Independent and Dependent variables are reported. Next, in terms of Analysis Objective 1, the nature of Strength of Relationship (SOR) in relation to the predictor variables of interest is reported in detail. A section on the nature of conflict and

levels of Agreement encompasses Analysis Objectives 2 and 3. The fourth section examines the nature of Strength of Relationship (SOR) in relation to student achievement (MEAP) as suggested by Analysis Objective 4. Finally, the hypotheses are restated with the apparent specific results summarized. A Glossary of Statistical Terms can be found as Appendix J as a ready reference.

Hypotheses, Analysis Objectives, and Results

Prior research suggested three hypotheses that guided this study. In order to investigate the three hypotheses, the researcher established four analysis objectives. These are summarized here with the major findings, which are detailed later in the chapter. Two data sets were used. The *Complete* data set was defined as responses wherein all variables in the survey questionnaire were answered in full with no missing data. The *Paired* data set was derived from the Complete data set wherein both the superintendent and the board president from the district responded in full.

Hypothesis 1

Hypothesis 1: Evaluation Method. The method of evaluation of the superintendent is a significant indicator (either positive or negative depending on the type of evaluation) of the strength of

relationship between school board and superintendent (AASA and NSBA, 1980; AASA, 1992).

Analysis Objective 1: Compare evaluation type (EvalID), political type (PolID), and demographic variables in terms of means on Overall Strength of Relationship (OverSOR) and the six Strength of Relationship subscales: Evaluation (Evalsor), Political Climate (Polsor), Conflict (Confsor), Superintendent's Influence (Gensor), Teaching and Learning (TLsor), and Training (Trainsor).

This analysis was performed with the use of the Complete data set. The finding was that when pluralistic board interaction was paired with data-driven superintendent evaluation, Strength of Relationship increased approximately two-fold in most cases. Furthermore, the stronger the superintendent's influence on the district, the higher the student achievement. The operating budget and the size of the district had little impact on Overall Strength of Relationship.

Hypothesis 2

Hypothesis 2: Conflict Levels. Low levels of conflict between the board and the superintendent correlate with the Data-driven evaluation type and the Pluralistic political climate type (Cuban, 1998; Lashway, 2002a; Ondrovich, 1997; Purdy, 2002).

Analysis Objective 2: Determine what type of conflict groups with what evaluation type (EvalID) and political climate type (PolID).

This analysis was performed with the use of the individual Complete data set. Political climate type emerged as a strong predictor of conflict level. Boards that work together in a pluralistic manner are 87–93% less likely to report conflict. When pluralistic political climate paired with data-driven evaluation, conflict decreased even more and student achievement passing rate was higher.

When conflict was reported, it centered first on role definition and fulfillment and second on financial issues. The most frequently reported conflict focused on staff negotiations, an issue that spans both role definition and financial considerations, as well as external (local, state, federal) and internal forces.

Hypothesis 3

Hypothesis 3: Agreement, Overall Strength of Relationship (OverSOR), and Student Achievement (MEAP). High levels of agreement and higher Overall Strength of Relationship (OverSOR) between the board and the superintendent correlate with higher Michigan Education Assessment Program (MEAP) passing rates.

Analysis Objective 3: Determine where board presidents and superintendents agree and disagree (on selected variables) and correlate that with district-level indicators.

This analysis was performed with the use of the district-level Paired data set. The finding was that the lower the level of disagreement between the board and the superintendent, the higher the MEAP passing rate. This remained the case regardless of district size, per-pupil expenditure, or socioeconomic status.

Analysis Objective 4: Determine the relationship of Overall Strength of Relationship with MEAP passing rate (student achievement) and other district variables.

This analysis was performed with the use of the district-level Paired data set. The finding was that the higher the Overall Strength of Relationship between the board and superintendent, the higher the MEAP passing rate. This remained the case regardless of district size, per-pupil expenditure, or socioeconomic status. Student achievement was shown to be as much as 3-4 times higher in districts exhibiting a pluralistic political type in conjunction with a data-driven superintendent evaluation method.

In consideration of these summarized main results, the researcher now reports the details of these findings.

Variables

The dependent variables for Analysis Objective 1 were Strength of Relationship (sor) for the six subgroups Evaluation (Evalsor), Conflict (Confsor), Political (Polsor), General (Gensor), Training (Trainsor), Teaching & Learning (TLsor), and Overall (OverSOR). Analysis Objective 2 considered four binary indicators of certain types of conflict (Money, Roles, Other, No Conflict) as dependent variables, while Analysis Objective 3 used amount of disagreement as the dependent variable (Disagreement). Analysis Objective 4 used Michigan Education Assessment Program (MEAP) scores as a dependent variable.

The independent variables posited to have a relationship with the dependent were type of evaluation (EvalID_m and its components A = Global, B = Judgment, C = Data-driven, and D = None), political climate type (PolID and its components E = Dominated, F = Factional, G = Pluralistic, and H = Inert), district size as measured by headcount (Size), per-pupil expenditure (Operating$), socioeconomic status as measured by free and reduced-priced lunch calculations (SES), student

achievement as measured by MEAP passing rate (this variable was both independent and dependent depending on the analysis objective), age of respondent (Age), and gender of respondent (Gender). Table 3 depicts these variables in an overview format.

Table 3

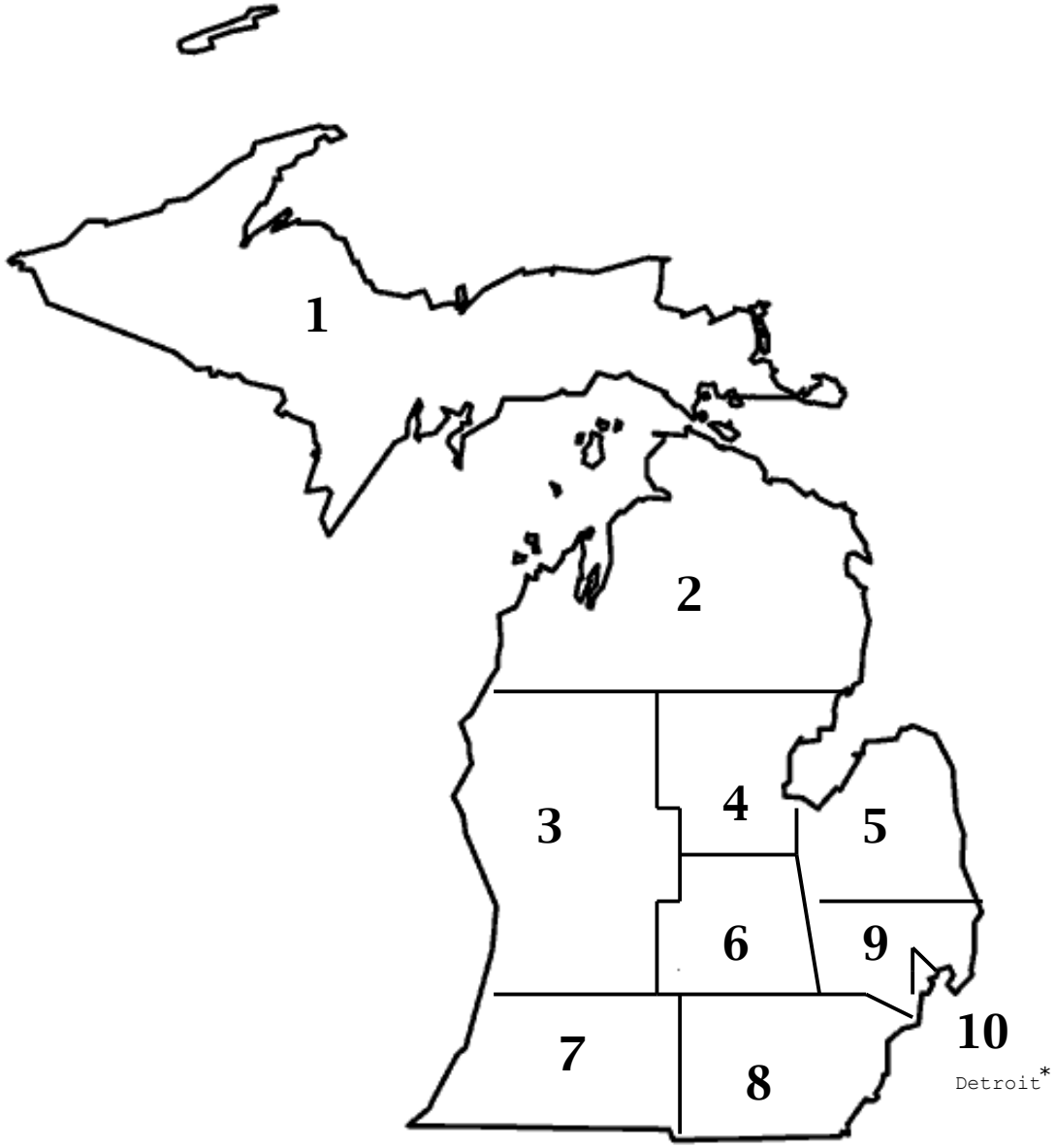
Analyses Variables and the Associated Objectives

Dependent variables/objective #	Independent variables
Evalsor / 1	EvalID
Confsor / 1	PolID
Polsor / 1	Size
Gensor / 1	$ / pupil
Trainsor / 1	SES
TLsor / 1	MEAP
OverSOR / 1,2,3,4	Age
Money / 2	Gender
Role / 2	
Other / 2	
No Conflict / 2	
Disagreement / 3	
MEAP / 4	

Sample Descriptive Statistics

Characteristics of the Population

All public school district superintendents and board presidents (N = 526 school districts, N = 1052 potential respondents) were invited to respond to the online survey. The Michigan Association of School Administrators (MASA) has identified and designated ten regions for purposes of organizing the large number of school districts in a conceptual working model. Figure 2 depicts the ten MASA Regions. Appendix I identifies the counties that compose each region.



* MASA designates Detroit as its own region, which was not reported in this study.

Figure 2. Michigan Association of School Administrators regional designations.

Sample Size and Representation

No attempt was made to preselect a sample. A self-selected sample (n = 1047, 99.5%) responded to the survey, and all Partial responses ($n_{partial}$ = 650, 62.1%) and Complete responses ($n_{complete}$ = 397, 37.9%) were tabulated. A Complete response meant that all 153 variables contained in 55 questions on the survey were responded to in full with no missing data, while all Partial responses had missing data on one or more variables.

Complete responses included Board Presidents (n_{board} = 165, 41.6%) and Superintendents (n_{super} = 232, 58.4%) from a majority of districts in the state ($n_{districts}$ = 308, 59%). Seven districts were not considered in the analysis because they were too small for effective analysis, resulting in there being 300 districts for analysis. Responses were received from board presidents and superintendents in rural (291 responses, 73.3%), suburban (105 responses, 26.4%), and urban (1 response, 0.3%) districts. Districts where both the superintendent and the board president responded were referred to as Paired responses (n_{paired} = 86, 28.6%).

Small Case Sets Eliminated

The final analyses eliminated cases reporting evaluation type "D" (None, 3 responses) and political type "E" (Dominated, 4 responses), as there were too few cases to analyze effectively. Only Complete cases were used ($n_{complete}$ = 390 respondents, $n_{districts}$ = 300 districts, n_{paired} = 86 pairs).

Chi-square Test

A Chi-square test was usd to determine if the sample distribution in terms of districts ($n_{districts}$ = 308), based on the Complete data set ($n_{districts}$ = 300), mirrored what was expected based on the distribution of districts in the MASA population. A Chi-square test is a nonparametric test of statistical significance appropriate when data are in the form of frequency counts; it compares frequencies actually observed with expected frequencies to see if they are significantly different (Agresti & Finlay, 1997). Figure 3 reports the results of a Chi-square test comparing the observed district counts in the nine regions of interest with expected district counts based on the population distribution. Figure 4 graphically depicts the close relationship of the sample to the population. The Chi-square statistic was not significant, suggesting that the observed sample distribution

of districts was statistically no different from what was expected on the basis of the population distribution.

Region

	Observed N	Expected N	Residual
1.00	26	32.4	-6.4
2.00	39	34.5	4.5
3.00	49	48.3	.7
4.00	28	22.8	5.2
5.00	36	31.8	4.2
6.00	20	21.9	-1.9
7.00	32	36.6	-4.6
8.00	31	28.5	2.5
9.00	39	43.2	-4.2
Total	300		

Test Statistics

	region
Chi-Square [a]	4.973
df	8
Asymp. Sig.	.761

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 21.9.

Figure 3. Chi-square test of the sample versus the population.

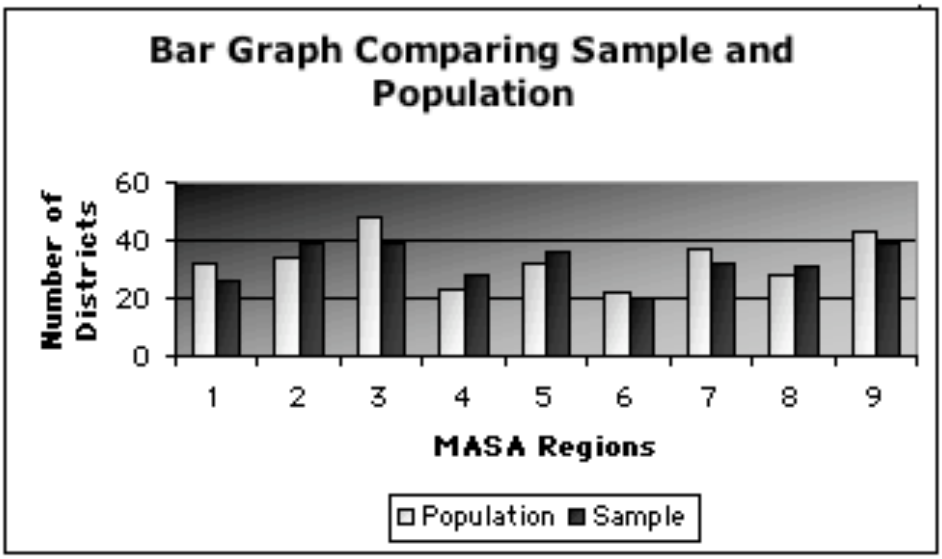


Figure 4. Bar graph of the sample relative to the population by MASA region, depicting an evenly distributed and representative sample of school districts.

On the basis of the results of the Chi-square test, the 37.9% Complete response rate was a representative sample population. The researcher made the decision to use only Complete responses for this analysis, as the Complete data provided an evenly distributed and representative sample of the entire state by region and district size from which the researcher might generalize with confidence.

Bivariate Analysis Results

A Bivariate Correlation Analysis yields a correlation coefficient, symbolized by the letter *r*, indicating the degree of relationship that exists between scores on two variables (Agresti & Finlay, 1997). Table 4 specifies the results of the Bivariate Correlation Analysis for Overall Strength of Relationship and the six subgroups (Evalsor, Confsor, Polsor, Gensor, TLsor, and Trainsor) that was conducted with district-level data as independent variables. Three distinctions became clear.

- First, the *lower* the socioeconomic status of a school district was, the *lower* the Overall Strength of Relationship and all six subgroups were.

- Second, the *higher* the Strength of Relationship in all subgroups was, the *higher* the student achievement as measured by the MEAP passing rate were.
- Furthermore, the *larger* the district size was, the *stronger* were the Conflict Strength of Relationship, Political Strength of Relationship, Teaching and Learning Strength of Relationship, and Training Strength of Relationship.

These results are depicted in Table 4.

Table 4

Bivariate Correlation Analysis Results

VARIABLES Dependent ↓	Independent → Size of district	SES	Per-pupil expenditure	MEAP
Evaluation sor	none	r=-.108, p=.033	none	r=.147, p=.004
Conflict sor	r=.104, p=.041	r=-.120, p=.018	none	r=.195, p=.000
Pollitical sor	r=.157, p=.002	r=.-192, p=.000	none	r=.223, p=.000
General sor	none	r=-.188, p=.000	none	r=.268, p=.000
Teach/Learn sor	r=.210, p=.000	r=-.215, p=.000	none	r=.239, p=.000
Training sor	r=.177, p=.000	r=-.176, p=.001	none	r=.114, p=.026
Overall SOR	none	r=-.145, p=.004	none	r=.233, p=.000

Correlation coefficients (r) and associated p-values (p) are reported in Table 4.

Relation of Predictor Variables with Strength of Relationship (SOR) Variables

Analysis Objective 1: Compare evaluation type (EvalID), political type (PolID), and demographic groups in terms of means on Overall Strength of Relationship (OverSOR), and the six Strength of Relationship subscales: Evaluation (Evalsor), Political Climate (Polsor), Conflict (Confsor), Superintendent's Influence (Gensor), Teaching and Learning (TLsor), and Training (Trainsor).

The first section on Evaluation Strength of Relationship will define statistical terminology as they are used. The following sections will not define these tests and terminology repeatedly. Appendix J: Glossary of Statistical Terminology serves as a reference for the later sections.

Evaluation strength of relationship modeling results. The Evaluation Strength of Relationship (Evalsor) outcome variable was considered in a multiple linear regression model with the aforementioned independent variables as predictors. A linear regression model is a statistical technique using a prediction equation

with two or more variables in combination to predict a criterion (Agresti & Finlay, 1997).

Fitting the initial model resulted in evidence of a violation of the normality assumption for the model residuals, as observed in the Normal Q-Q Plot of Standardized Residuals. Normality Assumptions consist of two aspects: (a) Constant Variance, or the same value for all individual cases within the extent to which scores differ from one another and (b) Normality of the Residuals, or a theoretical *bell shaped* distribution as found in *typical* populations (Agresti & Finlay, 1997). As a result, Box-Cox methodology (Box & Cox, 1964) was used in the SAS software package (PROC TRANSGREG) to determine a reasonable transformation of the response variable. Box-Cox methodology proposes algorithms for estimating optimal transformations for the achievement of normality of assumptions (Agresti & Finlay, 1997). The most reasonable transformation of the data to meet model assumptions based on the Box-Cox method was an (X+1)^3 transformation (add 1 to the outcome, and cube the result). Refitting the model with the transformed outcome variable resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Political type, evaluation type, and the interaction between the two were found to have significant relationships with Evaluation Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 5. The observed power of the sample to detect these effects was strong for each of these factors, with the standard being 80%. The *power* of a test is the probability that the null hypothesis will be rejected when there is a difference in the populations or the ability of a test to avoid Type II error (Agresti & Finlay, 1997).

Table 5

*Significant Effects on Evaluation Strength of Relationship (Evalsor)**

Variable	F value**	P value	Power
Political type	F (2,368)=13.453	p=.000	.998
Evaluation type	F (2,368)= 5.782	p=.003	.868
Interaction between Political type and Evaluation type	F (4,368)= 2.901	p=.022	.781

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

Pairwise multiple comparisons (a numerical index describing the relationship between predicted and actual scores using multiple regressions; the correlation between criterion and the *best combination* of predictors) of the estimated Evaluation Strength of Relationship (Evalsor) means in the political type (PolID) groups (Factional, Pluralistic, Inert) based on the regression model (with a Bonferroni adjustment to the significance level of 0.05 (refer to Appendix J for explanation of this adjustment) indicated the following:

- Significant differences were indicated between Pluralistic political type and both Factional (diff = -3.252, p = .005) and Inert (diff = 1.729, p=.020) political types.
- The mean in the Pluralistic group was higher than the mean in either the Factional or Inert groups.
- Pluralistic political type had a significantly higher mean Evaluation Strength of Relationship score than either Factional or Inert.

Pairwise multiple comparisons of the estimated Evaluation Strength of Relationship means in the evaluation type (EvalID) groups (Global, Judgment, Data-driven) based on the regression model (with a

Bonferroni adjustment to the significance level of 0.05) indicated the following:

- There was a significant difference between Global and Judgment (diff = -1.701, p = .005) where the mean for evaluation type Judgment was higher than the mean for evaluation type Global.
- The significant negative difference indicated that Global evaluation type had lower Evaluation Strength of Relationship in all cases in relation to Judgment evaluation type.
- In all cases, Pluralistic political type and Data-driven evaluation type were estimated to have the highest Evaluation Strength of Relationship means. However, due to the amount of variability in these estimates, these means were not found to be significantly different from the means in the other groups.
- Also significant was the interaction between how the board works together (political climate type) and how the board evaluates the superintendent (evaluation type). When combined, the two factors became highly predictive of Evaluation Strength of Relationship.

Figure 5 depicts the estimated marginal means of the transformed Evaluation Strength of Relationship (t_Evalsor) variable. Pluralistic political type (G) was consistently higher with every evaluation type.

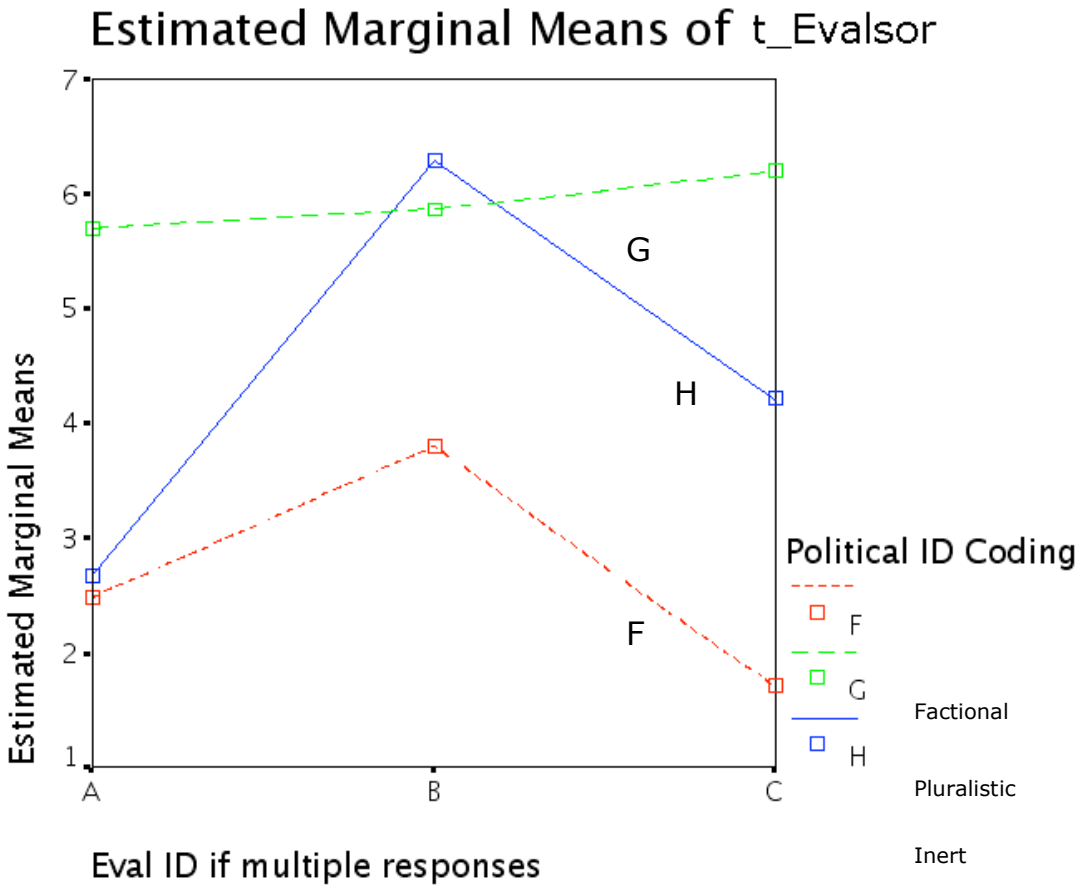


Figure 5. Estimated marginal means of transformed Evaluation Strength of Relationship (t_Evalsor).

Conflict strength of relationship modeling results. The Conflict Strength of Relationship (Confsor) outcome variable was considered in a multiple linear regression model with the aforementioned (Table 3, p. 121) independent variables as predictors. Fitting the initial model resulted in evidence of a violation of the normality assumption for the model residuals. The most reasonable transformation of the data to meet model assumptions based on the Box-Cox method was an (X+1)^1.75 transformation (add 1 to the outcome, and raise the result to the power of 1.75). Refitting the model with the transformed outcome variable resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Political type, evaluation type, and the continuous variable for student achievement (MEAP) (B = 0.012, p = .041) were found to have significant relationships with Conflict Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 6. The observed power of the sample to detect these effects was strong for the political type factor and weak for both evaluation type and MEAP (student achievement), with the standard being 80%.

Table 6

*Significant Effects on Conflict Strength of Relationship (Confsor)**

Variable	F value**	P value	Power
Political type	(2,368) =28.113	.000	1.000
Evaluation type	(2,368) = 2.908	.056	.566
Student achievement	(1,368) = 4.202	.041	.534

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

Pairwise multiple comparisons of the estimated Conflict Strength of Relationship (Confsor) means in the political type (PolID) groups (Factional, Pluralistic, Inert) based on the regression model (with a Bonferroni adjustment) indicated the following:

- A significant difference was indicated between Pluralistic political type and both Factional (diff = -1.728, p = .000) and Inert (diff = .993, p = .000) political types where the mean in the Pluralistic group was higher than the means in the Factional and Inert groups.
- Pluralistic political type had significantly higher level of Conflict Strength of Relationship (lower levels of conflict) than either Factional or Inert.

Pairwise multiple comparisons of the estimated Conflict Strength of Relationship (Confsor) means in the evaluation type (EvalID) groups (Global, Judgment, Data-driven) based on the regression model (with a Bonferroni adjustment to the significance level of 0.05) indicated the following:

- A borderline significant difference was indicated between Global evaluation type and Judgment evaluation type (diff = -.506, p =

.049) where the mean for Judgment evaluation type was higher than the mean for Global evaluation type.

- In terms of Conflict Strength of Relationship, Judgment evaluation type had consistently lower levels of conflict than Global evaluation type.
- In all cases, Data-driven evaluation type was estimated to have the highest Conflict Strength of Relationship means, and Global evaluation type was always lowest for all types of evaluation. However, due to the amount of variability in these estimates, these means were not found to be significantly different from the means in the other groups.
- Considering the MEAP passing rate (student achievement), the higher the Conflict Strength of Relationship was, the higher the MEAP passing rate was, or lower levels of conflict indicated higher student achievement.

Political strength of relationship modeling results. The Political Strength of Relationship (Polsor) outcome variable was considered in a multiple linear regression model with the aforementioned independent variables as predictors. Fitting the initial model resulted in evidence of a violation of the normality assumption for the model

residuals. The most reasonable transformation of the data to meet model assumptions based on the Box-Cox method was an (X+1)^2 transformation (add 1 to the outcome, and raise the result to the power of 2). Refitting the model with the transformed outcome variable resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Political climate type, evaluation type, and the continuous predictor variable indicating district Size (B = 0.00005, p = .045) were found to have significant relationships with Political Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 7. The observed power of the sample to detect these effects was strong for the political climate type factor and acceptable for the evaluation type and Size factors, with the standard being 80%.

Table 7

*Significant Effects on Political Strength of Relationship (Polsor)**

Variable	F value**	P value	Power
Political type	(2,368) =16.205	.000	1.000
Evaluation type	(2,368) = 3.970	.020	.710
Size of district	(1,368) = 4.029	.045	.517

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

Pairwise multiple comparisons of the estimated Political Strength of Relationship (Polsor) means in the political type (PolID) groups (Factional, Pluralistic, Inert) based on the regression model (with a Bonferroni adjustment) indicated the following:

- A significant difference was indicated between Pluralistic political climate type and both Factional (diff = 2.645, $p < .001$) and Inert (diff = 1.809, $p < .001$) political types with the mean in the Pluralistic group being higher than the mean in the Factional and Inert groups.
- Recognizing that the power to predict this effect was not strong, Pluralistic political type had a significantly higher level of Political Strength of Relationship than either Factional or Inert.

Pairwise multiple comparisons of the estimated Political Strength of Relationship (Polsor) means in the evaluation type (EvalID) groups (Global, Judgment, Data-driven) based on the regression model (with a Bonferroni adjustment to the significance level of 0.05) indicated the following:

- A significant difference was indicated between Global evaluation type and Judgment) (diff = -1.223, $p = .022$), where the mean for Judgment evaluation type was higher than the mean for Global

evaluation type. In terms of Political Strength of Relationship, Judgment evaluation type was consistently stronger than Global evaluation type.

- In all cases, Pluralistic political type and Data-driven evaluation type were estimated to have the highest Political Strength of Relationship means, and Factional political type was always lowest for all types of evaluation. However, the means for Data-driven evaluation type were not found to be significantly different from the means in the other evaluation type groups.
- Recognizing that the power to detect the Size effect was marginal, whether a district was rural, suburban, or urban had a significant positive effect on the Political Strength of Relationship. This suggests that larger districts tend to have higher Political Strength of Relationship.

General strength of relationship modeling results. The General Strength of Relationship (Gensor) outcome variable (the strength of the influence of the superintendent on the district) was considered in a multiple linear regression model with the aforementioned independent variables as predictors. Fitting the initial model, with the removal of one outlier, resulted in satisfaction of all key assumptions

behind the regression model (i.e., constant variance and normality of the residuals).

Political climate type and continuous variables Per-Pupil Expenditure (Operating$) (B = 0.00005, p = .040) and Student Achievement (MEAP) (B =0.009, p =.009) were found to have significant relationships with General Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 8. The observed power of the sample to detect these effects was strong for the political type factor and within acceptable parameters for the per-pupil expenditure and student achievement factors, with the standard being 80%.

Table 8

*Significant Effects on General Strength of Relationship (Gensor)**

Variable	F value**	P value	Power
Political type	(2,367) =11.988	.000	.995
Per-pupil expenditure	(1,367) = 4.240	.040	.537
Student achievement	(1,367) = 6.909	.009	.746

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

Pairwise multiple comparisons of the estimated General Strength of Relationship means in the political type groups (Factional, Pluralistic, Inert) based on the regression model (with a Bonferroni adjustment) indicated the following:

- A significant difference was indicated between Pluralistic political type and both Factional (diff = .707, $p < .001$) and Inert (diff = .418, $p = .008$) political types, with the mean in the Pluralistic group being higher than the mean in either the Factional or Inert groups.
- Pluralistic political type had a significantly higher level of General Strength of Relationship (stronger influence by the superintendent on the district) than either the Factional or Inert groups.
- In all cases, respondents reporting Pluralistic political type had higher means on General Strength of Relationship than did the other two political climate types. When respondents reported Data-driven evaluation type with Pluralistic political type, the means were approximately 3.5 times higher. Inert was always lowest.

- In terms of per-pupil expenditure, the more money a district has to spend in its operating budget, the higher the influence of the superintendent is.
- In terms of student achievement (MEAP passing rate), the stronger the influence of the superintendent was, the higher the student achievement was.

Teaching and learning strength of relationship modeling results. The Teaching and Learning Strength of Relationship (TLsor) outcome variable (higher TLsor relates to the *authentic* teaching and learning style, per Newman and Wehlage, 1995) was considered in a multiple linear regression model with the aforementioned independent variables as predictors. Fitting the initial model resulted in satisfaction of all key assumptions supporting the regression model (i.e., constant variance and normality of the residuals).

The continuous variable Size of District (B = 0.00002, p = .002) was found to have a significant relationship with Teaching and Learning Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 9. The observed power of the sample to detect this effect was strong, with the standard being 80%.

Table 9

*Significant Effects on Teaching and Learning Strength of Relationship (TLsor)**

Variable	F value**	P value	Power
Size of district	(1,367) = 9.325	.002	.861

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

The findings indicated that

- In terms of the effect of Size of District on Teaching and Learning Strength of Relationship, a positive significant relationship indicated that the larger the district was, the higher the Teaching and Learning Strength of Relationship was, or the higher the level of *authenticity* of teaching and learning was as defined by Newman and Wehlage (1995).
- Means of all political types increased when combined with Data-driven evaluation type, (e.g., Inert political type was approximately two times higher when combined with Pluralistic evaluation type).

Training strength of relationship modeling results. The Training Strength of Relationship (Trainsor) outcome variable was considered in a multiple linear regression model with the aforementioned independent variables as predictors. Fitting the initial model resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Age of Board Member and the continuous variable socioeconomic status (SES) ($B = -0.05$, $p = .002$) were found to have significant relationships with Training Strength of Relationship, and

continuous variable per-pupil expenditure (B = 0.0004, p = .056) was found to have a borderline significant relationship with Training Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 10. The observed power of the sample to detect these effects was strong for the Age and SES factors and was borderline for per-pupil expenditure, with the standard being 80%.

Table 10

*Significant Effects on Training Strength of Relationship (Trainsor)**

Variable	F value**	P value	Power
Age	(4,363) = 2.914	.021	.783
Per-pupil expenditure	(1,363) = 3.684	.056	.482
SES	(1,363) = 9.650	.002	.872

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

The findings indicated that

- In terms of Age, the younger the board member was, the higher the Training Strength of Relationship was, or the more likely board members were to receive training in boardsmanship.
- The higher the per-pupil expenditure was, the higher the Training Strength of Relationship was, or the higher the operating budget was, the more likely board members were to receive training in boardsmanship.
- The lower the socioeconomic status of the district was, the lower the Training Strength of Relationship was, or the less likely board members were to receive training in boardsmanship.

Overall Strength of Relationship modeling results. The Overall Strength of Relationship (OverSOR) outcome variable was considered in a multiple linear regression model with the aforementioned independent variables as predictors. The most reasonable transformation of the data to meet model assumptions based on the Box-Cox method was an (X+1)^1.75 transformation (add 1 to the outcome, and raise the result to the power of 1.75). The researcher made the decision to add 4 rather than 1 because a score of –3 was possible on this scale. The transformation was adjusted to

(X + 4) ^ 1.75. Refitting the model with the transformed outcome variable resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Political type, evaluation type, and continuous variable student achievement (MEAP) (B = 0.081, p = .005) were found to have significant relationships with Overall Strength of Relationship when adjusting for all of the other independent variables in the multivariable model, as indicated in Table 11. The observed power of the sample to detect these effects was strong for both political type and student achievement and within acceptable parameters for evaluation type, with the standard being 80%.

Table 11

*Significant Effects on Overall Strength of Relationship (OverSOR)**

Variable	F value**	P value	Power
Political type	(2,367) =30.527	.000	1.000
Evaluation type	(2,367) = 3.827	.023	.693
Student achievement	(1,367) = 7.929	.005	.802

* Nonsignificant results are not reported but were still considered in the model.

**Large effect = Large F

Pairwise multiple comparisons of the estimated Overall Strength of Relationship means in the political type (PolID) groups (Factional, Pluralistic, Inert) based on the regression model (with a Bonferroni adjustment) indicated the following:

- A significant difference was indicated between Pluralistic political type and both Factional (diff = 9.151, $p < .001$) and Inert (diff = 4.965 $p < .001$) political types, with the mean in the Pluralistic group being higher than the mean in either Factional or Inert groups.
- Pluralistic political type had a significantly higher level of Overall Strength of Relationship than either Factional or Inert.

Pairwise multiple comparisons of the estimated Overall Strength of Relationship means in the evaluation type (EvalID) groups (Global, Judgment, Data-Driven) based on the regression model (with a Bonferroni adjustment to the significance level of 0.05) indicated the following:

- A significant difference was indicated between Global evaluation type and Judgment (diff = -2.853, $p = .020$), where the means for Judgment evaluation type was higher than the means for Global evaluation type.

- Judgment evaluation type had consistently higher Overall Strength of Relationship means than did Global evaluation type.
- In all cases, Judgment evaluation type was estimated to have the highest Overall Strength of Relationship means except when combined with Pluralistic political type, in which case Data-driven evaluation type exhibited the highest means.
- In terms of student achievement (MEAP passing rate), the higher the Overall Strength of Relationship was, the higher the student achievement was.

The Nature of Conflict and Levels of Agreement

Analysis Objective 2: Determine what type of conflict groups with what evaluation type (EvalID) and political type (PolID).

The analyses of conflict type used the Complete subset of data wherein all survey questions were answered completely by the board presidents or the superintendents in the 301 districts ($n_{complete} = 390$ respondents). Conflict was reported in 120 cases, and No Conflict was reported in 266 cases. The remaining four cases had missing data on at least one of items considered in these analyses and were eliminated for this test. Table 12 depicts the broad range and frequency of

conflict reported, whereas Table 13 reports the grouping and frequency of conflict into four overarching conflict categories.

Table 12

Frequency of Conflict Type by Number of Citations

Conflict	Citations
Staff negotiations	112
Money	91
Roles	90
Hiring staff	61
Micromanagement	59
Communication	51
Leadership style	33
Discipline (staff & student)	29
Community	23
Athletics	22
Outside mandates	7
Schools of choice	5
Technology	4
Achievement	4
Total citations of conflict	591

Table 13

Conflict Types Grouped into Four Categories of Conflict

MONEY/citations		ROLES/citations		OTHER/citations		NONE/citations
Staff negotiations	112	Roles	90	Athletics	22	266
Money	91	Micromanagement	59	Mandates	7	
Hiring staff	61	Communication	51	Achievement	4	
Schools of choice	5	Leadership style	33	Technology	4	
		Discipline	29			
Total (Percentage of reported conflict)	269 (44.17%) *MONEY*		303 (49.75%) *ROLES*		37 (6.08%) *OTHER*	266(0%) *NONE*

Dummy variables indicating whether or not respondents reported the particular types of conflict identified in Table 13 (Money, Roles, Other, and No Conflict) were used to perform four separate multivariate logistic regression analyses. The dummy variables were evaluation and political types reported by the respondents and the district-level indicators as independent predictor variables (Size, SES, Operating$, and MEAP).

Money conflict analysis results. Conflict over money matters defined 44.17 percent of all conflict citations, and the logistic regression analysis indicated that none of the predictor variables had a significant relationship with the likelihood of citing money conflict.

Role conflict analysis results. Conflict over roles defined 49.75 percent of all conflict citations, and political type was found to be a significant predictor of the likelihood of reporting Role conflict when controlling for the other predictors [Wald Chi-square (2) = 23.708, $p < 0.001$]. Respondents reporting Factional political type were about 11.6 times more likely than respondents reporting a Pluralistic political type to report role conflict [Odds Ratio (OR) = 11.611, 95% CI = (3.422, 39.405)]. Respondents reporting a Inert political type were

about 9.9 times more likely than those reporting Pluralistic political type to report role conflict [OR = 9.879, 95% CI = (2.188, 44.600)].

Other conflict analysis results. Other conflict defined 6.08 percent of all conflict citations, and no predictor variables were found to have a significant relationship with the odds of reporting other types of conflict.

No conflict analysis results. No Conflict was reported in 266 cases, and political type was found to be a significant predictor of the likelihood of reporting No Conflict when controlling for the other predictors [Wald Chi-square (2) = 10.225, p = 0.006]. Respondents reporting a Factional political type were about 93% less likely than respondents reporting Pluralistic political type G to report No Conflict [Odds Ratio (OR) = 0.073, 95% CI = (0.010, 0.543)]. Respondents reporting Inert political type were about 87% less likely than those reporting Pluralistic political type to report No Conflict [OR = 0.130, 95% CI = (0.017, 1.003)].

Summary. The way a board works together and with its superintendent, also called the political type, was found to have a significant relationship with the odds of a board president or superintendent reporting either Role conflict or No Conflict. A

respondent in the Pluralistic political type was significantly *more* likely to report No Conflict and significantly *less* likely to report Role conflict. Table 14 depicts the Wald Chi-square statistics from the four categories of conflict models, and Table 15 depicts the Odds Ratios from the four categories of conflict models.

Table 14

Wald Chi-square Statistics from the Estimated Logistic Regression Models for the Four Categories of Conflict

	Money	Roles	Other	None
Data-driven	.953	.594	1.631	0.528
Pluralistic	.595	23.708*	.197	10.225*
Size of district	.011	2.486	2.582	1.407
SES	.062	.272	1.751	0.019
Per-pupil expenditure	1.498	1.140	.031	1.531
MEAP	1.410	2.031	.333	1.436

[1] Chi-square statistics have 2 df; remaining Chi-square statistics have 1 df.

* denotes $p < 0.001$

Table 15

Odds Ratios from the Estimated Logistic Regression Models for the Four Categories of Conflict*

	Money	Roles	Other	None
Global	0.085 (0.352, 1.838)	1.398 (0.582, 3.359)	4.125(0.465, 36.598)	0.801 (0.328, 1.960)
Judg-ment	1.087 (0.566, 2.085)	1.258 (0.628, 2.519)	3.469 (0.449, 26.798)	0.771 (0.382, 1.555)
Data-driven	ref	ref	ref	ref**
Faction	1.235(0.443, 3.445)	9.879(2.188, 44.600)	1.347 (0.278, 6.538)	0.130(0.017, 1.003)**
Plural	ref	ref**	ref	ref
Inert	0.780 (0.359, 1.692)	11.611(3.422, 39.405)**	0.845(0.185, 3.854)	0.073(0.010, 0.543)

* 95% Confidence Intervals (CI) have been included with the estimated odds ratios.

** Significant at $p < 0.05$

The Nature of Overall Strength of Relationship (OverSOR) and District-Level Variables and Disagreement

Analysis Objective 3: Determine the levels of board president and superintendent agreement and disagreement, and correlate that with district level indicators.

This analysis used the Paired data set (n_{paired} = 86 districts), wherein both the superintendent and the board president from each district answered all variables completely. Disagreement was calculated on the basis of the comparison of responses of the board president and the superintendent on 28 variables. Appendix H identifies the specific variables considered in the analysis of Agreement/Disagreement. Both agreement and disagreement scores were calculated, representing the number of agreements and the number of disagreements for a given district.

A multiple regression model was fitted to the data considering the number of Disagreements as the continuous dependent variable. The average of the two Overall Strength of Relationship scores for the two respondents from the district (factor scores based on Evalsor, Polsor, Confsor, and Gensor, as described in Chapter 3), as well as other district-level variables, were used as independent predictors.

Fitting the initial model resulted in evidence of a violation of the constant variance assumption for the model residuals. The most reasonable transformation of the response data to meet model assumptions based on the Box-Cox method, was a square-root transformation, which is often appropriate for stabilizing variance in count variables (such as the total number of disagreements). Refitting the model with the transformed outcome variable resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals). A problem with multicollinearity was also observed in the initial model in that MEAP (student achievement) scores were highly correlated with district-level SES (socioeconomic status). As a result, MEAP scores were retained in the model, and district-level SES was removed.

The results of the regression analysis indicated the following:

- Overall Strength of Relationship (OverSOR) was highly correlated with the number of Disagreements reported when controlling for the other district-level predictors.
- Overall Strength of Relationship had a significant negative relationship with the transformed total number of Disagreements ($B = -0.257$, $p = .001$), suggesting that a higher

mean Overall Strength of Relationship tends to result in a lower number of disagreements.

- None of the other district-level predictors were found to be significantly correlated with the total number of disagreements, but it is worth noting that MEAP scores had a borderline significant relationship with the total number of disagreements (B = -0.014, p = 0.103). This suggests that higher MEAP scores also result in a lower number of disagreements.

The Nature of Student Achievement (MEAP) and Overall Strength of Relationship (OverSOR)

Analysis Objective 4: Determine the relationship of Overall Strength of Relationship (OverSOR) with student achievement (MEAP).

The adjusted Paired data set (n_{paired} = 86 districts), wherein both the board president and the superintendent answered all variables completely, was used to construct a multiple regression model. The data were fitted to the model in consideration of student achievement (MEAP) as the continuous dependent variable. Overall Strength of Relationship (OverSOR), a factor score based on Evalsor, Polsor, Confsor, and Gensor, as described in Chapter 3, and other district-level variables (Size of District, SES, and per-pupil expenditure) were

modeled as independent predictors. The relationship between the average Overall Strength of Relationship reported by the respondents in the district and student achievement (MEAP) performance was estimated, controlling for other potential district-level predictors of MEAP performance. Fitting the initial model resulted in satisfaction of all key assumptions behind the regression model (i.e., constant variance and normality of the residuals).

Student achievement (MEAP) was found to have a significant positive relationship with Overall Strength of Relationship (OverSOR) ($B = 1.598$, $p = 0.024$) and a significant negative relationship with SES ($B = -0.411$, $p < 0.001$). Size of district was found to have a significant positive relationship with MEAP ($B = 0.0005$, $p = 0.014$) when adjusted for all of the other independent variables in the multivariable model.

In other words, the findings indicated that

- The higher the Overall Strength of Relationship was, the higher the student achievement was.
- The lower the socioeconomic status was, the lower the student achievement was.
- The larger the district size was, the higher the student achievement was.

A district achieved a high Overall Strength of Relationship by exhibiting (a) Data-driven methods of superintendent evaluation, (b) low levels of conflict, (c) pluralistic interaction among the board members, (d) high levels of influence by the superintendent, (e) authentic teaching and learning styles in classrooms, and (f) board members with more training. The closer a district came to this profile, the higher its MEAP passing rate was. Table 16 depicts the results of the regression analysis of student achievement (MEAP) with Overall Strength of Relationship (OverSOR) and the district-level predictors.

Table 16

Results of Regression Analysis of Student Achievement (MEAP) with Overall Strength of Relationship (OverSOR) and Other District Level Variables*

Variable	B	Std. error	Sig.
Overall SOR	1.598	.697	p = 0.024
SES	-.411	.049	p < 0.001
Size of district	0.0005	.001	p = 0.014

* Nonsignificant results are not reported but associated predictors were still considered in the model.

Hypotheses' Results Summarized

Hypothesis 1: Evaluation Method. Method of evaluation of the superintendent is a significant indicator (either positive or negative depending on the type of evaluation) of the strength of relationship between school board and superintendent. Findings were positive and more complex than suggested by Hypothesis 1.

Considering Overall Strength of Relationship (OverSOR) and its six subgroups (Evaluation, Conflict, Political Climate, Superintendent's Influence, Teaching and Learning Style, and Board Training), the following summarizes the results of the multiple linear regression analyses.

1. *Evaluation Strength of Relationship (Evalsor)* was significantly influenced by
    a. Political climate type (PolID) [$F(2,368) = 13.453$, $p < 0.001$],
    b. Evaluation method type (EvalID) [$F(2,368) = 5.782$, $p = .003$], and
    c. the interaction between the two factors (PolID*EvalID) [$F(4,368) = 2.901$, $p = .022$].

In all cases Pluralistic political type and Data-driven evaluation type were estimated to have the highest means. Pluralistic political type was

significantly higher than both Factional (diff = -3.252, p < 0.001) and Inert (diff = 1.523, p = .020). Judgment evaluation type was significantly higher than Global (diff = 1.701, p = .005).

2. *Conflict Strength of Relationship (Confsor*) was significantly influenced by
    a. Political climate type (PolID) [F (2,368) = 28.113, p < .001],
    b. Evaluation method type (EvalID) [F (2,368) = 2.908, p = .056], and
    c. Student achievement (MEAP, a continuous variable) [B =0.012, F (1,368) = 4.202, p = .041]. The higher the Conflict Strength of Relationship was, the higher the student achievement was, or the lower the rate of Conflict was, the higher the MEAP passing rate was.

In all cases, Pluralistic political type and Data-driven evaluation type were highest, with Pluralistic political type being significantly higher than either Factional (diff = 1.728, p < .001) or Inert (diff = .993, p < .001).

3. *Political Strength of Relationship (Polsor)* was significantly influenced by
    a. Political climate type (PolID) [F (2,368) = 16.205, p < .001],
    b. Evaluation method type (EvalID) [F (2,368) = 3.970, p = .020], and
    c. Size of District (continuous variable) [B = 0.00005, F (1,368) = 4.029, p = .045]. The larger the district was, the higher the Political Strength of Relationship was.

In all cases, Pluralistic political type and Data-driven evaluation type exhibited the highest Political Strength of Relationship means, with Pluralistic political type being higher than Factional (diff = 2.645, p < .001) and Inert (diff = 1.809, p < .001) and with Global evaluation type being lower than Judgment (diff = -1.223, p = .022).

4. *General Strength of Relationship (Gensor, the influence of the superintendent on the district)* was influenced by
    a. Political climate type (PolID) [F (2,367) = 11.988, p < .001],
    b. Per-pupil expenditure (Operating$, a continuous variable) [B = 0.00005, F (1,367) = 4.240, p = .040], wherein the higher the operating budget was, the higher the influence of the superintendent was, and

c. Student achievement (MEAP, a continuous variable) [B = 0.009, F (1,367) = 6.909, p = .009], wherein the stronger the superintendent influence on the district was, the higher the student achievement was.

In all cases, Pluralistic political type had higher means when combined with Data-driven evaluation type, and Inert was always lowest. Pluralistic political type was always significantly higher than Factional (diff = .707, p < .001) and Inert (diff = .418, p = .008).

5. *Teaching and Learning Strength of Relationship (TLsor)* was influenced by

   a. Size of District (a continuous variable) [B = 0.00002, F (1,367) = 9.325, p = .002], wherein the larger the district was, the higher the Strength of Teaching and Learning Relationship was.

   b. Means in all Political type cases (Factional, Pluralistic, Inert) increased when combined with Data-driven evaluation method type; for example, Political type Inert was approximately two times higher when combined with Data-driven.

6. *Training Strength of Relationship (Trainsor, pertaining to board members)* was influenced by
    a. Age of board member [F(4,363) = 2.914, p = .021], wherein the younger the board member was, the more training was obtained,
    b. Per-pupil expenditure (Operating$, a continuous variable) [B = 0.0004, F (1,363) = 3.684, p = .056], wherein the higher operating budget was, the more training was received, and
    c. Socioeconomic status (SES, a continuous variable) [B = -0.05, F (1,363) = 9.650, p = .002], wherein low SES indicated less training.
7. *Overall Strength of Relationship (OverSOR, a factor score including scores in all six strength of relationship groups)* was influenced by
    a. Political climate type (PolID) [F (2,367) = 30.527, p < .001],
    b. Evaluation method type (EvalID) [F (2,367) = 3.827, p = .023], and
    c. Student achievement (MEAP, a continuous variable) [B = 8.112, F (1,367) = 7.929, p = .005], wherein the higher the Overall Strength of Relationship was, the higher the student achievement (MEAP passing rate) was.

In all cases, Pluralistic political type was significantly higher than either Factional (diff = 9.151, $p < .001$) or Inert (diff = 4.965, $p < .001$). When combined with Data-driven evaluation type, Pluralistic political type means increased approximately 33% over Factional and 66% over Inert.

Hypothesis 2: Conflict Levels. Low levels of Conflict between the board and the superintendent correlate with Data-driven evaluation type and Pluralistic political climate type.

Conflict was reported by 120 of 386 respondents. Citations of conflict were grouped into four dummy variables (Money, Roles, Other, None). The dummy variables were used to fit a multivariate logistic regression model. The results were as follows:

1. Money

a. Conflict over money drew 269 citations, 44.17% of all citations of conflict.
b. The likelihood of reporting Money conflict had no significant relationship with predictor variables. In other words, conflict over money crosses all district categories.

2. *Roles*

    a. Role conflict drew 303 citations, 49.75% of all citations of conflict.

    b. Political climate type (PolID) was a significant predictor of the likelihood of reporting Role conflict [Wald Chi-square (2) = 23.708, $p < 0.001$].

    c. Factional political type was 11.6 times as likely to report Role conflict than was Pluralistic political type [OR = 11.611, 95%CI = (3.422, 39.405)].

    d. Inert political type was 9.9 times as likely to report Role conflict than was Pluralistic political type [OR = 9.879, 95%CI = (2.188, 44.600)].

3. *Other*

    a. Other conflict drew 37 citations, 6.08% of all reported conflict.

    b. The likelihood of reporting Other conflict had no significant relationship with predictor variables.

4. *No Conflict*

    a. There were 266 citations of No Conflict.

b. Political climate type (PolID) was a significant predictor of the likelihood of reporting No Conflict [Wald Chi-square (2) = 10.225, p = 0.006].
c. Factional political type was 93% less likely to report No Conflict than was Pluralistic political type [OR = 0.073, 95%CI = (0.010, 0.543)].
d. Inert political type was 87% less likely to report No Conflict than was Pluralistic political type [OR = 0.130, 95%CI = (0.017, 1.003)].
e. Pluralistic political type was significantly more likely to report No Conflict and significantly less likely to report Role conflict.

Hypothesis 3: Agreement, Overall Strength of Relationship (OSOR), and Student Achievement (MEAP). High levels of agreement and higher Overall Strength of Relationship between the board and the superintendent correlate with a higher district MEAP passing rate.

The relationship between the average Overall Strength of Relationship reported by the respondents in the district and student achievement (MEAP performance) was calculated in a model that controlled for other potential district-level predictors of MEAP performance. The results were as follows:

1. Overall Strength of Relationship (OverSOR) was found to have a significant positive relationship with MEAP (B = 1.598, p = 0.024).
2. Socioeconomic status (SES) was found to have a significant negative relationship with MEAP (B = -0.411, p < 0.001).
3. The findings indicated that the stronger the relationship between the board and superintendent was, the higher the student achievement was, and the lower the socioeconomic status of a district was, the lower the student achievement was.

Conclusion

In this chapter, each of the three Hypotheses was tested through execution of its parallel Analysis Objective(s). Detailed results of the data analyses were reported and then summarized.

The final chapter (Chapter 5: Discussion and Conclusions) will apply these results to existing research and the body of knowledge in the field and present implications of the study in terms of context, further research, and professional practice.