APPENDIX C:

REGRESSION MODEL

APPENDIX C
REGRESSION MODEL

C.1 *An Overview of Multivariate Regression and Description of Analytical Model*

In Chapter 5.0, multivariate regression was employed to examine the influence of selected student characteristics on student developmental progress by third grade. Characteristics included student membership in one of the two study cohorts (the preschool cohort or the comparison cohort) and race, gender, socioeconomic status. A fourth variable—intensity of student services, described in Chapter 5.0—was also factored into the analysis. For this analysis, the student's teacher rating for a given variable was the dependent variable, or the variable to be explained by the presence, absence, or strength of selected characteristics variables, known as independent or explanatory variables. The term dependent is used to describe this variable because its value depends on the values of the independent or predictor variables.

In the following explanation of the statistical models employed in this study, we attempt to accommodate two audiences, the reader versed in statistics and the lay reader, who is not. For the latter, regression findings were reported in Chapter 5.0 for the three developmental measures of student progress by the time they reached third grade. Examples were provided of how the values derived from the regression analyses affected teacher ratings of student progress by the third grade. In this appendix, we offer some more detailed examples. For the reader with a statistics background, a theoretical justification is provided for the regression models employed in the analyses. For this discussion, since statistics jargon is unavoidable, the lay reader is advised to focus on the section in which descriptive examples are provided to better understand the application of the regression models used to analyze third grade outcomes.

C.1.1 *Description of Regression Models Employed for the Analyses in Chapter 5.0*

C.1.1.1 *Analyzing the Part A and T-CRS Using Linear Regression*

The dependent variable (the variable to be explained by the independent variables in the model) was defined operationally as third grade teacher ratings for the Part A, the T-CRS and the WSS measures for the two study cohorts. For the Part A measure, it will be recalled that teacher ratings were assigned in four categories of assistance required by the student in the general education—No Teacher Assistance Required, assigned a value of 1; Periodic Teacher Assistance Required, assigned a value of 2; Frequent Teacher Assistance Required, assigned a value of 3; and, Continuous Teacher Assistance Required, assigned a value of 4. For the T-CRS, although tables in the previous section reported frequencies and percentages of students who achieved teacher ratings in four domains in the 51^{st} to 99^{th} Percentile, T-CRS data was also provided to MGT as rating values or integers, on a scale of 1 to 40. For the WSS, teacher ratings were reported for the three WSS domains based on a dichotomy of student progress As Expected, assigned a value for the analysis of 1 and Needs Improvement, assigned a value of 0. These variables for each measure respectively were the dependent variables of interest in terms of assessing the effect of the independent student characteristic variables and, most importantly, the effect of receiving preschool services or not.

C.1.1.2 <u>Independent Variables</u>

For the analyses reported in Chapter 5.0, the independent (i.e., explanatory) variables were those characteristics hypothesized as having an effect on the dependent variable (teacher ratings for the three outcome measures). The independent variables included:

- *Student's Minority or Nonminority status* – testing the hypothesis there was a relationship between student race/ethnicity and teacher developmental ratings.
- *Student's Gender* – testing the hypothesis there was a relationship between student gender and teacher developmental ratings.
- *Student's socioeconomic status* – testing the hypothesis there was a relationship between student socioeconomic status, measured as a function of whether or not students were eligible to participate in a school district's Free and Reduced Meal (i.e., FRM) program, and teacher developmental ratings.
- *Student's preschool special education experience* – testing the hypothesis there was a relationship between teacher ratings for students who received preschool special education services and those who did not.
- *Intensity of student services, kindergarten through third grade* – testing hypothesis there was a relationship between the location, type, frequency and time spent in receipt of kindergarten through third grade special education services.

Linear regression analysis permits simultaneous examination not only of the effects on the dependent variable of ***all*** independent variables in the multivariate model, but also the effect of each, unique variable by neutralizing, or controlling for, the effects of the other independent variables in the equation. The effect of each predictor (independent) variable on the dependent variable (teacher ratings) is expressed as the magnitude of the change in the dependent variable (y) for each unit change in the independent variable (x) plus an error term. Since the independent variable is not a perfect predictor of the dependent variable—that is, X is expressed as an imperfect predictor of Y such that one unit change in X does not leads to one unit change in Y—the error term, ε, is postulated to acknowledge the residual change in the value of Y that X, which the predictor, or independent variables, cannot explain. The goal in sound regression modeling, therefore, is to minimize this error or the residual values associated with the independent variables and to maximize the explanatory power of the independent variables hypothesized as having an effect on teacher ratings of student developmental progress.

C.1.2 Part A and T-CRS Models

Since the Part A (Participation in the General Education Classroom) and the Teacher-Child Rating Scale (T-CRS) teacher ratings were integer values, linear regression was the regression approach chosen for the analysis of these two measures. For the reader who is not versed in statistics, we are simply trying to quantify the effect of the independent variables—membership in the preschool cohort or comparison cohort, race, gender, socioeconomic status and intensity of school-age special education services—on teacher ratings of student progress by the third grade.

For the statistics-minded reader, mathematically, the multivariate linear regression model is expressed as:

$$\mathbf{Y} = \beta_0 + \beta_I X_I + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \varepsilon$$

Where: Y = teacher rating for a given student on the Part A and T-CRS.
β_0 = the constant, representing the value of Y when $X_I = 0$
β_I = coefficient representing the magnitude of X_I's effect on Y
X_I = the independent variables, race/ethnicity, gender, FRM status, the intensity of K through 3 students services, and the student's cohort membership—preschool or comparison.
ε = the error term, representing the variance in Y unexplained by X_i

This equation describes the hypothesized relationship between the dependent variable and the independent variables and was used to test the hypothesis that third grade teacher developmental ratings were higher for third grade, special education students who received preschool special education services (Y_1) than for those who did not (Y_2):

$$H : Y_1 > Y_2$$

Where: H represents the hypothesis,
Y_1 = Part A or T-CRS teacher rating values for students who received preschool special education services,
Y_2 = Part A or T-CRS teacher rating values for students who received no preschool special education services,

As reported in Chapter 5.0, we were able to accept this hypothesis in instances in which findings indicated that students in the preschool cohort received significantly more favorable teacher ratings of developmental progress than students in the comparison cohort. Results are significant (i.e., statistically significant) if it is determined that the probability of students in the preschool cohort of achieving more favorable developmental ratings, due to chance, which is calculated, was less than 5 in 100 (i.e., $p < .05$).

C.1.3 Applying Data to the Linear Model for the Part A and T-CRS: Examples

In Chapter 5.0, the discussion of independent variable effects was limited to independent variables for which the statistical impact on the Part A teacher ratings was statistically significant—that is, student cohort and school-age service intensity. This decision was made after it was determined that the inclusion of other independent variables in the model (minority status, gender and FRM status) *significantly* reduced its predictive power (i.e., what statisticians refer to as model fit) as well as the size of the subsamples analyzed due to missing data for one or all of the these independent variables. This decision applied not only to models for the Part A measure, but also to the T-CRS and the WSS, for which predicted teacher ratings could be calculated as described above for the Part A. In other words, the report of regression findings reported in Chapter 5.0 based on a model that includes only the effects of student cohort and student school-age service intensity on teacher ratings of student development for each of the three measures—the Part A, the T-CRS and the WSS—can be viewed not only as adequate for the analysis but superior to the model which included all of the independent variables described in the example above.

C.1.3.1 Predicting Student Part A Teacher Ratings Based on the Full Linear Model

An example of how a linear regression model was applied to derive predicted teacher ratings of student progress is described below for the Part A subdomain, Language and Literacy. Recalling the equation for the regression model above with β, **X** and **Y** categories articulated in the equation for a minority, male, special education student who received preschool related services and who participated in school-age, Free and Reduced Meal (FRM) programs, the following was analyzed:

Y (student's predicted Part A Language and Literacy
rating of teacher assistance required) =

the calculated value of Y when X = 0 (β_0)
+
the calculated β-value for a student who received preschool related services
+
the calculated value for a student in this category of the effect of the intensity of his school-age services
+
the calculated β-value for a FRM recipient
+
the calculated β-value for a minority student
+
the calculated value for a male student.

Based on this equation, Part A teacher rating data for the Language and Literacy subdomain were submitted to the analysis, yielding results reported in **Exhibit C-1**.

EXHIBIT C-1
PART A REGRESSION RESULTS

Independent Variables	Unstandardized Coefficients	
	B	Std. Error
Constant	1.750218	0.304814
Intensity of School-age special education services	**0.00052**	**0.00006**
Comparison Cohort (0)/Preschool Related Services (1)	**-0.10274**	**0.273505**
FRM (0 received/1 did not received)	0.094582	0.138565
Gender (0 Female/1 Male)	0.010177	0.110654
Ethnicity (0 Nonminority/1 Minority)	0.07168	0.13076

Note: Categories and values in bold text connote statistically significance ($p < .01$). In the list of independent variables, variables with only two categories were assigned values of 0 and 1 referred to in the discussion below.

To derive the predicted Part A teacher rating from the regression model for this student in comparison with a peer who received no preschool special education services, values from **Exhibit C-1** are substituted in the equation as follows.

Predicted Part A Teacher Rating = 1.75 + (-.10 x 1) + (0.00052 x 361.14[1]) + (0.09 x 0) + (0.01 X 1) + (0.07 x 1)

Predicted Part A Teacher Rating = 1.9

Recalling the Part A rating scale for teacher assistance required in the third grade classroom-- No Assistance Required = 1, Periodic Assistance Required = 2, Frequent Assistance Required = 3, Continuous Assistance Required = 4 -- the predicted teacher rating for a minority, male student who received preschool Related Services and who participated in school-age, Free and Reduced Meal (FRM) programs of 1.9 suggests this student would require periodic assistance in general education.

Applying values for the independent variables derived from the analysis using the linear regression model described above, predicted Part A teacher ratings of assistance required in the general education (i.e., the dependent variable) can be derived for students for any combination of independent variables (i.e., cohort, school-age service intensity, gender, FRM and ethnicity status).

C.1.3.2 <u>Predicting Student T-CRS Teacher Ratings Based on the Full Linear Model</u>

An example of how a linear regression model was applied to derive predicted teacher ratings of student progress is described below for the T-CRS subdomain, Task Orientation. Once again recalling the equation for the regression model above with β, **X** and **Y** categories articulated in the equation for a minority, male, special education student who received preschool related services and who participated in school-age, Free and Reduced Meal (FRM) programs, the following was analyzed:

Y (student's predicted T-CRS Task Orientation

[1] For this hypothetical student, the mean value representing intensity of school-age special education services received by student who received preschool related services was chosen.

Teacher rating of student developmental progress) =

the calculated value of Y when X = 0 (β_0)
+
the calculated β-value for a student who received preschool related services
+
the calculated value for a student in this category of the effect of the intensity of his school-age services
+
the calculated β-value for a FRM recipient
+
the calculated β-value for a minority student
+
the calculated value for a male student.

Based on this equation, Part A teacher rating data for the Language and Literacy subdomain were submitted to the analysis, yielding results reported in **Exhibit C-2**.

EXHIBIT C-2
T-CRS REGRESSION MODEL RESULTS

Independent Variables	Unstandardized Coefficients	
	B	Std. Error
Constant	25.06738	1.204335
Intensity of School-age special education services	**0.002085**	**0.000299**
Comparison Cohort (0)/Preschool Related Services (1)	0.662421	0.877065
FRM (0 received/1 did not received)	0.838379	0.980191
Gender (0 Female/1 Male)	-1.33412	0.721582
Ethnicity (0 Nonminority/1 Minority)	1.393289	0.884918

Note: Categories and values in bold text connote statistically significance ($p < .01$). In the list of independent variables, variables with only two categories were assigned values of 0 and 1 referred to in the discussion below.

To derive the predicted T-CRS teacher rating of student developmental progress the regression model for this student in comparison with a peer who received no preschool special education services, values from **Exhibit C-2** are substituted in the equation as follows.

Predicted Part A Teacher Rating = 25.1 + (0.0021 x 361.14[2]) + (0.66 x 1) + (0.83 x 0) + (-1.33 x 1) + (1.39 x 1)

Predicted Part A Teacher Rating = 26.5

Based on the full T-CRS regression model, therefore, the predicted teacher rating of developmental progress for a minority, male student who received preschool Related Services and who participated in school-age, Free and Reduced Meal (FRM) programs is 26.5 of a total of 40 points possible on the Teacher-Child Rating Scale.

[2] For this hypothetical student, the mean value representing intensity of school-age special education services received by student who received preschool related services was chosen.

The following section applies the full regression model to derive predicted teacher ratings of student development based on the WSS analysis. Because the WSS employed logistical regression rather than linear regression, a brief description of logistical regression and how it works precedes that discussion.

C.1.4 Deriving the Logistical Regression Model from the Simple Linear Model to Analyze WSS Findings

Because teacher ratings for the Work Sampling System were of just two categories (i.e., teacher development ratings of Needs Improvement assigned a value of 0 and As Expected, a value of 1) a logistical regression model was employed to assess the effect of the student characteristic variables and preschool special education and no-preschool special education services on teachers ratings. For the reader who is not versed in statistics, we are trying to quantify the likelihood that a student who progressed As Expected was influenced by the student having received preschool special education services or not, and the student's race, gender and socioeconomic status (i.e., the independent variables, or those variables which explain the likelihood of the student receiving a rating of 0 or 1).

The logistical regression model can be derived with reference to the equation for the simple linear regression model described above and summarized as:

$$E(Y) = \mu = \sum_{k=1}^{K} \beta_k \; x_k$$

in which Y is the dependent variable and μ represents the expected values of Y as a function of the effects of β, the explanatory variables. In linear regression, when we study a random distribution of Y values of the dependent variable (unlike logistical regression which examines categorical variables, linear regression requires that the dependent variable be an integer value), we specify its expected values as the function of a linear combination of K unknown parameters and the covariates or explanatory variables. When this model is applied to data in the analysis, we are able to quantify the link between the dependent variable and the explanatory or independent variables.

The logistical model can be derived from the linear model above by introducing a new term, η, into the linear regression equation described above such that $\eta = \mu = \sum_{k=1}^{K} \beta_k \; x_k$.

Recalling that linear regression was the appropriate model for an analysis of integer values (1, 2, 3, 4, 5, etc.) when the data are randomly distributed, when we employ linear regression, the link between η and μ is linear. However, in the case of the WSS, since WSS teacher ratings consisted of only *two* possible, categorical values -- 0 (Needs Improvement/ 1 (As Expected) these were *binomially* distributed (i.e., not distributed randomly as integers). To accommodate the binomial nature of this variable, the link between η and μ becomes $\eta = \log[\mu/(1-\mu)]$ and logistical regression is utilized to determine the relationship between the dependent variable and the explanatory variables, calculated as a probability value (e.g., the probability of receiving a rating of 1, or progressing As Expected if one received preschool special education services). The logistical regression model is expressed mathematically as:

$$\log[\mu/1(1-\mu)] = \alpha + \beta_i X_n + \varepsilon$$

- Where:

$(\mu/1-\mu)$ = the probability of receiving a teacher developmental rating of 1, or As Expected

α = a constant value

β_i = a coefficient representing the slope of the line, based on the calculated relationship between the dependent and independent variables (i.e., race, gender, socioeconomic status and intensity of school-age special education services)

X_n = values of the independent variables for each student in relation to their WSS teacher rating

ε = the error term, representing the variance in the distribution of Y teacher ratings unexplained by that cannot be explained by associated values of the independent variable, X_i

Applying this model to determine the relationship between a single categorical variable (0 = teacher rating of Needs Improvement/1 = teacher rating of As Expected) and a set of characteristics hypothesized to influence the probability of finding a 0 or 1 value for the categorical variable, we were able to illustrate not only the extent to which a characteristic can increase or decrease the likelihood that the categorical variable will be 0 or 1, but also if the effect of each of the influencing characteristics is positive or negative in relation to membership in either the preschool services cohort or the comparison cohort.

C.1.4.1 Predicting Student WSS Teacher Ratings of Student Development Based on the Logistic Model

An example of how the logistic regression model was applied to derive predicted teacher ratings of student progress is described below for the WSS subdomain, Language and Literacy. Since WSS teacher ratings of student development progress represented categories (As Expected and Needs Development) and, unlike ratings for the Part A and T-CRS, were not number values, per se, the interpretation of the logistic model differs somewhat from the interpretation of the examples described above. That is, for the logistic model, effects of the independent variables on the dependent variable are described in terms of the odds that they will lead to a developmental rating of As Expected. **Exhibit C-3** below reports regression results for the WSS.

EXHIBIT C-3
WSS REGRESSION MODEL RESULTS

Independent Variables	Unstandardized Coefficients		
	B	Std. Error	Exp(B)
Constant	1.317763	0.285197	3.735058
Intensity of School-age special education services (0, < median/1, > median)	0.655484	0.240223	1.926074
Comparison Cohort (0)/Preschool Related Services (1)	0.368165	0.296664	1.44508
FRM (0 received/1 did not receive)	-0.3236	0.271412	0.723537
Gender (0 Female/1 Male)	-0.00396	0.199683	0.996049
Ethnicity (0 Nonminority/1 Minority)	**-0.7105**	**0.271584**	0.491397

Note: Categories and values in bold text connote statistically significance ($p < .05$). In the list of independent variables, variables with only two categories were assigned values of 0 and 1 referred to in the discussion below.

Applying the findings reported in **Exhibit C-3** to derive the predicted odds of achieving an As Expected teacher developmental rating for the example of the minority, male, special education student who received preschool related services and who participated in school-age, Free and Reduced Meal (FRM) programs, among the five independent variables on the variables, only intensity of school-age services and minority/ nonminority status achieved statistically significant effects. These effects are expressed in terms of odds ratios [Exp(B) values] for service intensity and ethnicity, 1.926074 and 0.491397, respectively. In other words, when the effect of the other independent variables in the equation were controlled, the odds of a minority student receiving a developmental rating of As Expected in the WSS domain of Language and Literacy were approximately half the odds of a majority student [Exp(B)=0.491397] receiving an As Expected rating. Similarly, the odds of a student who received less intense school-age special education services of receiving a developmental rating of As Expected in the WSS domain of Language and Literacy were nearly two times greater than those who received more intense school age special education services [Exp(B)=1.926074]. From these findings for the Language and Literacy subdomain, it can be concluded that being a member of a minority, as compared to majority students, regardless of cohort, significantly reduced the odds of receiving an As Expected developmental rating *and* that students who received less intense school-age special education services were significantly more likely to have received an As Expected teacher developmental rating. The latter of these two findings would seem to support the tautology that, regardless of cohort, the needs of students who received less intense school-age special education services were also substantially of a lesser degree than students who required more intense special education services *and* that for this subdomain receiving or not receiving preschool special education services had no bearing on results.

C.1.5 Concluding Comments

The preceding application of the linear regression model in the cases of the Part A and T-CRS and the logistic model in the case of the WSS were intended to demonstrate how these models are used to determine predicted student teacher ratings in each case. Since findings reported in Chapter 5.0 represented superior predictive models including only the school-age service intensity and preschool special education and comparison variables, the preceding discussion which included other demographic variables was intended largely as a demonstration of how regression works, rather than as a basis for any inferences regarding the effects on students outcomes of any of the demographic variables originally employed in the analysis. A model in which a larger number of theoretically justifiable variables-- in this case, a model with the cohort variable, the intensity variable *and* the demographic variables-- would have been preferable to one including only the first two (as reported in Chapter 5.0). Missing information in any one of the demographic variable categories had the effect of excluding those cases from the analyses, especially for the comparison cohort, reducing the number of cases to be analyzed and, ultimately, the predictive power of the regression models. Consequently, any inferences regarding variable effects should be restricted to the findings reported in Chapter 5.0.