DOCUMENT RESUME

ED 137 385 TM 006 191

AUTHOR	Gabriel, Blyth Riegel; And Others
TITLE	An Empirical Examination of Three Models for Estimating the Effects of No-Treatment.
PUB DATE	[Apr 77]
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977) ; Not available in hard copy due to marginal legibility of original document
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; *Achievement Gains; *Compensatory Education Programs; Control Groups; Criterion Referenced Tests; Elementary Education; *Mathematical Models; Mathematics; Multiple Regression Analysis; Norm Referenced Tests; Norms; Program Effectiveness; *Program Evaluation; Reading Achievement; Test Interpretation
IDENTIFIERS	Elementary Secondary Education Act Title I

ABSTRACT

An evaluation system for use with Elementary and Secondary Education Act (ESEA) Title I programs has been developed by RMC Research Corporation of Mountain View, California under contract with USOE. The system presently addresses cognitive achievement impact using three statistical designs, each of which may be implemented using either norm referenced or nonnormed tests. The reading and mathematics components of the District of Columbia Title I program was evaluated using both types of tests and two of the three models. The third analysis design was initially considered for implementation, but serious violations of its requirements by the data disqualified that model for eventual usage with this year's District of Columbia data. This paper presents a description of the three models and the results derived from implementing two of the three. Additionally, a differential growth rate associated with development as evidenced in the norms tables of various currently used instruments is discussed and potential areas of further research are highlighted. (Author/RC)

AN EMPIRICAL EXAMINATION OF THREE MODELS FOR ESTIMATING THE EFFECTS OF NO-TREATMENT

Blyth Riegel Gabriel
Alfred Jackson Stenner IV
Joan B. Troy

An evaluation system for use with ESEA Title I programs has been developed by RMC Research Corporation of Mountian View, California under contract with USOE. The system presently addresses cognitive achievement impact using three statistical designs, each of which may be implemented using either norm-referenced or nonnormed tests. The reading and mathematics components of the District of Columbia Title I program were evaluated using both types of tests and two of the three models. The third analysis design was initially considered for implementation, but serious violations of its requirements by the data disqualified that model for eventual usage with this year's District of Columbia data. This paper presents a description of the three models and the results derived from implementing two of the three. Additionally, a differential growth rate associated with development as evidenced in the norms tables of various currently used instruments is discussed and potential areas of further research are highlighted.

Paper presented at the American Educational Research Association Convention, New York City, April 6, 1977.

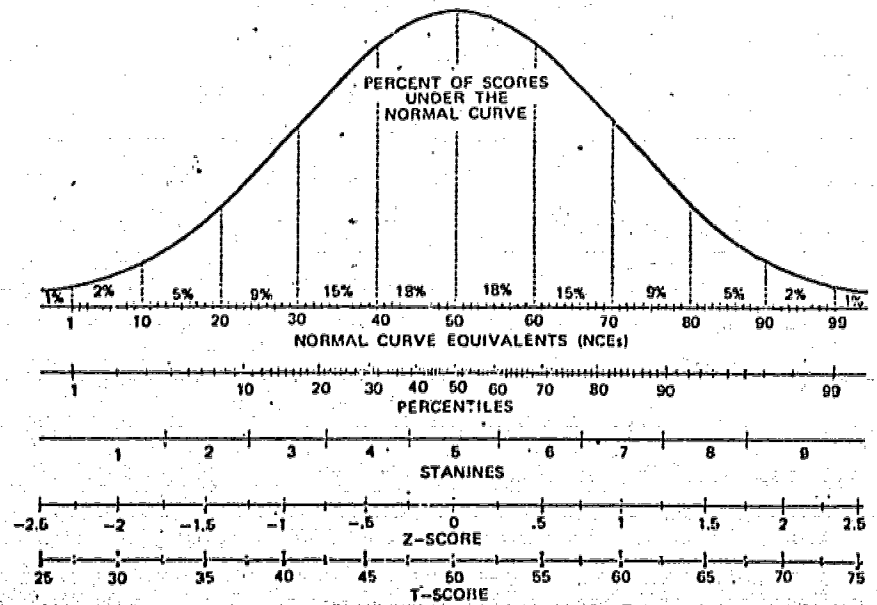
Session 17.11a

An Empirical Examination of Three Models for Estimating the Effects of No-Treatment

The United States Office of Education (USOE) contracted with RMC Research Corporation of Mountain View, California, three years ago to develop an evaluation and reporting system for nationwide use with ESEA Title I programs. The resulting evaluation package consists of three statistical models, each of which may be implemented using either norm-referenced or criterion-referenced tests (NRTs or CRTs, respectively). These models address the cognitive impact of Title I programs as measured by achievement gains.

The metric used to assess program impact in this system is the normal curve equivalent (NCE). This metric is a normalized standard score which has been linearly transformed to match the percentile rank scale at the 1st, 50th, and 99th percentile points. The NCE scale is simply a standard score scale which, for ease of interpretation, may be viewed as an equal interval percentile scale. NCEs have a range of 1-99, a mean of 50, and a standard deviation of 21.06. One advantage of NCEs is that, due to their equal interval characteristic, any mathematical operations may be performed. Another is that gain scores are easily computed, whereas grade equivalents and percentiles, which are not equal interval, do not lend themselves so easily to gain score analysis. A further inducement to use NCEs is that, in the near future, USOE will probably recommend that they become part of the evaluation system. At the same time, one drawback of the NCE score is that it can easily be misinterpreted to be a percentile score, and vice versa; this type of misunderstanding facilitates both improper interpretation and manipulation of both NCEs and percentiles. Figure 1 illustrates the relationship between NCEs, percentiles, stanines, and Z-scores. A more complete discussion of these interrelationships can be found in Chiang and Rosen, 1970.

Figure 1



Before discussing the cognitive achievement evaluation results, a brief description of the District of Columbia Title I program is necessary. The District of Columbia Title I program served approximately 17,000 students in grades K-3 and 7 during the 1975-76 school year. They attended Title I eligible schools and fell below the fiftieth percentile on the Comprehensive Test of Basis Skills (CTBS), Form S using the level of this instrument appropriate for the students' respective grade levels. National norms were used for kindergarten and first grade selection, while local norms were implemented in the three upper grade levels. Both reading and mathematics were emphasized by the Title I program. Participating students in the program were exposed to supplementary instructional strategies, both in their regular classrooms and in special resource laboratories. A primary objective of the District of Columbia Title I program is to effect significantly enhanced levels of achievement in both reading and mathematics.

To assess the impact in cognitive achievement, the CTBS/S was given to the Title I students in the fall and spring. An existent districtwide testing program additionally supplied spring criterion-referenced test (CRT) scores for Title I students, both for the 1974-75 and the 1975-76 school years. The CRTs used were the Prescriptive Mathematics Test (PMT) and the Prescriptive Reading Test (PRT). These CRT scores enabled a spring-spring analysis to be performed, in addition to the fall-spring analysis using the CTBS/S scores.

Program Impact in Reading and Mathematics

In terms of evaluation, at least two types of information are needed to determine whether a Title I project has resulted in improved student performance. The first involves an assessment of how the project students performed on outcome measures such as reading comprehension and mathematical computation after participating in the Title I project. The second requires an estimate of expected student accomplishment, given the provision that the students have not participated in the program. If the observed accomplishment of project students exceeds their expected performance, and if the difference is both statistically significant (manifesting a greater difference than can be attributed to chance fluctuation in the scores) and practically relevant (large enough to be educationally meaningful), then the Title I project is considered to be educationally salient.

It is a relatively straightforward procedure to calculate how well the project students performed on the outcome measures, but it is considerably more difficult to estimate how the project students would have performed with no treatment. Several approaches are available for assessing "no-treatment performance" or what the student would have achieved had there been no special project. This next section presents the results of two such approaches to estimating program impact.

Earlier it was stated that a primary objective of the Washington, D. C. Title I program is to improve reading and mathematics achievement among participating students to an extent that is statistically and educationally significant. Within this framework, treatment effect is the observed posttest performance minus the expected no-treatment posttest performance. Thus,

TREATMENT EFFECT	=	OBSERVED POST TREATMENT PERFORMANCE	-	EXPECTED POST NO-TREATMENT PERFORMANCE

The observed post treatment performance is simply the mean post-test score for Title I students on either the CTBS/S or the PRT and PMT. The no-treatment expectation is derived using two complementary models in an effort to converge on a valid estimate of impact (Bessey, Rosen, Chiang, and Tallmadge, 1976).

Norm-Referenced Model Results

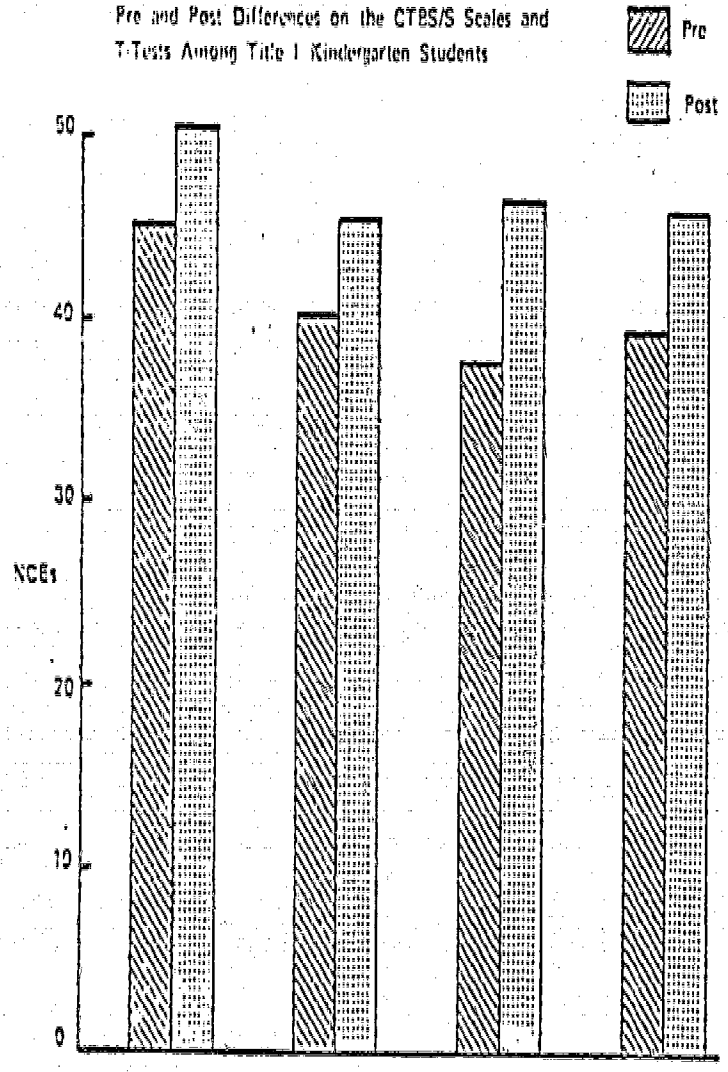
With the norm-referenced model, the impact of the Title I program was computed as follows. The pretest percentiles of each student within the treatment group were converted to NCEs and averaged. A similar procedure was followed for posttest scores. Finally, the average pre and post NCE values were compared under the assumption that, without the Title I program, the treatment group would maintain its standing relative to the norm group. Stated another way, the pretest and posttest mean NCE scores should have been similar if the project had had no impact.

There are four assumptions which should be met if this model is to yield an unbiased estimate of program impact: (1) the pretest should not be used to select project participants; (2) the test must be given at the time(s) of the year when the test was normed; (3) comparable pretest and posttest forms must be used; and (4) only those students having both pretest and posttest scores should be used in the analyses. The present application of this model at the second, third, and seventh grade levels meets all but one of the assumptions. Both the kindergarten and first grade data, however, satisfy all of the requirements. The CTBS/S was normed only in the spring for the second, third, and seventh grade levels; subsequently, fall norms were linearly interpolated from the spring data. To the extent that student learning throughout the year is nonlinear, the model may yield a biased estimate of program impact at the second, third, and seventh grade levels.

The pre- and posttest results expressed in NCEs for kindergarten, first, second, third, and seventh grade Title I students will be utilized in order to illustrate the gains in achievement which Title I students enjoy. Figures 2, 3, 4, 5, and 6 present the actual data for grades K-3 and 7, respectively, of the D. C.

Figure 2

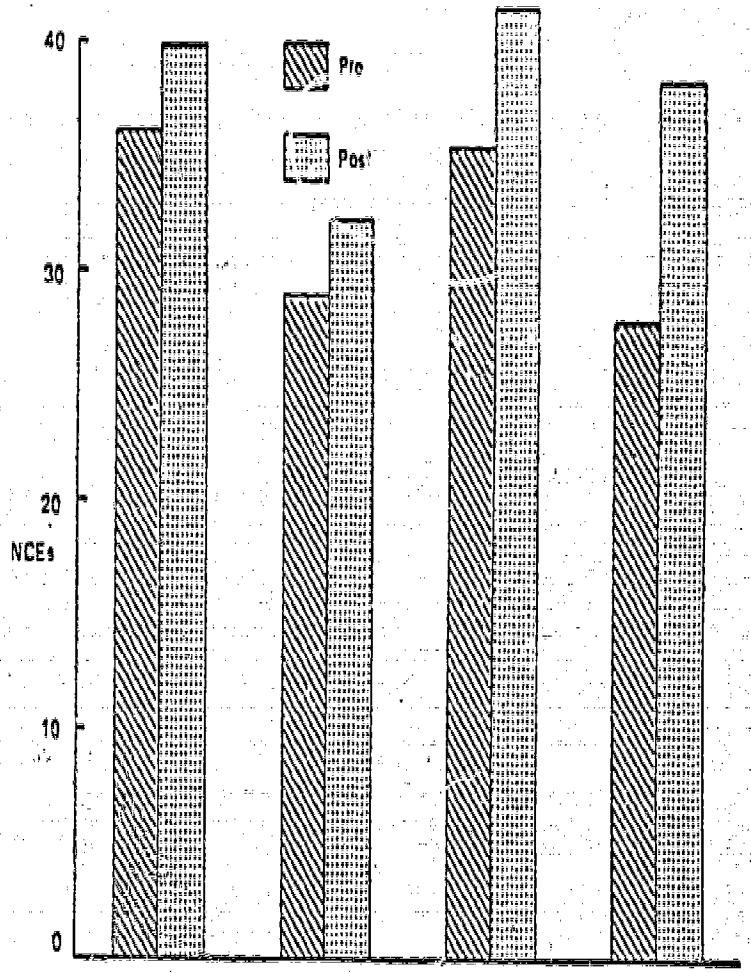
Pre and Post Differences on the CTBS/S Scales and T-Tests Among Title I Kindergarten Students



	Alphabet Skills		Visual & Auditory Discrimination		Pre-Reading		Mathematics	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
$\bar{X}$	45.1	50.8	40.2	45.2	37.7	46.5	39.1	45.9
S.D.	15.46	19.30	16.94	22.10	16.57	19.07	18.14	21.18
Difference	5.7		5.0		8.8		6.8	
N	1203		1167		963		1235	
T	11.30		6.12		16.54		10.42	
p	<0.001		<0.001		<0.001		<0.001	

Figure 3

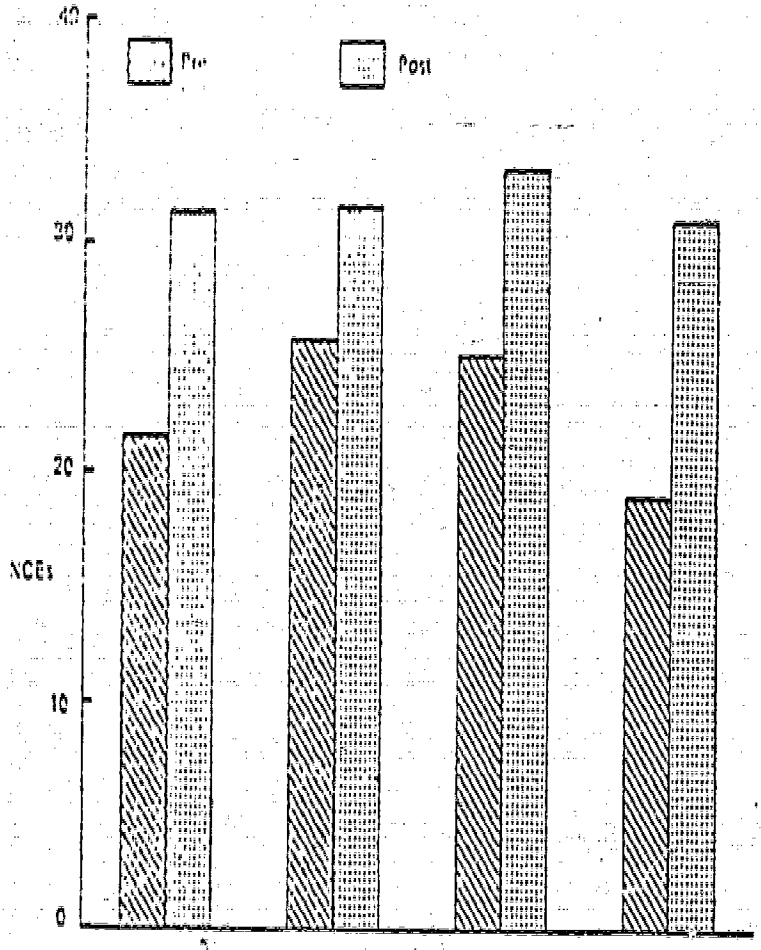
Pre and Post Differences on the CTBS/S Scales and T Tests Among Title I First Graders



	Reading		Language		Mathematics		Total Battery	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
$\bar{X}$	36.1	39.9	28.9	32.2	35.6	41.8	27.9	38.5
S.D.	17.43	17.56	14.17	16.15	16.99	19.77	15.09	10.6
Difference	3.9		3.3		6.2		10.6	
N	1041		1013		1013		829	
T	6.31		6.13		9.14		17.29	
p	<0.001		<0.001		<0.001		<0.001	

Figure 4

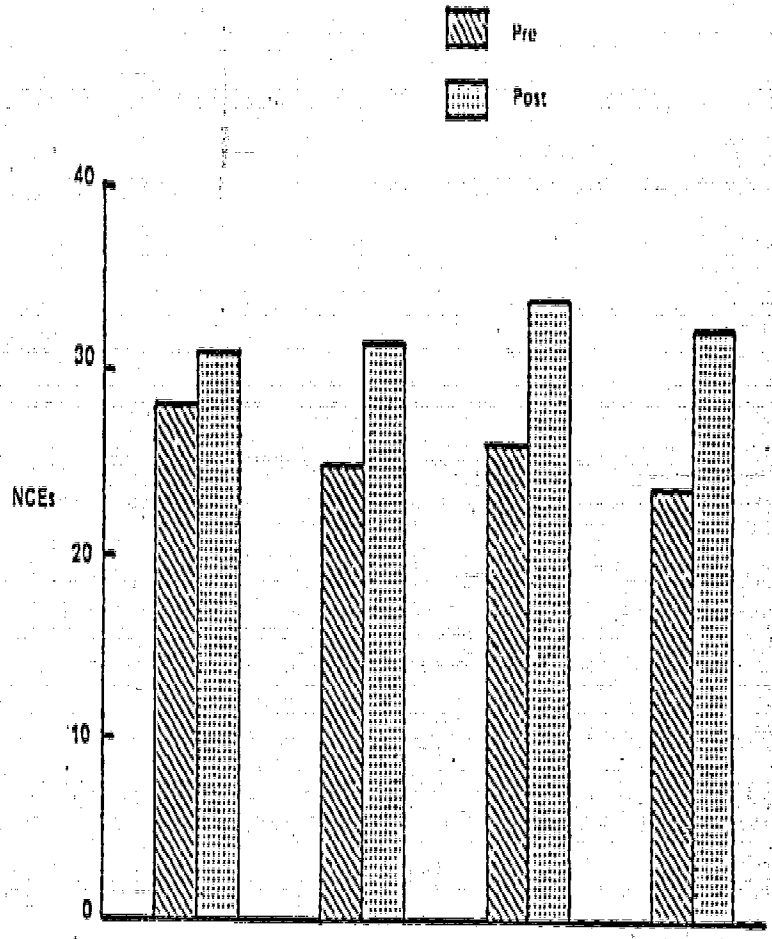
Pre and Post Differences on the CTBS/S Scales and T Tests Among Title I Second Graders



	Reading		Language		Mathematics		Total Battery	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
$\bar{X}$	21.6	31.5	25.9	31.7	25.1	33.2	19.0	31.0
S.D.	12.07	15.85	13.45	17.36	14.10	16.98	12.37	15.60
Difference	9.9		5.8		8.1		12.0	
N	980		966		947		774	
T	26.94		10.33		15.43		25.32	
p	<0.001		<0.001		<0.001		<0.001	

Figure 5

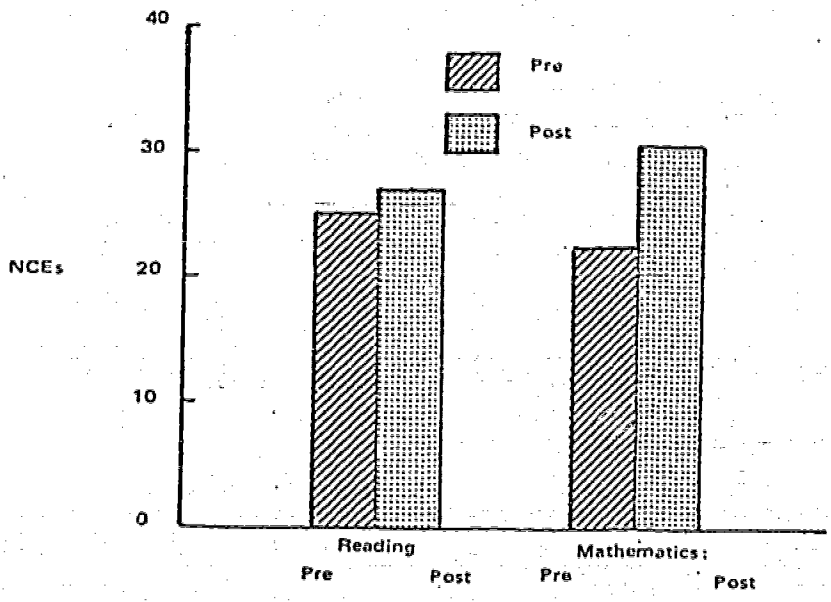
Pre and Post Differences on the CTBS/S Scales and T Tests Among Title I Third Graders



	Reading		Language		Mathematics		Total Battery	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
$\bar{X}$	28.0	31.1	24.9	31.4	26.0	33.9	23.5	32.1
S.D.	11.05	13.10	12.59	14.10	12.05	13.77	11.05	12.71
Difference	3.1		6.5		7.9		8.6	
N	1210		940		1094		821	
T	8.26		14.43		20.43		22.60	
p	<0.001		<0.001		<0.001		<0.001	

Figure 6

Pre and Post Differences on the CTBS/S Scales and T-Tests Among Title I Seventh Graders



	Total Reading		Mathematics	
	Pre	Post	Pre	Post
$\bar{X}$	25.2	26.8	22.5	30.6
S.D.	15.09	14.90	14.91	16.88
Difference	1.6		8.1	
N	395		305	
T	2.69		10.40	
p	<0.007		<0.001	

Title I program.* The mean pretest-posttest differences for all CTBS/S scales presented in these figures are statistically significant at a confidence level greater than 0.999 ($p < 0.001$) except for the Reading scale for seventh grade. This scale (see Figure 6) displays a statistically significant difference at a confidence level of 0.99 ($p < 0.01$). The mean differences range from 1.6 on Reading in grade three to 12.0 on the Total Battery for grade two. The median of these mean differences is approximately 6.65 across the five grade levels. The Mathematics mean differences tend to surpass those on the Reading scale for grades 1, 3, and 7 but not the second grade. Using a rule of thumb applied by Resource Management Corporation, exemplary gains are denoted by mean pre to post differences of 7.0 NCEs or more. Hence, using at least the Total

* Each scale on the CTBS/S, including total scales, are standardized and normed separately. Hence, the total battery score is normed by taking the score derived from all the items on the CTBS/S and not by forming a linear composite of the three skill areas total scale scores.

Battery scales, exemplary gains have been shown in grades 1-3 and in the prereading component in kindergarten. In the mathematics component of seventh grade, the gain can also be called exemplary.

Relative to the scale standard deviations, the pre- and posttest differences depict even more sharply their significance. Thirteen of the eighteen scores for scales represented in Figures 2 through 6 have mean differences which are at least one-third as large as the corresponding standard deviations. The Visual and Auditory Discrimination scale in kindergarten, Reading and Language scales at first grade, and the Reading scales at the third and seventh grades do not have mean differences which are at least one-third as large as the scale standard deviations. On nine of the eighteen scales, the mean differences are at least half the size of the corresponding standard deviations. The total Battery mean differences at the first and second grade levels particularly illustrate this point. All of these results for the five grade levels lend firm support to the contention that treatment effect is distinctly visible.

As noted earlier, the data for grades two and three violate one of the assumptions of the norm-referenced model. However, the kindergarten and first grade data, which do satisfy all of the requirements of this model, reflect statistically significant differences between pre- and posttest means for all of the CTBS/S scales. A similar result is documented by the second and third grade data, although the results at these levels are somewhat more substantial than those at the kindergarten and first grade. Thus, it is possible that the violation of the one assumption at the upper grade levels does not seriously and adversely affect the inferences which may be drawn from the second, third, and seventh grade data.

Control Group Model Results

The second analysis design implemented in this evaluation is called the control group model. As its name suggests, this evaluation design calls for the construction of control and treatment groups, both selected at random from an initial population of

eligible Title I students. The initial population should be as similar as possible with respect to all educationally relevant characteristics, such as age, sex, race, ethnicity, socioeconomic status, and measured, pretreatment achievement levels. After assignment to the treatment or control group, each student is taught and treated equally, the single exception being the application of the Title I program services to those students in the treatment group. The observed post treatment effect is derived from the actual average performance of the treatment group. The expected no-treatment effect is represented by the measured average performance of the control group.

In the present application of the control group model, raw scores on criterion-referenced reading and mathematics tests are compared. NCE gains can be derived, algebraically, by dividing the difference between the treatment group's posttest raw score mean and the no-treatment expectation by the standard deviation of the national sample and subsequently multiplying by 21.06. Through this procedure raw score gains can be converted to NCE gains. Unfortunately, there is no national sample standard deviation for the PRT and PMT, and it becomes necessary to make the following assumption: the ratio of the treatment group's standard deviation to the standard deviation of the national sample on the norm-referenced test is equal to the ratio of the treatment group's standard deviation to the national sample's standard deviation on the PMT and PRT. That is,

$$\frac{S_{NRT}}{\sigma_{NRT}} = \frac{S_{CRT}}{\sigma_{CRT}}$$

where "S" represents the treatment group's standard deviation and "σ" represents the national sample's standard deviation. Since the two treatment group standard deviations can be calculated from the collected data and the standard deviation of the national sample on the normed test can be obtained from that test's technical manual, the estimated national sample's standard deviation on the PMT and PRT test can easily be derived (Tallmadge and Wood, 1976). Under the above assumption, the raw score gains have been converted to NCE gains to permit comparisons between treatment effect estimates yielded by the norm-referenced model and control group model, respectively.

Title I schools are selected according to a weighted index comprised of the total number and percentage of economically disadvantaged students as indicated by eligibility for free lunch and low family income. The control group model contrasts thirteen schools immediately below the cutoff with thirteen schools immediately above the cutoff. The rationale for the model is this: among the schools near the cutoff, it is largely chance which determines eligibility for Title I services. In other words, the schools immediately above and below the line do not substantially differ on educationally relevant variables. Thus, those schools not receiving Title I services can fairly act as a control group for those schools which operate Title I programs. This is because

students within ineligible schools, even though their achievement levels might indicate a need for supplementary aid, do not receive any Title I services. Tables 1 through 6 give the pretest, posttest, no-treatment expectation, and treatment effect on the Prescriptive Reading Test and Prescriptive Mathematics Test, respectively, for non-Title I and Title I schools for first, second, and third grades.*

The means for Title I and non-Title I third graders were not statistically, significantly different. Many of the students in non-Title I schools this year participated in the Title I program last year; they are enrolled in schools which are not eligible for Title I funds this year but were eligible last year. Hence, some of these students actually received supplementary services last year. If the Title I program was effective in the 1974-75 school year, then the current second and third grade students in non-Title I schools near the cutoff might be expected to display higher scores, as a group, than they would have had their schools not received Title I services in the previous school year. In other words, the treatment effects of the Title I program in the 1974-75 school year would continue to influence the achievement scores of those students who had been in the treatment group that year. This effect is sometimes called statistical contamination: the non-Title I second and third grade students in the current school year are not free from the influence of the previous year's Title I program.

Table 1
First Grade Prettest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=339) and Students in Non-Title I Schools (N=339) On the Prescriptive Reading Test Subtests

Subtest	Prettest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Auditory Perception						
$\bar{X}$	10.9	11.3	15.7	14.8	14.3	7.2
S.D.	3.2	3.0	3.4	3.7		
Visual Perception						
$\bar{X}$	10.5	10.6	13.0	12.7	12.6	3.1
S.D.	3.7	3.3	2.3	2.6		
Comprehension Development						
$\bar{X}$	17.7	18.3	21.8	21.2	20.6	4.6
S.D.	4.5	4.2	4.6	4.1		
Total Reading						
$\bar{X}$	39.0	40.0	50.5	48.7	47.7	5.4
S.D.	9.3	9.2	9.2	9.3		

In this Table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

*Seventh grade students were not included in the control group model analyses because the appropriate control schools were not designated in time to be included in the comparisons.

Table 2

First Grade Pretest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=405) And Students in Non-Title I Schools (N=874) On The Prescriptive Mathematics Test Subtests

Subtest	Prettest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Sets and Numbers						
$\bar{X}$	12.5	13.1	20.3	19.0	18.4	7.1
S.D.	5.0	5.5	5.3	5.4		
Numeration						
$\bar{X}$	5.4	5.8	9.1	8.8	8.4	5.5
S.D.	2.4	2.7	2.5	2.7		
Operations						
$\bar{X}$	8.7	9.4	16.9	15.8	14.8	7.8
S.D.	4.2	4.3	5.3	6.0		
Problem Solving						
$\bar{X}$	1.8	2.0	3.8	3.5	3.2	6.5
S.D.	1.3	1.3	1.7	1.7		
Measurement						
$\bar{X}$	7.2	7.3	9.5	9.4	9.3	1.7
S.D.	2.4	2.5	2.2	2.4		
Geometric Concepts						
$\bar{X}$	3.6	3.6	4.5	4.5	4.5	0.0
S.D.	1.3	1.4	1.3	1.4		
Total Mathematics						
$\bar{X}$	39.2	41.0	64.0	61.0	58.9	6.7
S.D.	12.7	14.3	15.1	16.2		

In this Table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

The no-treatment expectation is an estimate of the posttest score that Title I students would have attained had they not participated in the program. The no-treatment expectation was determined by adjusting the observed posttest for differences in pretest scores between students in Title I and non-title I schools. An examination of pretest differences between students in Title I and non-Title I schools revealed that this adjustment was important, because students in non-Title I schools consistently outperformed students in Title I schools on the pretest. A straight comparison of posttest scores for the two groups would be inappropriate, given that students in non-Title I schools had an initial advantage. Because it was expected that students in non-Title I schools would show an initial advantage, a principal axis adjustment rather than a covariance adjustment was employed (see Kenny, 1975).

On the average, first grade students in non-Title I schools have slightly higher pretests than first grade students in Title I schools. However, this pattern reverses on the posttest, with students in Title I schools showing higher Total Reading and Total Mathematics scores than students in non-Title I schools. This finding provides a strong argument for treatment effect at the first grade level. The results for the control group model do not indicate a Title I program impact at the second grade, although a moderate level of impact is found at the third grade. One plausible explanation for the absence of an effect at the upper grade levels is that some of these students benefited from the Title I program in previous grades. Given the strong effects at the second and third grade levels yielded by the norm-referenced model, it seems possible

that the control group model is failing to identify an effect because the control group is contaminated with last year's treatment. However, the finding of a moderate effect at third grade raises doubts about the possibility of a contaminated control group. Another explanation is that certain assumptions underlying the present application of the control model are faulty; thus, the model yields an inaccurate estimate of program effect.

Table 3

Second Grade Pretest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=467) and Students in Non-Title I Schools (N=293) on The Prescriptive Reading Test Subtests

Subtest	Pretest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Word Perception						
$\bar{X}$	28.1	27.7	34.1	33.7	34.0	0.3
S.D.	6.8	5.9	5.1	5.0		
Comprehension and Interpretation						
$\bar{X}$	30.4	27.9	42.4	41.4	43.8	-2.2
S.D.	9.8	10.2	10.0	9.3		
Study Reading						
$\bar{X}$	14.7	13.5	18.6	18.5	19.6	-3.5
S.D.	4.7	5.2	4.5	4.1		
Total Reading						
$\bar{X}$	73.2	69.0	95.1	93.5	97.3	- 2.0
S.D.	18.9	18.9	17.6	16.1		

In this Table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

Table 4

Second Grade Pretest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=482) and Students in Non-Title I Schools (N=287) on The Prescriptive Mathematics Test Subtests

Subtest	Pretest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Sets and Numbers						
$\bar{X}$	18.1	17.3	23.5	22.3	23.1	1.4
S.D.	4.8	5.0	4.8	4.3		
Numeration						
$\bar{X}$	5.6	5.3	7.8	7.6	7.9	-1.1
S.D.	2.0	1.8	2.3	2.2		
Operations						
$\bar{X}$	19.5	19.0	26.8	26.1	26.7	0.3
S.D.	5.9	6.0	6.7	7.4		
Problem Solving						
$\bar{X}$	3.2	3.1	4.6	4.6	4.7	-1.2
S.D.	1.6	1.6	1.4	1.5		
Measurement						
$\bar{X}$	7.0	6.8	9.7	9.1	9.3	2.6
S.D.	2.7	2.4	2.6	2.6		
Geometric Concepts						
$\bar{X}$	5.2	5.0	6.5	6.1	6.3	2.1
S.D.	1.8	1.7	1.7	1.6		
Total Mathematics						
$\bar{X}$	58.6	56.4	79.9	75.8	79.1	0.9
S.D.	14.7	13.8	15.8	15.7		

In this Table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

Table 5

Third Grade Pretest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=313) and Students in Non-Title I Schools (N=98) on The Prescriptive Reading Test Subtests

Subtest	Pretest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Word Perception						
$\bar{X}$	31.04	31.14	34.38	34.40	34.3	0.2
S.D.	6.60	5.72	5.85	5.45		
Comprehension and Interpretation						
$\bar{X}$	28.87	29.76	33.19	33.18	32.4	1.5
S.D.	8.10	7.37	6.81	7.10		
Study Reading						
$\bar{X}$	24.33	25.40	29.17	29.65	28.6	1.2
S.D.	6.50	6.19	6.49	6.34		
Total Reading						
$\bar{X}$	84.24	86.30	96.74	97.23	94.4	2.3
S.D.	18.65	17.09	16.74	16.98		

In this table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

Table 6

Third Grade Pretest, Posttest, and "No-Treatment" Posttest Expectation for Students in Title I Schools (N=326) and Students in Non-Title I Schools (N=109) on The Prescriptive Mathematics Test Subtests.

Subtest	Pretest Raw Score		Posttest Raw Score		No-Treatment Posttest Expectation Raw Score	Treatment Effect In Normal Curve Equivalents
	Title I	Non-Title I	Title I	Non-Title I		
Sets and Numbers						
$\bar{X}$	14.02	14.89	16.42	16.18	15.4	4.4
S.D.	3.58	3.23	3.21	3.03		
Numeration						
$\bar{X}$	7.27	8.09	9.47	10.02	9.2	1.2
S.D.	3.03	3.16	3.12	2.85		
Operations						
$\bar{X}$	22.04	24.27	31.73	32.78	20.7	1.9
S.D.	7.80	8.33	7.67	6.67		
Problem Solving						
$\bar{X}$	5.37	5.91	6.74	6.77	6.4	2.9
S.D.	2.41	2.38	1.82	1.80		
Measurement						
$\bar{X}$	13.86	14.69	16.44	16.76	16.0	1.6
S.D.	3.88	4.16	3.64	3.36		
Geometric Concepts						
$\bar{X}$	1.86	2.10	2.30	2.25	2.0	3.5
S.D.	1.12	1.13	1.20	1.15		
Total Mathematics						
$\bar{X}$	64.42	69.96	83.10	84.75	79.5	3.0
S.D.	16.58	17.92	16.55	14.44		

In this table "Non-Title I" refers to students in non-Title I schools, whereas elsewhere the term refers to ineligible students within eligible schools.

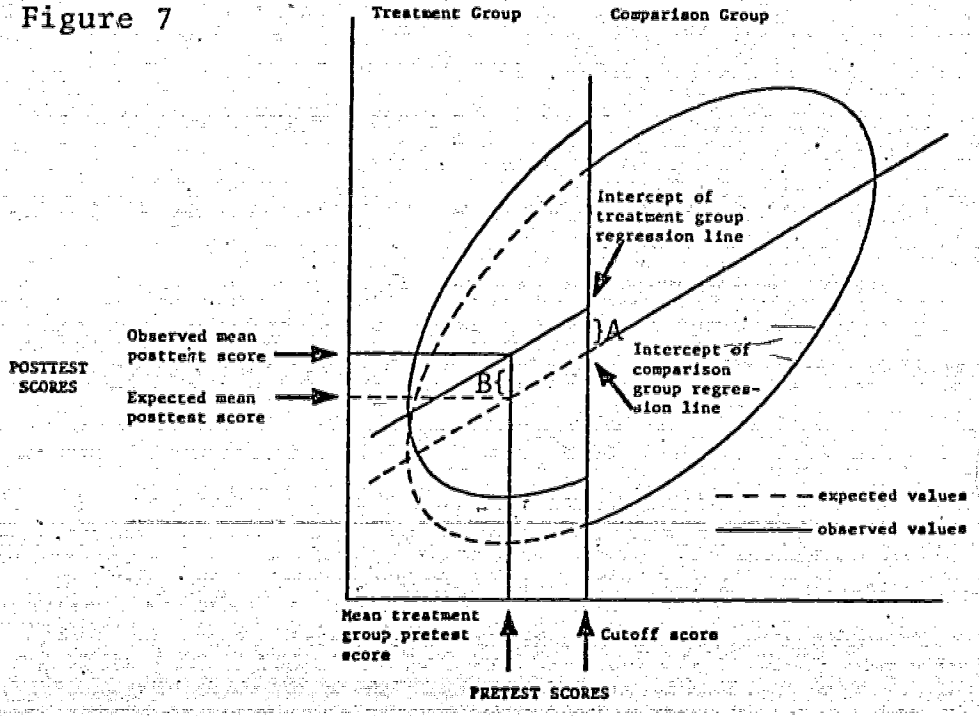
The first grade and third grade results from the control group model corroborate the findings of the norm-referenced model and confirm that the Title I program is having a statistically and educationally significant impact on student reading and mathematics achievement. The fact that two models using different achievement tests converged on a similar estimate of treatment effect strongly indicates that the estimate is valid. The fact that the two models do not converge on a similar estimate of treatment effect at the second grade, in light of the findings at the other two grades, is best considered a sampling anomaly. Replication of this analysis next year should afford additional insight into these second grade results.

Special Regression Model

This statistical design, as its name implies, is based on regression methodology. As with the two previously discussed models, the method which is used to derive the no-treatment expectation determines the model which is actually used. In this case, the treatment and comparison groups are selected on the basis of their pretest scores using a fi mly established, strictly enforced cutoff. The treatment group is then given the benefits of the Title I program. Both treatment and comparison groups are posttested using an instrument which correlates highly with the pretest device. The observed post treatment performance is actually the treatment group's average posttest score. The no-treatment expectation is derived from a projection of the regression line determined by the comparison group's pre- and posttest scores. The actual treatment effect is measured at two points, as indicated in Figure 7:

A. At the pretest cutoff score.
B. At the treatment group's average pretest score.

Figure 7



The purpose of measuring the treatment effect at these two points along the projected regression line of the comparison group is to determine whether the relationship between pre- and posttest for both treatment and comparison groups was the same.

There are five assumptions of the model which should be met if this model is to yield an unbiased estimate of program impact: (1) the pretest and the selection test must be the same instrument; (2) the pretest or posttest should be given at the empirical normative data point of the instrument; (3) the pretest and posttest should be highly correlated ($r \geq 0.60$); (4) there should be a strict cutoff score for determination of placement in the treatment group and the comparison group; (5) only those students having both pretest and posttest scores should be used in the analyses.

The model is predicated on the supposition that the no-treatment expectation of the treatment group can be calculated using the regression line determined from the comparison group. A strict cutoff score for determining placement in each of the groups is necessary. Otherwise, the pretest standard deviation is inflated and the regression line of the comparison group is flattened. This will systematically bias the results against showing a positive treatment effect for Title I programs.

This model was rejected for implemantation in the evaluation after determining that one of these requirements was violated by the data, a violation which would render the resulting measured treatment effects questionable if not totally misleading. The evaluators found that a clearly defined cutoff was not enforced in every case. Primarily for this reason, the special regression model was not implemented using the District of Columbia data. A poorly implemented evaluation model can be not only confusing but totally misleading and should be avoided for this reason.

Standardized Growth Expectations: Some Findings and Implications

Most educational evaluations, including the present one, ignore what may be a critical factor when estimating whether a program has had an impact or whether students have learned more at one grade level than another. It is typically assumed that a treatment effect of seven NCEs (one-third standard deviation) has the same meaning in first grade as in seventh grade. What is not typically considered is that the expected growth is different in first grade and seventh grade. Another way of viewing the issue is to ask whether a student would lose the same amount (relative to national norms) in reading achievement if he/she fell asleep for all of first grade or all of seventh grade. This is the same as asking, how much growth does the average student make in reading achievement during first grade, and is it the same as the growth realized by the average seventh grade student during one school year?

An answer to the above question can be approximated by assuming that a student will attain the same raw score on the pretest and posttest if no learning has taken place. If the pretest raw score is equivalent to a national percentile of 50 and the same raw score is entered into the posttest percentile table, the resulting percentile score will be less than 50. The difference between the pretest percentile and the posttest percentile expressed in standard score form is the standardized growth expectation (SGE). The SGE is the amount that a student learns over a period of time or, conversely, what the student would lose if he/she fell asleep and learned nothing. An example may help to clarify the computation procedures used to calculate SGEs. Table 7 presents a raw score to percentile conversion for beginning of first grade and end of first grade on the Total Reading scale of the CTBS/S. The average (50th percentile) beginning first grade student attains a raw score of 31 on Total Reading. Under the assumption that this average student learns nothing in the first grade, he/she would be expected to obtain again a raw score of 31 on the posttest. Whereas a raw score of 31 is equivalent to a beginning first grade percentile of 50, it represents an end of first grade percentile of 9. If both percentiles are converted to NCEs (50 → 50; 9 → 21.8) and subtracted, the result is an SGE of 29.2. In other words, if an average student falls asleep and learns nothing during the first grade, he/she would be expected to lose 29.2 NCEs because that is the amount of standardized growth exhibited by the national norm group during the first grade. Yet another way of viewing the SGE is to consider it as an estimate of the effect of school, home, and social forces (such as radio and television) on first grade students' reading achievement.*

*The SGE differs slightly depending upon where in the pretest distribution the raw score is selected to be entered into the posttest percentile distribution. For ease of presentation, this difference is ignored since it does not influence the general conclusions.

Table 7

Raw Score to Percentile Table for Beginning and End of First Grade on CTBS/S, Level B

Total Reading

Beginning of First Grade		End of First Grade	
Raw Score	Percentile	Raw Score	Percentile
73-84	99	84	99
86-72	98	84	98
65-67	97	84	97
61-64	96	84	96
59-60	95	84	95
57-58	94	83	94
55-56	93	83	93
53-54	92	82	92
52	91	82	91
31	50	59	50
31	49	58	49
31	48	58	48
31	47	57	47
31	46	56	46
30	45	55	45
30	44	54	44
30	43	53	43
30	42	53	42
30	41	52	41
20	10	32	10
19	9	31	9
18	8	30	8
18	7	29	7
18	6	28	6
18	5	27	5
17	4	25-26	4
16	3	24	3
15	2	21-23	2
0-14	1	0-23	1

Table 8 presents SGEs in reading and mathematics for grades 1,2,3, and 7. To facilitate comparison of information from the findings of this section with the rest of the chapter, all SGEs are presented as normal curve equivalents (X=50; S.D.=21.06). All SGEs in Table 8 are computed from the norms in the publisher's manual for the CTBS/S and CTBS/T (for seventh grade). The procedure used to compute SGEs is identical to that described in the paragraph above.

Table 8

Standardized Growth Expectations (SGEs) Expressed as NCEs in Total Reading and Total Mathematics for Grades 1, 2, 3, and 7

	Grade 1	Grade 2	Grade 3	Grade 7
Total Reading	29.2	9.3	8.7	3.7
Total Mathematics	21.8	12.3	11.7	3.2

The period of growth in each case is fall to spring of the school year. The 50th percentile point was used for entry into the pretest tables.

Table 8 reveals that the SGE for second grade Total Reading is only one-third of the SGE at the first grade level. Similarly, the second grade SGE for Total Mathematics is about four times larger than the SGE for grade seven. This seems to indicate that the rate of growth is different from grade to grade and, in particular, the rate slows with each additional year of schooling. The SGEs for height and weight computed from birth to eighteen years of age follow a similar pattern of deceleration. The largest SGEs appear during the first few years of life and gradually diminish until eighteen years of age when the SGE is less than one NCE point. Table 9 illustrates this phenomenon for height, viewing ages four to eighteen.

Table 9

Means, Standard Deviations, and SGEs for Weight Expressed in Pounds at Various Age Levels

AGE (In Years)	MEAN Weight Pounds	STANDARD DEVIATION	SGE
4	35.7	4.2	19.6
5	40.3	4.9	17.1
6	45.3	5.9	16.7
7	50.8	6.8	13.5
8	56.6	8.2	14.2
9	62.6	10.1	13.5
10	68.9	12.3	13.5
11	75.9	16.2	15.6
12	86.2	20.0	14.9
13	99.2	22.5	14.9
14	113.2	24.2	13.5
15	126.5	25.1	11.7
16	137.7	25.2	8.1
17	145.1	24.6	2.1
18	147.2	23.9	-

Tables 10 and 11 further depict the differential growth rate across development in the cognitive domain. The various inconsistencies which exist across several tests in measuring the same cognitive construct is illustrated in Table 10. Notice the disparate rates of decline on each scale for the SGE using the ITBS and the CTBS. Also, consider the different SGEs on the CTBS alone, depending on whether national or big city norms tables are used. Apparently, something happens to the normative populaton* at the end of third grade which causes the mean achievement scores for each percentile group during the successive year to increase, thereby inflating the SGE. This is probably caused by an unusual proportion of the lower achieving students somehow being excluded from the normative population after third grade: when a large percentage of low scorers drop out of the norms, then the average score at all percentile levels increases. This occurs again at the end of grade nine and is most likely due to the large number of low scoring dropouts exiting the educational system.

This "dropout effect" discussed above does not seem to occur in the area of cognitive ability at either grade four or ten. Rather, the SGE steadily declines, year by year, for the verbal area. Quantitative ability behaves similarly, but the SGE hops up noticeably in the eleventh grade. This may also be a reflection of the "dropout effect" as evidenced with the achievement instruments. Although nonverbal IQ for one test (IPAT) shows a steady decline, its evanescence on the other test is not so consistent. This discrepancy is due either to the differences in item content between the two tests, cohort effects, or differences in norming samples.

Table 11 displays the decline in SGEs for ability as measured by the Wechsler Adult Intelligence Scales (WAIS) from ages sixteen to seventy-four. Notice that the standard deviation for the norm group on all three measures of ability (verbal, performance, full scale) remains relatively stable across this vast age range. Also, it is easy to recognize the effects of aging and senility upon the tested ability level of the senior citizens group (sixty-five and older): the SGE loss suddenly doubles (triples in the case of verbal ability) after age sixty-four. Interestingly, the SGE loss for performance ability is nearly equally substantial between ages twenty-five and fifty-four.

Table 12 introduces two new variables into the investigation of developmental SGE decline. If the amount of growth is calculated separately, both for different percentile ranks and for the regular school year and summer, yet another type of differential growth is revealed. The Metropolitan was empirically normed in both the fall and spring, thus permitting a comparison between achievement during the summer and that during the regular school year. Summer achievement gains approximate or surpass those made during the regular school year for the upper percentile students. The fact that 25th percentile students seem to keep pace during the school year but lag behind during the summer has some interesting implications for compensatory education programs.

*The possibility of cohort effects must also be considered (Baltes and Schaie, 1973).

Table 10
SGEs for Several Achievement/Ability Subtests
Expressed as Normal Curve Equivalents ($\bar{X}$ = 50; S.D. = 21.06)

Ages at Successive Grade Levels

	Age:	7	8	9	10	11	12	13	14	15	16	17	18
	Grade:	1	2	3	4	5	6	7	8	9	10	11	12
Standardized Achievement Tests													
Vocabulary													
ITBS			18.5	9.3	12.9	12.3	11.7	9.7	6.4				
CTBS: National Norms			21.8	15.6	10.1	10.4	7.0	6.7	6.4	4.8	5.3	5.9	5.3
CTBS: Big City Norms				(14.9)	(7.0)	(5.3)	(6.4)	(5.9)	(1.1)				
Reading Comp.													
ITBS			20.1	9.9	13.5	13.5	12.3	8.7	5.3	3.7	3.7	3.2	3.2
CTBS: National Norms		29.6	22.2	14.9	9.9	8.1	4.8	5.3	4.8	3.7	7.0	5.3	3.7
CTBS: Big City Norms				(13.5)	(6.4)	(4.8)	(5.3)	(4.8)	(1.4)				
Total Language													
ITBS			27.0	11.0	15.6	11.7	10.4	8.1	5.0				
CTBS: National Norms		15.6	24.7	14.2	8.7	8.7	4.2	4.8	4.8	1.4	7.5	4.8	4.2
CTBS: Big City Norms				(14.9)	(6.4)	(5.3)	(5.3)	(5.9)	(1.6)				
Total Math													
ITBS			24.7	13.5	19.3	17.0	14.3	11.4	8.1	2.6	2.6	1.1	2.1
CTBS: National Norms		27.0	24.7	20.9	11.0	12.9	7.5	6.4	6.4	6.4	7.0	3.5	2.6
CTBS: Big City Norms				(13.5)	(6.4)	(7.5)	(7.5)	(5.3)	(2.6)				
Cognitive Abilities Test													
Verbal				19.3	13.5	12.9	10.4	7.5	7.0	7.0	6.4	4.8	
Quantitative				17.3	13.5	10.4	8.9	7.0	7.5	6.4	3.7	10.4	
Nonverbal				9.9	8.7	7.0	4.2	5.3	3.7	5.9	4.8	3.2	
IPAT Culture Fair Test													
(Nonverbal IQ)			8.4	11.7	9.5	6.8	6.8	6.8					

Table 11

SGEs Expressed in Standard Deviation Units (Z Scores) for the Wechsler Adult Intelligence Scales (WAIS)

AGE GROUP	VERBAL IQ		PERFORMANCE IQ		FULL SCALE IQ	
	SGE	S.D of Norm Group	SGE	S.D. of Norm Group	SGE	S.D. of Norm Group
16-17	+0.13	13.9	+0.07	11.3	+0.13	25.2
18-19	+0.13	14.9	+0.07	11.8	+0.13	25.7
20-24	+0.13	15.2	-0.13	12.0	+ .00	24.8
25-34	-0.07	14.6	-0.30	11.8	-0.13	24.8
35-44	-0.13	14.9	-0.40	11.3	-0.26	26.2
45-54	-0.13	16.2	-0.33	11.3	-0.26	25.8
55-64	-0.13	16.4	-0.20	10.8	-0.20	
65-69	-0.40		-0.45		-0.40	
70-74	-0.26		-0.45		-0.40	

Table 12

SGEs Expressed as NCEs ($\bar{x}$=50; SD=21.06) for Metropolitan Achievement Test (MAT), Total Reading and Total Mathematics for the 25th, 50th, and 75th Percentiles

GRADE LEVEL	READING			MATHEMATICS		
	25th	50th	75th	25th	50th	75th
Summer	11.6	11.0	13.1	9.5	8.7	10.0
2	16.9	17.7	10.5	9.6	16.3	8.3
Summer	5.9	9.3	10.0	4.3	9.0	8.7
3	6.7	7.5	10.0	6.7	14.0	10.0
Summer	5.1	6.4	8.9	5.9	5.3	5.5
4	7.6	6.4	6.7	11.6	12.3	13.7
Summer	4.3	5.3	4.9	5.9	1.1	9.4
5	7.6	6.4	5.5	9.5	7.5	9.4
Summer	2.8	3.7	4.3	4.3	6.4	6.1
6	5.9	5.3	4.9	5.9	6.4	7.2
Summer	1.8	5.3	4.9	4.3	7.5	3.8
7	-	2.1	3.8	0.7	3.2	3.8
Summer	6.3	7.5	6.1	5.9	7.0	10.0
8	-	1.1	1.1	0.0	1.1	0.7
Summer	-	2.5	2.5	0.0	1.1	1.3

As pointed out earlier, Title I evaluations assume that an NCE gain, or treatment effect, of seven points means the same thing if it occurs in the first grade or the seventh grade. The assumption is that it is just as difficult to improve a first grade treatment group by seven NCEs as it is to improve a seventh grade treatment group by seven NCEs. An examination of Table 8 suggests that a gain of seven NCEs in Total Mathematics at the seventh grade level represents a 200 percent increase in achievement rate, whereas the same seven point gain at the first grade represents a 33 percent increase in achievement rate. The question arises as to whether achievement rate expressed in SGEs must be considered in interpreting a treatment effect. If all the impact of school, community, home, and social forces can only cause a Total Reading SGE of 3.7 NCEs for the average seventh grader nationally, then is it fair to expect a Title I program to show a treatment effect of seven NCEs above and beyond the SGE of 3.7? Perhaps the ratio of treatment effect to SGE would provide a more comparable index across grades, tests, and subtests. When the SGE is considered, a number of difficult questions arise regarding the meaning of treatment effect and the wisdom of aggregating across either grades or tests. At the present time not enough is known to judge the value of the SGE as a statistic for communicating treatment effects in a comparable unit. A special report is forthcoming that will present this concept in fuller detail and, hopefully, discuss the contributions, if any, that the SGE promises to make to evaluation methodology.

Summary

Table 13 summarizes the treatment effect in Total Reading and Total Mathematics as documented by the norm-referenced and control group models. It is interesting to note that the two models yield similar estimates of treatment effect for first grade reading and mathematics achievement, but substantial differences in treatment effect are evident at second and third grade. The fact that all assumptions of both the norm-referenced and control group models were met at the first grade level generated confidence in the accuracy of these estimates of treatment effect. The widely divergent estimates for second grade and the moderately similar estimates at third grade suggest that both models may be highly sensitive to the types of assumption violations which are encountered in typical applications of these two models. It seems plausible that when low achievers in a class are removed for one to two hours per day of special laboratory instruction, the remaining group can progress at a faster pace. This is because the student/teacher ratio is lowered and because the group is academically more homogeneous. Also, the likelihood is that equipment, materials, and teacher inservice (bought by Title I funds) benefit all students in the class. It becomes

Table 13

Summary of Treatment Effects (in NCEs) at Grades 1, 2, and 3 for Total Reading and Total Mathematics

	Total Reading			Total Mathematics		
GRADE	1	2	3	1	2	3
Norm Referenced Model Estimates of Treatment Effect	3.9	9.9	3.1	6.2	8.1	7.9
Control Group Model Estimates of Treatment Effect	5.5	-2.0	2.3	6.7	0.9	3.0
Average Effect	4.7	6.0	2.6	6.5	4.5	5.5

clear that non-Title students are receiving a "treatment" by virtue of the fact that Title I students obtain project services. Although the full magnitude of this unexpected, positive outcome is not yet known, this year's findings confirm that non-Title students are achieving much better than would be expected from their pretest scores. A fuller treatment of the results supporting this finding will be forthcoming in a special report. In addition, the issue of differential growth as measured by SGEs across various grade levels definitely merits serious study.

One puzzlement continues and should be investigated further. Why do Title I students do so well in the D. C. Title I program and then lose a major portion of their newfound advantage over the summer? This phenomenon is nationwide and should not be considered an anomaly of the D. C. program. However, its widespread appearance does not reduce local responsibility for finding an explanation.

BIBLIOGRAPHY

Baltes, Paul B. and K. Warner Schaie. Life Span Developmental Psychology: Personality and Socialization. New York: Academic Press, 1973.

Bessey, B. L., L. D. Rosen, A. Chiang, and G. K. Tallmadge. Further Documentation of State ESEA Title I Reporting Models and Their Technical Assistance Requirements: Phase I Report. Mountain View, California: RMC Research Corporation, 1976.

Chiang, Alice and Larry D. Rosen. "The Normal Curve Equivalent: A Metric to Measure Instructional Project Impact." Paper written at RMC Research Corporation, Mountain View, California, 1975.

Kenny, David A. "A Quasi-Experimental Approach to Assessing Treatment Effects in The Non-Equivalent Control Group Design," Psychological Bulletin, 1975, V. 82, No. 3, pp. 345-362.

Tallmadge, G. K. and C. T. Wood. Users Guide-ESEA Title I Evaluation and Reporting System. Washington, D. C.: Department of Health, Education, and Welfare, USOE, OPBE, 1976.