A YEAR OF CONSEQUENCES: A LONGITUDINAL STUDY OF DELAWARE'S STUDENT ACCOUNTABILITY PLAN

JANUARY 2003

AUDREY J. NOBLE, PH.D.
DIRECTOR

LISA A. BANICKY, PH.D.
EDUCATIONAL RESEARCHER FOR POLICY ANALYSIS

MICHELE KREISMAN, PH.D.

Delaware Education Research & Development Center
University of Delaware
Newark, DE 19716

PUBLICATION T03.01.1

ACKNOWLEDGEMENTS

The author extends a special thanks to the following individuals who provided data and assistance necessary to complete this study:

Terry Anderson
Ann Case
Shirley Dear
Pamela Brown Stazesky
Tommy Tao
Robin Taylor
Ximena Uribe-Zarain

This research was made possible through the support of the Delaware State Board of Education.

TABLE OF CONTENTS

LIST OF TABLES	4
EXECUTIVE SUMMARY	6
INTRODUCTION	9
SECTION 1: STUDENT ACHIEVEMENT	10
SECTION 2: BEHAVIORAL EFFECTS	24
SECTION 3: RETENTION	29
SECTION 4: ATTRITION	37
POLICY CONSIDERATIONS	39
APPENDIX A: PSYCHOMETRIC ISSUES	40

LIST OF TABLES

Table 1c1: Secondary Cohort: All Students 16

Table 1c2: Secondary Cohort: All Female Students 17

Table 1c3: Secondary Cohort: All Male Students 17

Table 1c4: Secondary Cohort: African American Students 18

Table 1c4: Secondary Cohort: Asian Students 18

Table 1c6: Secondary Cohort: Caucasian Students 19

Table 1c7: Secondary Cohort: Hispanic Students 19

Table 1c8: Secondary Cohort: Special Education Students 20

Table 1c9: Secondary Cohort: Students from Low Income Families 20

Table 1c10: Secondary Cohort: Students enrolled in Title I Programs 21

Table 1c11: Overall Changes in Performance of All Students 21

Table 1c12: Performance of Students by Gender 22

Table 1c13: Performance of Students by Race 22

Table 1c14: Performance of Students from Low-Income Families 23

Table 1c15: Performance of Students from Title I Programs 23

Table 2a1: Elementary Cohort: Out-of-School Suspensions by Gender, 2000-01 and 2001-02 24

Table 2a2: Elementary Cohort: Out-of-School Suspensions by Race, 2000-01 and 2001-02 25

Table 2a3: Elementary Cohort: Out-of-School Suspensions by Income Status, 2000-01 and 2001-02 25

Table 2a4: Secondary Cohort: Out-of-School Suspensions by Gender, 2000-01 and 2001-02 26

Table 2a5: Secondary Cohort: Out-of-School Suspensions by Race, 2000-01 and 2001-02 26

Table 2a6: Secondary Cohort: Out-of-School Suspensions by Income Status, 2000-01 and 2001-02 27

Table 2b1: DSTP Performance of Secondary Students Suspended in 2001-02 28

Table 3a1: Gender of Retained Students in Elementary Cohort (grade 6) 29

Table 3a2: Race of Retained Students in Elementary Cohort (grade 6) 30

Table 3a3: Income Status of Retained Students in Elementary Cohort (grade 6) 30

Table 3a4: Retained Students in Secondary Cohort by Gender (grade 9) 30

Table 3a5: Retained Students in Secondary Cohort by Race (grade 9) 31

Table 3a6: Retained Students in Secondary Cohort by Income Status (grade 9) 31

Table 3a7: Retained Kindergarten Students by Race from 1992 – 2002 32

Table 3a8: Retained 1st Grade Students by Race from 1992 –2002 32

Table 3a9: Retained 2nd Grade Students by Race from 1992 –2002 33

Table 3a10: Retained 3rd Grade Students by Race from 1992 –2002 33

Table 3a11: Retained 4th Grade Students by Race from 1992 –2002 33

Table 3a12: Retained 5th Grade Students by Race from 1992 –2002 34

Table 3a1:. Retained 6th Grade Students by Race from 1992 –2002 34

Table 3a14: Retained 7th Grade Students by Race from 1992 –2002 34

Table 3a15: Retained 8th Grade Students by Race from 1992 –2002 35

Table 4.1: Demographic Characteristics of Students Identified as "Missing" within the Elementary Cohort 37

Table 4.2: Demographic Characteristics of Students Identified as "Missing" within the Secondary Cohort 37

EXECUTIVE SUMMARY

This ongoing study, conducted at the request of the Delaware State Board of Education, is intended to monitor Delaware's student accountability plan and provide longitudinal information about its effects on students. It addresses four key areas of potential impact:
1) student achievement, 2) behavioral effects, 3) retention rates, and, 4) completion rates. Highlights are organized according to these categories.

STUDENT ACHIEVEMENT

- Findings for the secondary cohort (students in grade 8 in 2000 and grade 10 in 2002):
  - Overall, the results revealed a statistically significant decrease for SAT9 reading scores, with a medium effect size;[1]
  - Female students showed statistically significant decrease in reading scores, with a medium effect size;
  - African American students demonstrated statistically significant decreases in reading, with a medium effect size;
  - Caucasian students showed a statistically significant decrease with a medium effect size for reading;
  - Hispanic students demonstrated a statistically significant decrease in reading scores, with a medium effect size;
  - Between 2000 and 2002, the percentages of students scoring "below the standard" (performance levels 1 and 2) remained relatively stable across all three content areas. The overall percentage of students scoring "above the standard" (performance levels 4 and 5) decreased in all three content areas between 2000 and 2002.
  - With the exception of Asian students' math performance, between 2000 and 2002, there is a decrease in the percentage of students in all racial groupings scoring "above the standard" in all 3 content areas. There is an increase in the percentage of African-American, Caucasian, and Hispanic students scoring "below the standard". This decline is particularly pronounced in reading among Hispanic students.
  - Title 1 students were more likely to improve by at least one performance level in reading and writing between 2000 and 2002.

[1]Only statistically significant results revealing medium or large effect sizes should be explored further because the large numbers of students in these analyses increase the possibility that even trivial differences between scores will appear to be statistically significant.

BEHAVIORAL EFFECTS

- More than half of the students who had been suspended during the 2001-2002 school year had been retained at least once.

- For the elementary cohort

    - Gender, minority status, and income status were statistically significant predictors of suspensions given during the 2001-2002 school year
        - Males were more likely to be suspended than females.
        - The odds of being suspended were:
            - 2.39 times greater for African American and Hispanic students; and
            - 2.64 times greater for low income students.

- For the secondary cohort
    - Gender, minority status, and income status were statistically significant predictors of suspensions given during the 2001-2002 school year.
        - Males were more likely to be suspended than females.
        - The odds of being suspended were:
            - 2.39 times greater for African American and Hispanic students; and
            - 1.89 times greater for low income students.

RETENTION

- During the 2000-2001 and 2001-2002 school years, most of the retained students in the elementary cohort were male, African American, and from low income families.

- The largest number of students retained within the secondary cohort was male and African American.

- Over the past ten years, African American and Hispanic students have been retained at rates that are consistently higher than their Caucasian and Asian peers;

- In 2002, the first year that the state's student accountability plan went into full effect, there was a major increase in the number of 8th grade students retained. This trend was even more notable among African American and Hispanic students; and,

- During the past 2 years, 2001 and 2002, there appears to be an increase in the incidence of kindergarten retentions across all demographic groups.

- Placement into special education does NOT appear to be a prevalent practice in response to retention in grade.

- Retention was a significant predictor of suspension. and that the odds of being suspended were:

    - At the elementary level, students who had been retained were *2.09 times more likely to* receive out-of-school suspension; and,

    - The odds of being suspended were *6.33 times greater* for students in the secondary cohort who had been retained than students in the secondary cohort who had been promoted.

COMPLETION RATES

Since neither of the cohorts has progressed to the age when they are able to dropout of school, this year's analysis explored the characteristics of students who are labeled as "missing". These students are not officially classified as dropouts yet they are not be accounted for in the state's accountability system.

- As of September 30, 2001, 15% of each of the two cohorts studied have been labeled as "missing". (i.e., 1233 students within the elementary cohort and 1336 from the secondary cohort)

- As compared to the portion of the state's population that each group represents, there appears to be over-representation of:

    - Hispanic students within the elementary cohort; and,

    - Students classified as Limited English Proficient (LEP), students from low-income families, and Hispanic students within the secondary cohort.

INTRODUCTION

This report is the third in a series of annual reports designed to monitor the impact of Delaware's Student Accountability Plan on students in Delaware. More than ten years of educational reform in Delaware has led to the creation of a performance based accountability system composed of rigorous standards, a statewide assessment, and consequences for performance.

With increased accountability for performance comes the responsibility of monitoring systems for both intended and unintended consequences. According to the National Research Council: *"high stakes testing programs [should] routinely include a well-designed evaluation component. Policymakers should monitor both the intended and unintended consequences of high-stakes assessment on all students and on significant sub-groups of students including minorities, English language learners, and students with disabilities"* (p. 281).[1]

This ongoing study, conducted at the request of the Delaware State Board of Education, is intended to monitor Delaware's student accountability plan and provide ongoing information about its effects on students. Based on previous research in the area of high stakes testing, several key issues were identified during the planning of the study as areas for monitoring. These issues included:

- Student Achievement;
- Behavioral Effects;
- Retention Rates; and,
- Completion Rates.

Study Design

To address these issues, students enrolled in 3rd or 6th grade during the 1997-98 school year were selected as cohorts for study. New to this year's report was an examination of student retention data collected over the past decade for grades kindergarten through eight. Both of these approaches represent a longitudinal approach to examining student outcomes with the first focusing on the same students over time, and the second, examining the effects of a single policy with different groups of students over a ten-year period. Questions and issues associated with the findings appear at the end of the report in the section entitled "Policy Considerations".

SECTION 1: STUDENT ACHIEVEMENT

The primary goal of the student accountability system is to improve the academic achievement of all Delaware students. This year's longitudinal study focuses on the performance of the secondary cohort only. At the time of the spring 2002 administration of the DSTP, the students in the secondary cohort were in the 10th grade; the elementary cohort would have been in 7th grade in 2002. The analyses that were conducted to produce the following are similar to those performed in 2000 for the elementary cohort (see *Putting the 'Student' Back in Student Accountability*. November 2000).

<u>A Word of Caution About Statistical Interpretations</u>

The goal of many statistical analyses is to show that there is some difference between sets of observations, and that the difference is due to something other than chance factors. For example, when examining the elementary cohort's average SAT9 math score from 1998 (Mean=53.88) to their average SAT9 math score in 2000 (Mean=57.35) there is an increase of 3.43. Finding such a difference does not necessarily mean that it is a meaningful difference. This difference may simply reflect the amount of variability in the data. Statistical analyses are set up in such a way as to compare the difference found, in this case 3.43, to a measure of how much of a difference is expected simply due to chance factors. The extent to which the observed difference (3.43) is over and above the amount of difference expected simply due to chance determines whether or not a result is statistically significant. Therefore a statistically significant result simply means that an outcome, in this example a difference of 3.43, is unlikely to be due to chance factors and instead may represent an actual improvement in scores.

Recently many researchers have argued that significance tests can be misleading because with very large sample sizes, even the smallest difference between two sets of observations can result in a significant finding. Therefore, testing for statistical significance is often viewed as the first step in data analysis with the second step focused on the size of the "effect".[9]

To use an analogy, testing for statistical significance is like using a magnifying glass to locate an object. The size of the sample determines the "magnification" of the lens. Consequently, larger samples result in even the smallest difference appearing quite large. Effect size can be thought of as a ruler that researchers use to measure the size of their findings. In the case of the magnifying glass analogy, we may locate an object that appears to be quite large, but when the ruler (i.e. effect size) is placed next to it under the magnifying glass the size of the object is placed in a more meaningful context.

Conventional rules of thumb indicate that an effect size of .2 is small, an effect size of .5 is medium, and an effect size of .8 is large.[10] For the purposes of this study, a statistically significant finding with an effect size of less than .20 is not considered to represent a real difference or change in scores.

With these caveats in mind, the analyses reported in the following sections provide information as to the statistical significance of the findings as well as the size of the effect found for each analysis.

Question 1a: Are students' test scores improving over time?

Method

An examination of improvements over time was conducted through use of paired samples t-tests. In this analysis, students' scores from the spring of 2000 (when they were in 8th grade) were compared to their scores in the spring of 2002 (when they were in 10th grade). This analysis required that a student have a score at both points in time in order to be included.

When conducting statistical analyses researchers select a probability value indicating how unlikely an outcome needs to be to consider it as resulting from something other than chance factors. In most cases, .05 is chosen as the probability value. If an outcome of an analysis has a probability of occurring that is less than this value then it is considered statistically significant and the researcher can be 95% confident that the outcome reflects a true difference and not simply chance factors.

When multiple analyses are conducted on the same set of data, in this case multiple dependent t-tests, adjustments to the probability level must be made. This adjustment is necessary to prevent the researcher from capitalizing on chance factors. It works to limit the likelihood that a researcher will falsely conclude that a difference is statistically significant. One method of adjusting the probability value is to divide it by the number of planned comparisons. For example, when looking at the relationship of SAT9 Math scores to Gender, one dependent t-test was conducted for males and one for females. In this case the probability value (.05) was divided by two resulting in a probability value of .025. In this case, an outcome had to have a probability less than .025 to be considered statistically significant. Similar adjustments were made for each of the dependent t-tests reported.[11]

Results

OVERALL PERFORMANCE CHANGES OVER TIME

- When examining the secondary cohort overall, the results revealed a statistically significant decrease for SAT9 reading scores, with a medium effect size (d=.54).

- Although there was a statistically significant decrease in mathematics as well for this cohort, the effect size was negligible (d=.07).

- There was no statistically significant difference in writing scores.

GENDER COMPARISONS

- Both male and female students in the secondary cohort showed statistically significant decreases in reading scores. The male analysis demonstrated a small effect size and the female analysis showed a medium effect size (d=.42 and d=.67, respectively).

- Although there were statistically significant decreases in mathematics scores for secondary males and females, the effect sizes were negligible (d=.06 and d=.08, respectively).

- In addition, there was a statistically significant decrease in writing scores for females, however, this analysis also exhibited a negligible effect size (d=.08). There was no statistically significant difference in writing scores for males.

RACE COMPARISONS

- African American Students
    - Regarding race[2], African American students demonstrated statistically significant decreases in reading and mathematics, however, the reading analysis demonstrated a medium effect size (d=.52) and the mathematics analysis showed a negligible effect size (d=.12).
    - There was no statistically significant difference in writing scores for African American students.

- Asian Students
    - There was a statistically significant decrease in reading scores, with a small effect size (d=.47).
    - There were no statistically significant differences in mathematics or writing scores for Asian students between 2000 and 2002.

- Caucasian Students
    - There were statistically significant decreases in all three subjects:
        - with a medium effect size for reading (d=.55),
        - negligible effect sizes for mathematics (d=.06) and writing (d=.04).

[2] The number of American Indian students was too small to be included in this analysis.

- Hispanic Students
  - There was a statistically significant decrease in reading scores, with a medium effect size (d=.50).
  - There were no statistically significant differences in mathematics or writing scores for Hispanic students.

SPECIAL EDUCATION COMPARISONS

Regarding students classified into specific programs[3], special education students evidenced:

- A statistically significant decrease in reading scores with a small effect size (d=.27).
- There were no statistically significant differences in mathematics or writing scores for special education students between the 2000 and 2002 test administrations.

INCOME LEVEL COMPARISONS

Students from low-income[4] families demonstrated:

- A statistically significant decrease in reading with a small effect size (d=.49);
- A statistically significant decrease in mathematics with an negligible effect size (d=.08); and,
- No statistically significant difference in writing scores between the 2000 and 2002 test administrations

TITLE I COMPARISONS

Between 2000 and 2002, students enrolled in Title I programs showed:

- A statistically significant decrease in reading with a small effect size (d=.35);
- There was no statistically significant difference in mathematics scores; and,
- A statistically significant increase in writing scores; however, the effect size was negligible (d=.17).

[3] The number of Limited English Proficient students was too small to be included in this analysis.
[4] Low income status is based on students' application for free/reduced lunch programs.

SUMMARY

The results of the paired samples t-tests revealed that all student groups analyzed showed statistically significant decreases in SAT9 reading scores from 2000 to 2002, with small to medium effect sizes. The remaining mathematics and writing analyses evidenced negligible effect sizes, even when statistically significant differences were exhibited. Only statistically significant results revealing medium or large effect sizes should be explored further because the large numbers of students in these analyses increase the possibility that even trivial differences between scores will appear to be statistically significant.

Question 1b: Do test score gains differ as a function of gender, race, or income status?

The previous question, 1a, compared the performance differences between 2000 and 2002 *within* each student group. Previously, test score gains were examined as a function of gender, race, income status, special education status, and Title I classifications. However, several of these categories have too few students in them to allow for a reasonable test of the question under consideration. Therefore, test score gains were examined as a function of gender, low-income status, and race. For the purpose of this analysis, only African Americans and Caucasians were included in the analysis because of the small number of students represented within the other race categories.

Method

Test score gains were examined by first computing a difference score for each student by subtracting the student's score in 2000 from their score in 2002. Three separate 2 (gender: male vs. female) X 2(race: African American vs. Caucasian) X 2 (income status: low income vs. not low income) factorial analyses of variance (ANOVA) were conducted. The gender, race, and income status of the student were used as categories for comparing change scores. This analysis explored differences between low income and non-low income students, between males and females, and between African Americans and Caucasians. The analysis also allowed the researcher to determine if the variables under examination had a combined influence, or interactive effect.

Results

GENDER COMPARISONS

The results of the factorial ANOVA examining the rate of change in SAT9 reading scores revealed a statistically significant effect for gender such that:

- Females had larger change scores than males;
- Both groups had change scores in the negative direction; and,

- The effect size for this statistically significant analysis was negligible (partial eta squared =.006).

The results of the factorial ANOVA examining the rate of change in SAT9 mathematics scores showed no statistically significant effects.

RACE COMPARISONS

The results of the factorial ANOVA examining the rate of change in writing scores evidenced a statistically significant effect for race such that:

- Caucasian students had larger change scores than African American students;
- Caucasian students demonstrated a decline in writing scores, while,
- African American students evidenced an increase in writing scores;
- However, the effect size for this analysis was negligible. (partial eta squared=.001)[5]

The results of the factorial ANOVA examining the rate of change in SAT9 reading and mathematics scores showed no statistically significant effects.

INCOME LEVEL COMPARISONS

The results of the factorial ANOVA examining the rate of change in SAT9 reading scores revealed a statistically significant effect for low-income status was found such that:

- Students considered "not low income" had larger change scores than those considered "low income";
- Both groups had change scores in the negative direction.
- The effect size for this statistically significant analysis was negligible (partial eta squared =.001).

The results of the factorial ANOVA examining the rate of change in SAT9 mathematics scores showed no statistically significant effects.

SUMMARY

Female students demonstrated larger decreases in reading than males between 8th and 10th grade. In addition, students classified as "not low income" demonstrated larger decreases

[5] It should be noted that although the paired t-test conducted for question 2 also revealed an increase in writing scores for African American students, this effect was not statistically significant.

than low income students in SAT9 reading scores from 2000 to 2002. Caucasian students demonstrated lower performance than African American students on the writing portion of the DSTP. Although these analyses were statistically significant, the effect sizes were very small.

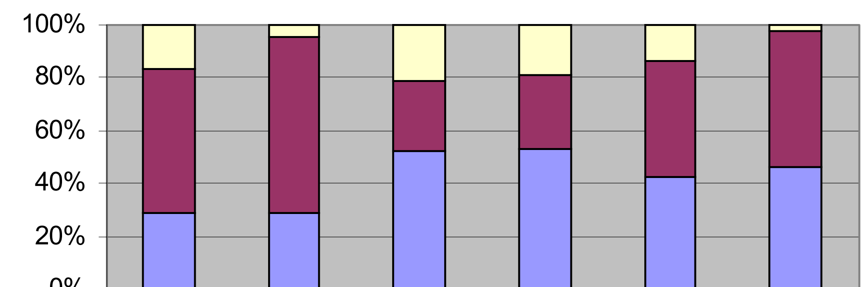
Question 1c: How much movement is there in performance levels on the DSTP from one test administration to the next?

Method

Descriptive statistics were used to examine how much movement in DSTP performance levels occurred from 2000 to 2002. Students were classified as performing above the standard (performance levels 4 and 5), meeting the standard (performance level 3), or below the standard (performance levels 1 and 2) based on the cut-scores associated with each content area on the DSTP.

Secondary Cohort: Performance Levels for 2000 and 2002

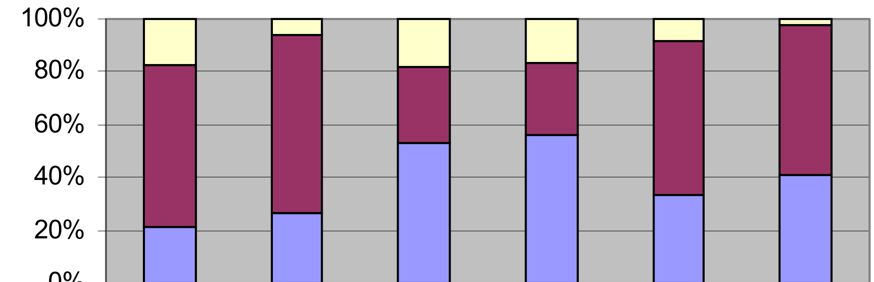
Table 1c1: Secondary Cohort: All Students



	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	17	4	22	19	13	2
Meets	55	67	26	28	44	52
Below	29	29	53	53	43	47

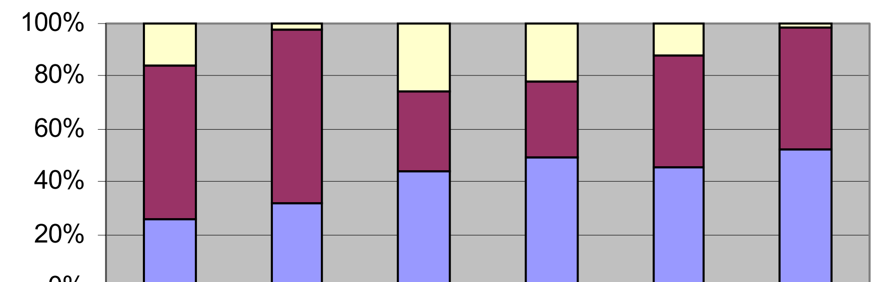
Table 1c1 above shows that between 2000 and 2002, the percentages of students scoring "below the standard" remained relatively stable across all three content areas. In addition, the overall percentage of students scoring "above" the standard decreased in all three content areas between 2000 and 2002.

Table 1c2: Secondary Cohort: All Female Students



	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	17	6	18	16	8	2
Meets	62	67	29	28	58	57
Below	21	27	53	56	33	41

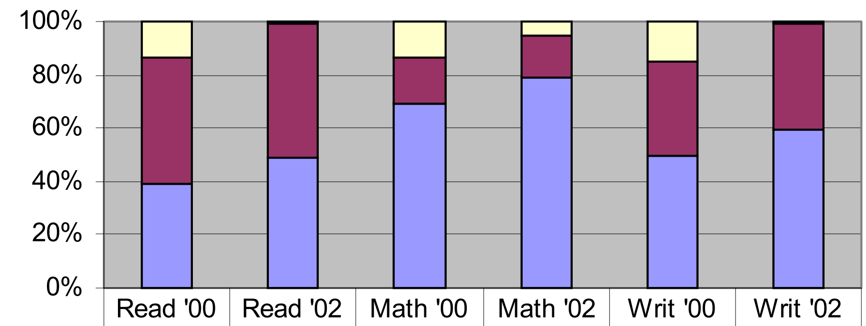
Table 1c3: Secondary Cohort: All Male Students



	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	16	3	26	22	12	2
Meets	58	66	30	29	42	46
Below	26	32	44	49	46	52

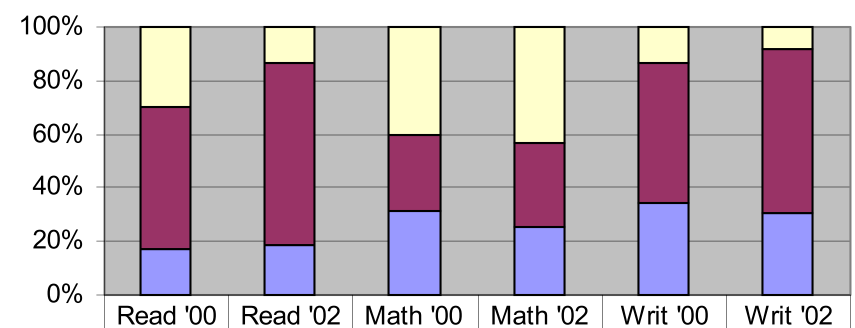
As seen in Tables 1c2 and 1c3, from 2000 to 2002, there was a decrease in the percentage of both male and female students scoring “above the standard” in all three content areas. The number of students “meeting the standard” increased in reading and writing but this may be due to fewer students scoring 4 or 5. The percentages of male and female students increased at the “below” level, performance levels 1 and 2, between the 8^{th} and 10^{th} grade testing periods.

Table 1c4. Secondary Cohort: African American Students



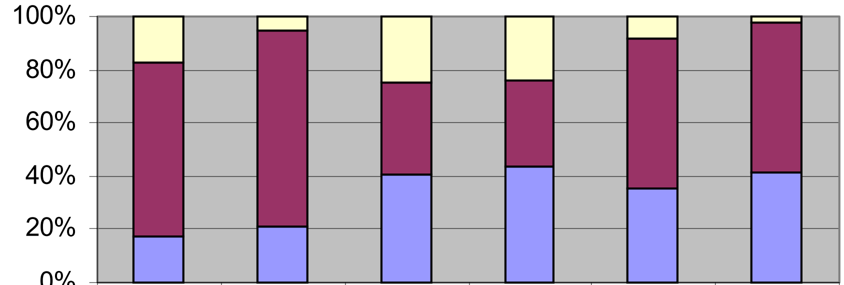
	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	13	1	14	5	15	1
Meets	47	50	17	16	35	40
Below	39	49	69	79	50	60

Table 1c5. Secondary Cohort: Asian Students



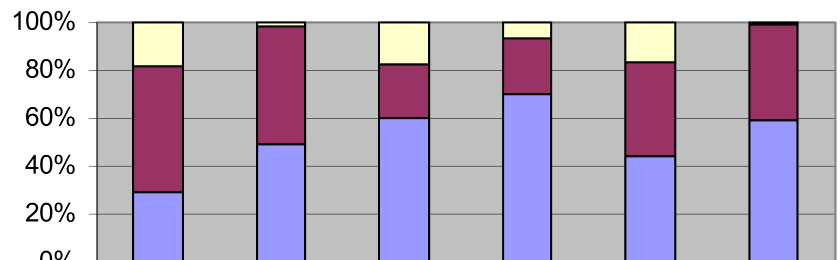
	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	30	13	41	43	13	8
Meets	53	68	28	32	53	61
Below	17	19	31	25	34	31

Table 1c6. Secondary Cohort: Caucasian Students



	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	18	6	25	24	8	2
Meets	65	73	35	32	56	56
Below	17	21	41	44	35	42

Table 1c7. Secondary Cohort: Hispanic Students

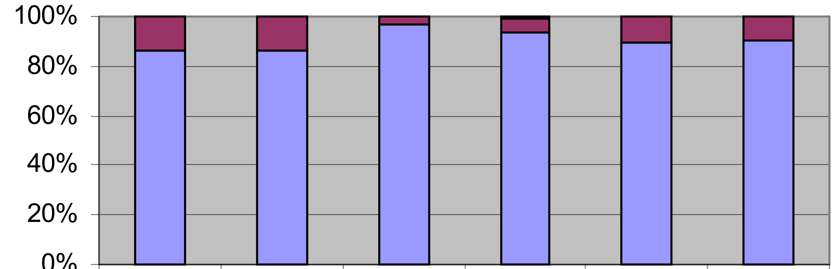


	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	18	2	18	7	17	1
Meets	53	49	22	23	39	40
Below	29	49	60	70	44	60

With the exception of Asian students' performance in mathematics, between 2000 and 2002, there is a decrease in the percentage of students scoring "above the standard" in all 3 content areas. There also appears to be an increase in the percentage of African-American, Caucasian, and Hispanic students scoring at performance levels 1 and 2 (below) between the two testing periods. This decline is particularly pronounced in reading among Hispanic students, where the percentage increased from 29% in 2000 to 49% in 2002.

Special Education Comparisons

Table 1c8. Secondary Cohort: Special Education Students

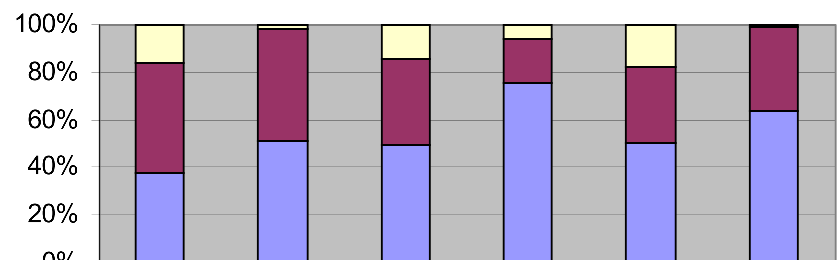


	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	0	0	0	1	0	0
Meets	14	14	4	6	11	9
Below	86	86	96	93	89	90

Between 2000 and 2002, there was little change in the performance of special education students in any of the content areas. Significant numbers (86% to 96%) are scoring at Performance Levels 1 and 2.

Income Level Comparisons

Table 1c9. Secondary Cohort: Students from Low Income Families

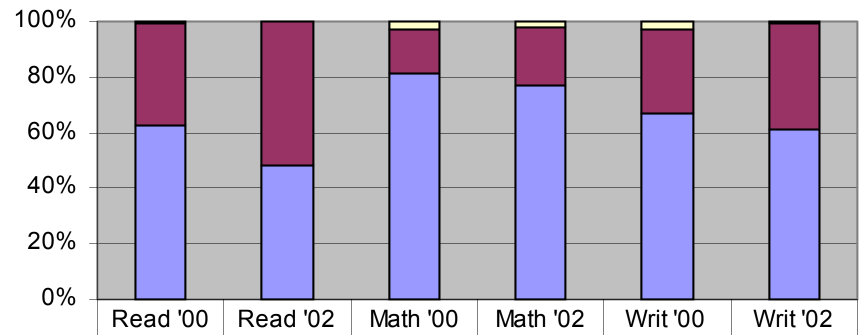


	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	16	1	18	6	18	1
Meets	46	47	47	19	32	36
Below	38	51	65	75	50	64

Table 1c9 illustrates that between 8th and 10th grade, there was a decline in performance among students from low-income families in all content areas.

TITLE I COMPARISONS

Table 1c10. Secondary Cohort: Students enrolled in Title I Programs



	Read '00	Read '02	Math '00	Math '02	Writ '00	Writ '02
Above	1	0	3	3	3	0
Meets	37	52	16	20	31	38
Below	63	49	81	77	67	62

There was an improvement in reading among students in Title I programs with a decrease among those scoring at the "below the standard" levels; yet very few of these students improved to the level of exceeding the reading, mathematics, or writing standards.

Secondary Cohort: Performance Level Changes between 2000 and 2002

The previous analysis provided information on the percentage of students classified into each of the various performance levels in 2000 and 2002; it provided a means to view overall changes in performance of a specific group of students. Hence, it does not address how much movement occurred on an individual basis. In order to examine this issue, simple counts were made of the number of students who improved by at least one performance level, students who performed at the same level both times, and the number of students who declined by at least one performance level.

Table 1c11. Overall Changes in Performance of All Students

	IMPROVED		REMAINED THE SAME		DECLINED		TOTAL	
	N	%	N	%	N	%	N	%
WRITING	836	16%	3142	60%	1244	24%	5222	100%
READING	421	8%	3633	71%	1092	21%	5146	100%
MATH	764	15%	3601	70%	812	16%	5177	100%

Table 1c12: Performance of Students by Gender

	Improved		Remained the Same		Declined		Total	
Male								
	N	%	N	%	N	%	N	%
Writing	451	18%	1477	58%	610	24%	2538	100%
Reading	214	9%	1733	70%	544	22%	2491	
Math	396	16%	1692	67%	423	17%	2511	100%
Female								
Writing	385	14%	1665	62%	634	24%	2684	100%
Reading	207	8%	1900	72%	548	21%	2655	100%
Math	368	14%	1909	72%	389	15%	2666	100%

Table 1c13: Performance of Students by Race

	Improved		Remained the Same		Declined		Total	
	N	%	N	%	N	%	N	%
African American								
Writing	223	18%	701	55%	345	27%	1269	100%
Reading	122	10%	853	68%	283	22%	1258	100%
Math	110	9%	957	75%	202	16%	1269	100%
Asian								
	N	%	N	%	N	%	N	%
Writing	17	16%	68	65%	20	19%	105	100%
Reading	7	7%	71	68%	27	26%	105	100%
Math	17	16%	80	76%	8	8%	105	100%
Hispanic								
	N	%	N	%	N	%	N	%
Writing	32	20%	74	47%	53	33%	159	100%
Reading	9	6%	110	69%	40	25%	159	100%
Math	125	16%	100	64%	32	20%	157	100%
Caucasian								
	N	%	N	%	N	%	N	%
Writing	558	15	2290	62	822	22	3670	100%
Reading	280	8	2587	72	738	20	3605	100%
Math	609	17	2451	68	567	16	3627	100%

Table 1c14: Performance of Students from Low-Income Families

	Improved		Remained the Same		Declined		Total	
	N	%	N	%	N	%	N	%
Writing	162	16%	529	53%	310	31%	1001	100%
Reading	80	8%	667	68%	241	24%	988	100%
Math	104	10%	705	71%	190	19%	999	100%

Table 1c15: Performance of Students from Title I Programs

	Improved		Remained the Same		Declined		Total	
	N	%	N	%	N	%	N	%
Writing	56	24%	134	57%	44	19%	234	100%
Reading	56	24%	154	66%	22	9%	232	100%
Math	30	13%	189	80%	16	7%	235	100%

Overall, the majority of students performed at approximately the same performance level in 2000 and 2002, with the next largest group likely to have declined by at least one performance level. The only notable exception was that students in Title I programs were more likely to improve by at least one performance level in reading and writing scores over the other groups.

SECTION 2: BEHAVIORAL EFFECTS

Question 2a: What are the demographic characteristics of students who are given out-of-school suspension?

Elementary Cohort: Suspensions During the 2000-2001 School Year

Within the *elementary cohort*, 904 students were suspended during the 2000-2001 school year. The total number of out-of-school suspensions received by any single student ranged from 1-15 with 50% of the students (n=454) receiving only one out-of-school suspension and 21% (n=185) receiving two. The total number of days spent out of school ranged from 1-134 days, with 54% spending 3 or less days out of school and 75% spending 6 days or fewer out of school.

Elementary Cohort: Suspensions During the 2001-2002 School Year

Within the *elementary cohort*, 1378 students were suspended during the 2001-2002 school year. The total number of out-of-school suspensions received by any single student ranged from 1-15 with 48% of the students receiving only one out-of-school suspension and 20% receiving two. The total number of days spent out of school ranged from 1-46 days, with 50% of students spending 3 days or fewer and 77% spending 7 days or fewer out of school.

The demographic characteristics of the students given out-of-school suspensions over the two- year period are as follows:

Table 2a1: Elementary Cohort: Out-of-School Suspensions by Gender, 2000-01 and 2001-02

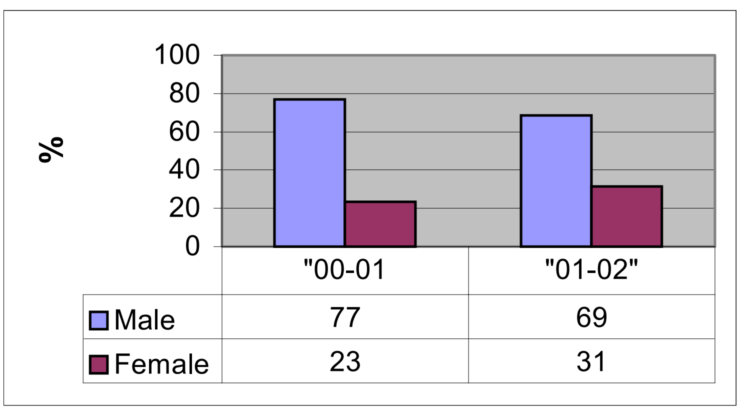
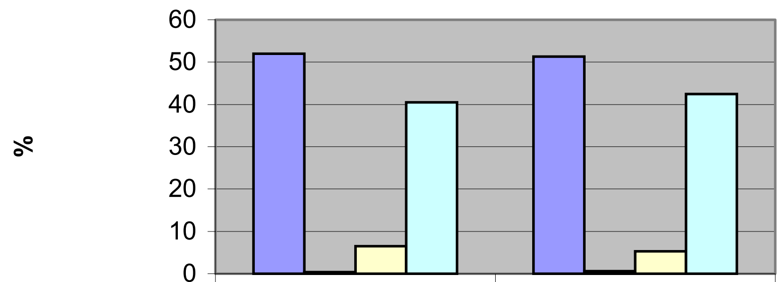
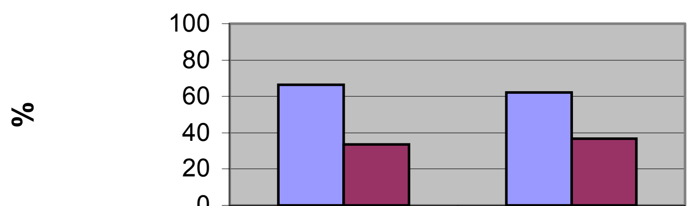


Table 2a2: Elementary Cohort: Out-of-School Suspensions by Race, 2000-01 and 2001-02



	"00-01	"01-02"
Afr. Amer.	52	51
Asian	0	1
Hispanic	7	5
Caucasian	41	43

Table 2a3: Elementary Cohort: Out-of-School Suspensions by Income Status, 2000-01 and 2001-02



	"00-01	"01-02"
Low Income	66	62
Non-Low Income	34	37

Secondary Cohort: Suspensions During the 2000-2001 School Year

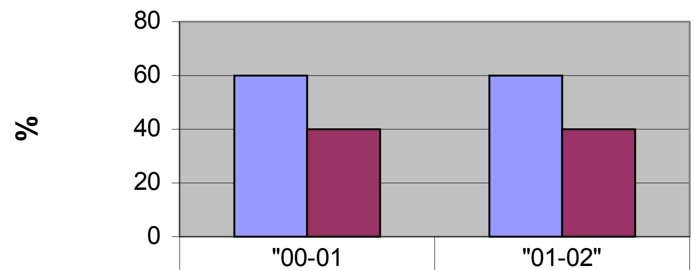
Within the *secondary cohort*, 1471 students were suspended during the 2000-2001 school year. The total number of out-of-school suspensions received by any single student ranged from 1-15 with 45% of the students receiving only one out-of-school suspension and 21% receiving two. The total number of days spent out of school ranged from 1-185 days, with 52% of students spending 4 days or fewer and 78% spending 8 days or fewer out of school.

Secondary Cohort: Suspensions During the 2001-2002 School Year

Within the *secondary cohort*, 1439 students were suspended during the 2001-2002 school year. The total number of out-of-school suspensions received by any single student ranged from 1-15 with 46% of the students receiving only one out-of-school suspension and 19% receiving two. The total number of days spent out of school ranged from 1-90 days, with 49% of students spending 3 days or fewer and 78% spending 8 days or fewer out of school.

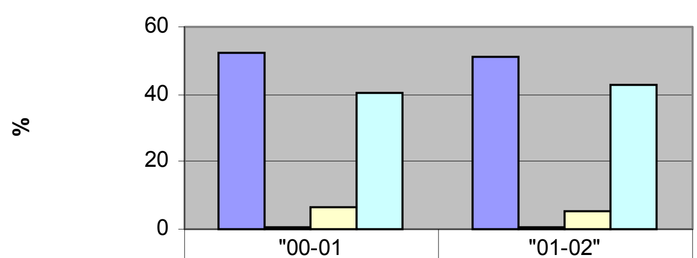
The demographic characteristics of the students given out-of-school suspensions over the two- year period are as follows:

Table 2a4: Secondary Cohort: Out-of-School Suspensions by Gender, 2000-01 and 2001-02



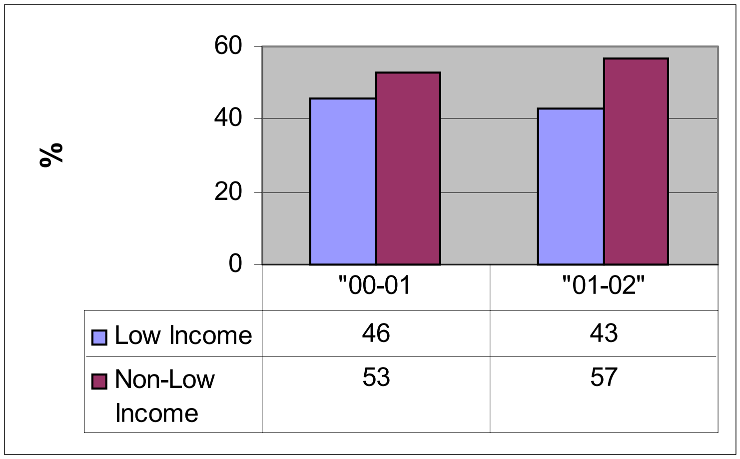
	"00-01	"01-02"
Male	60	60
Female	40	40

Table 2a5: Secondary Cohort: Out-of-School Suspensions by Race, 2000-01 and 2001-02



	"00-01	"01-02"
Afr. Amer.	52	51
Asian	0	1
Hispanic	7	5
Caucasian	41	43

Table 2a6: Secondary Cohort: Out-of-School Suspensions by Income Status, 2000-01 and 2001-02



A Closer Examination of the Demographic Characteristics Associated with OSS

An examination of the demographic characteristics associated with out-of-school suspension suggested that students who are low income, of a minority status, or male may be over-represented in suspensions. In order to address this issue statistically, logistic regression analyses were performed.

Logistic regression is a statistical procedure for estimating the relationship between one or more predictor variables and the likelihood that an individual is a member of a particular group. For the purposes of the current investigation, group membership was defined as whether or not a student had been given a suspension during the school year. Gender, minority status comprised of traditionally higher performing students (Asian and Caucasian students) and traditionally lower performing students (African American, American Indian and Hispanic, and income status (low income vs. not low income) were used to predict whether or not a student was suspended.

For the elementary cohort, the results of the logistic regression revealed that gender, minority status, and income status were statistically significant predictors of suspensions given during the 2001-2002 school year. According to the logistic regression, males were more likely to be suspended than females. In addition, the odds of being suspended were 2.39 times greater for historically lower performing students compared to historically higher performing students and 2.64 times greater for low income students than students who were not low income.

For the secondary cohort, the results of the logistic regression revealed that gender, minority status, and income status were statistically significant predictors of suspensions given during the 2001-2002 school year. According to the logistic regression, males were more likely to be suspended than females. In addition, the odds of being suspended were 2.39 times greater for historically lower performing students compared to historically higher performing students and 1.89 times greater for low income students than students who were not low income.

Question 2b: What is the relationship of out-of-school suspension to performance on the DSTP?

Considering the impact that out-of-school suspension might have on a students' opportunity to learn and subsequently upon his/her academic performance, it is important to examine the relationship between suspension and DSTP performance.

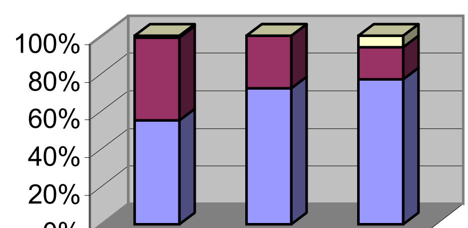
Method

In addition to examining the demographic characteristics of suspended students, the current investigation examined the relationship between suspension and student performance on the DSTP for the secondary cohort during the 2001-2002 school year. This study also examined the association between suspension and grade level (retention) as of September 30, 2001 for the secondary cohort during the 2001-2002 school year.

Results

According to the data, 44% of the secondary students (10th grade) suspended during the 2001-2002 school year met the standard in reading on the DSTP and only 1% performed at levels 4 or 5. On the writing portion of the DSTP, 28% of the 10th grade students suspended during the 2001-2002 school year met the standard, that is, scored at performance level 3, and none exceeded it. In mathematics, only 17% of the secondary students suspended during the 2001-2002 school year met the DSTP mathematics standard and only 6% exceeded it.

Table 2b1. DSTP Performance of Secondary Students Suspended in 2001-02



	Reading	Writing	Math
Above	1	0	6
Meets	44	28	17
Below	55	72	77

In addition, it was found that 56% of the students in the secondary cohort who were suspended during the 2001-2002 school year were enrolled at the 9th grade level or below on September 30, 2001. The rest (44%) of the cohort was enrolled in tenth grade, that is, were at the expected grade level. Consequently, more than half of the students who had been suspended during the 2001-2002 school year had been retained at least once.

SECTION 3: RETENTION

The National Research Council in its report on high-stakes assessment systems cites research that indicates that retention rates frequently increase under those conditions. There is significant research that also indicates that retention in grade, when typically practiced, is ineffective in improving students' achievement and frequently leads to students dropping out of school. In addition, other research (cited in the November 2000 study) shows that retention is more likely to occur among certain groups of students; i.e., it is more common among African American than among Caucasians, males are more likely to be retained than females, and students from poor families are more likely to be retained.

For these reason, we continue in this report to track the incidence of retention and demographic characteristics of retained students. It was examined in two ways.

Question 3a: What are the demographic characteristics of students who are retained?

Method #1

Demographic characteristics were examined for students in the elementary cohort who were enrolled in grade 6 during the 2000-2001 and 2001-2002 school years. They were also examined for students in the secondary cohort who were enrolled in grade 9 during the 2000-2001 and 2001-2002 school years.

Results

Elementary Cohort

The demographic characteristics of students in the *elementary cohort* who were enrolled in grade 6 during the 2000-2001 and 2001-2002 school years (N=267) indicated that most of the students were male, African American, and low income.

Table 3a1: Gender of Retained Students in Elementary Cohort (grade 6)

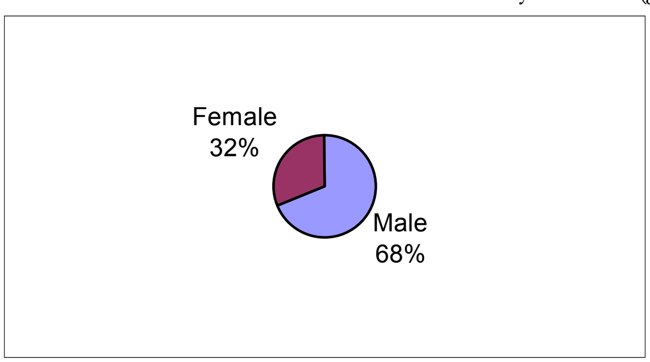


Table 3a2: Race of Retained Students in Elementary Cohort (grade 6)

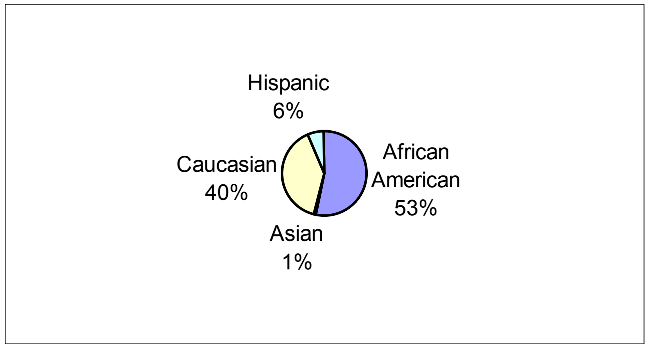
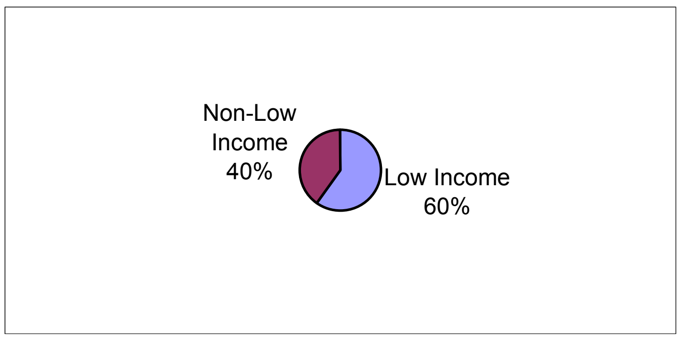


Table 3a3: Income Status of Retained Students in Elementary Cohort (grade 6)



Secondary Cohort

The demographic characteristics of students in the *secondary cohort* who were enrolled in grade 9 during the 2000-2001 and 2001-2002 school years (N=863) indicated that the largest number of students were male and African American.

Table 3a4: Retained Students in Secondary Cohort by Gender (grade 9)

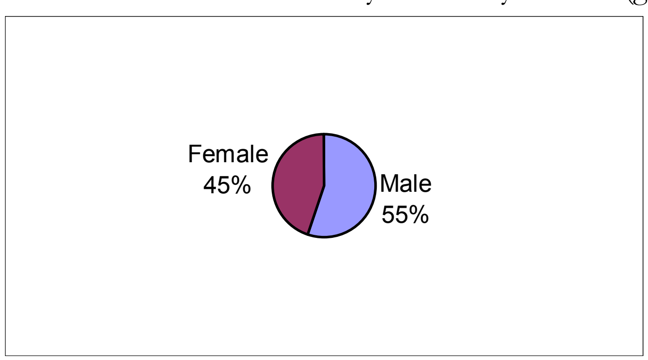


Table 3a5: Retained Students in Secondary Cohort by Race (grade 9)

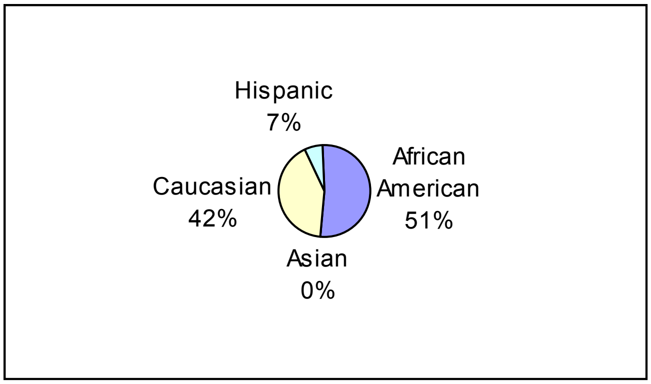
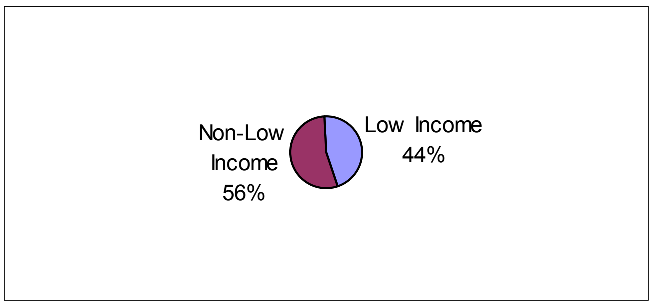


Table 3a6: Retained Students in Secondary Cohort by Income Status (grade 9)



Method #2

The second method used to examine retention involved the analysis of cross-sectional data sets, not cohorts. Ten years of student retention data beginning in 1992 through 2002 were included in this analysis. Data were examined for kindergarten through 8th grade. Due to the complexity of issues regarding grade placement that begins in grade 9, that is, grade level based on credit accumulation, the analysis is limited to k-8. In this analysis, a student was considered "retained" if he/she were assigned to the same grade level in the spring and then the following fall of the same year. While this analysis does not capture multiple retentions, it does provide another means for us to examine the original question, "What are the demographic characteristics of students who are retained?"

In addition, this analysis provides a different perspective as compared to method #1 in regards to examining each demographic grouping. Here we present the percentage of students retained in each grade over a 10-year period in relation to the entire student population from that demographic grouping[6]. For example, in Table 3a7 below, the analysis shows that in 1992, 5% of all African American students were retained at the end of their

---

[6] Note, in this and the subsequent analyses, the American Indian students have not been represented due to the very low numbers of American Indians in the state's population (the total N ranges from 11 to 40 over the ten-year period across all grades examined).

kindergarten year and only 2.5% of all Caucasian students were held back. Moreover, with the exception of 1997 and 2001, over the ten-year period, the percentage of African-American and Hispanic students retained remained higher than either the Asian or Caucasian student population.

Table 3a7. Retained Kindergarten Students by Race from 1992 – 2002

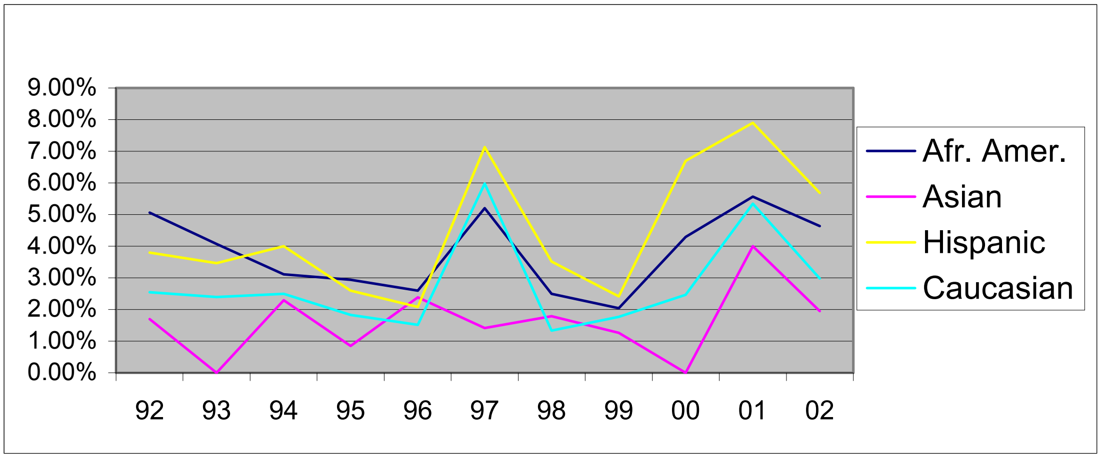


Table 3a8. Retained 1st Grade Students by Race from 1992 –2002

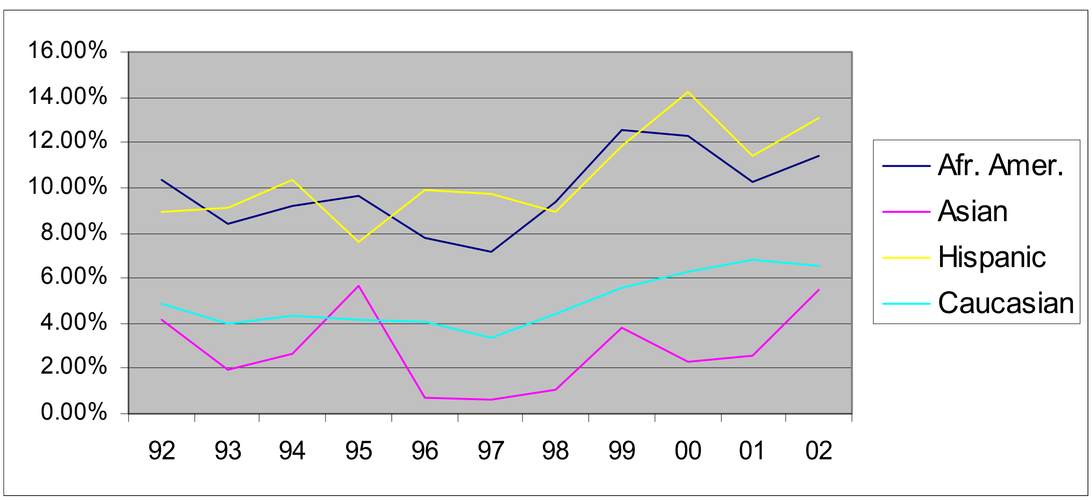


Table 3a9. Retained 2nd Grade Students by Race from 1992 –2002

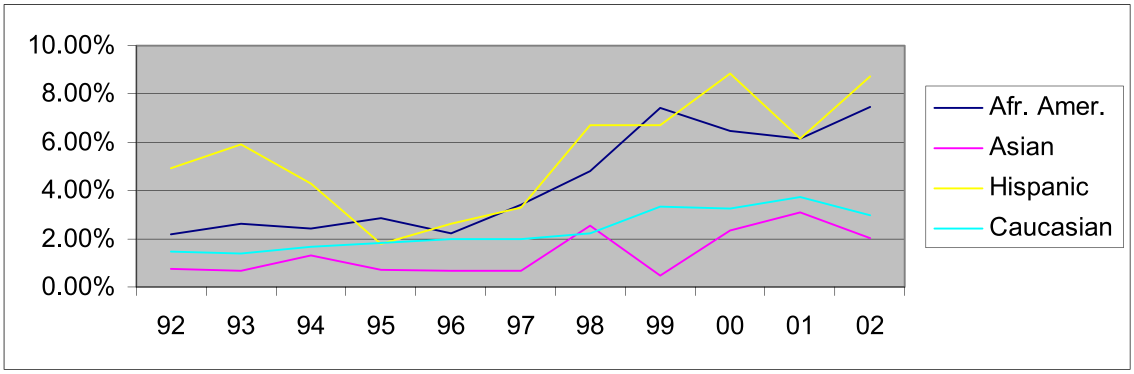


Table 3a10. Retained 3rd Grade Students by Race from 1992 –2002

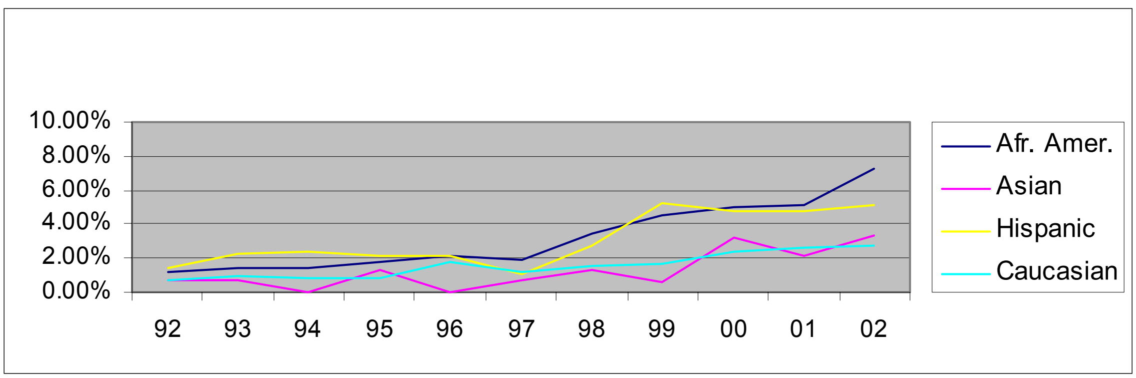


Table 3a11. Retained 4th Grade Students by Race from 1992 –2002

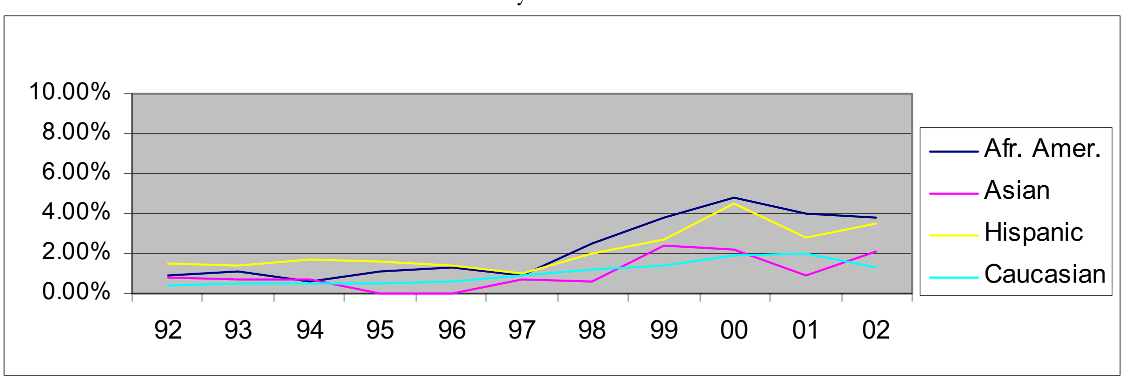


Table 3a12. Retained 5th Grade Students by Race from 1992 –2002

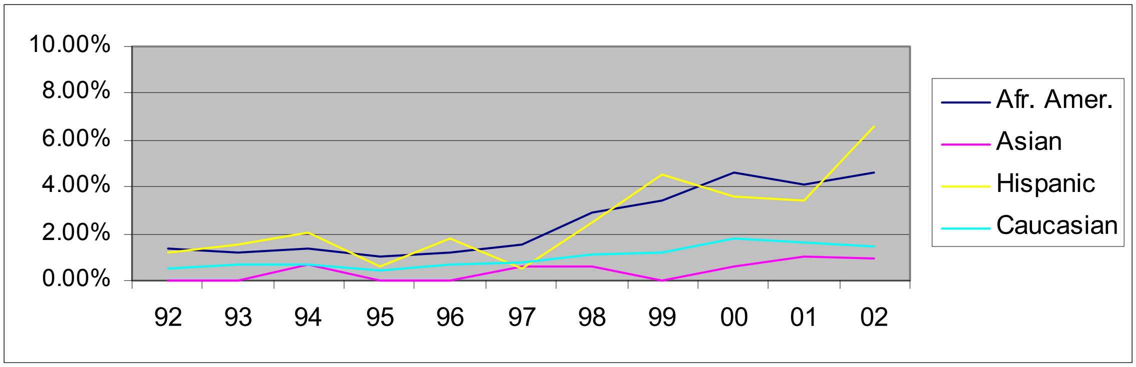


Table 3a13. Retained 6th Grade Students by Race from 1992 –2002

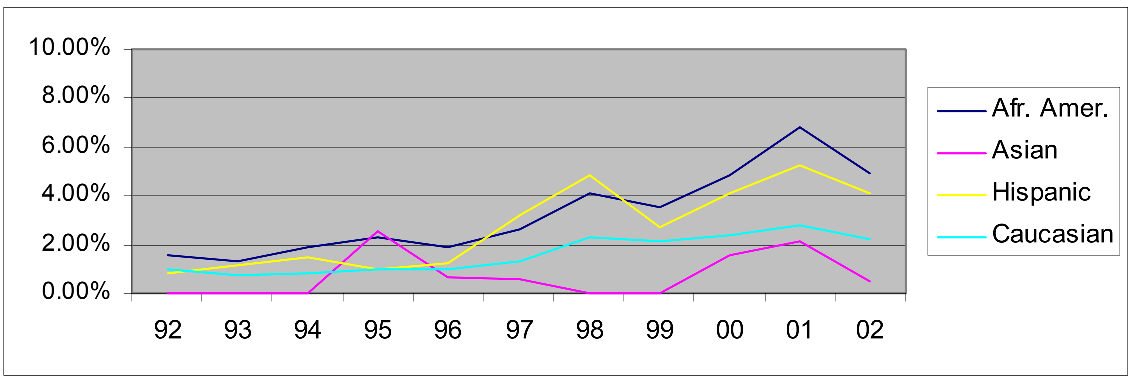


Table 3a14. Retained 7th Grade Students by Race from 1992 –2002

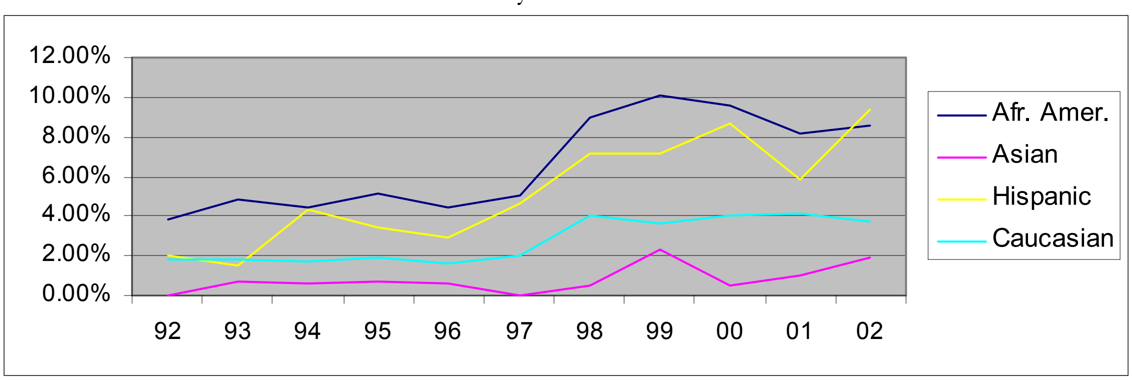
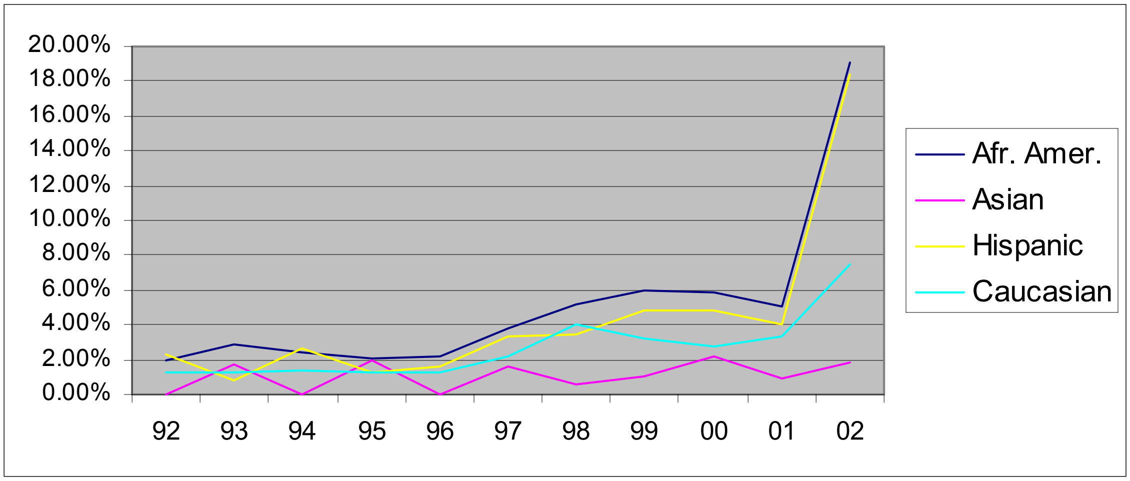


Table 3a15. Retained 8th Grade Students by Race from 1992 –2002



One can notice a few trends from the 10-year analyses in the previous tables:

1) Over the past ten years, African American and Hispanic students have been retained at rates that are consistently higher than their Caucasian and Asian peers;

2) In 2002, the first year that the state's student accountability plan went into full effect, there was a major increase in the number of 8th grade students retained. This trend was even more notable among African American and Hispanic students; and,

3) During the past 2 years, 2001 and 2002, there appears to be an increase in the incidence of kindergarten retentions across all demographic groups.

Question 3b: After being retained, how do students fare with respect to behavioral effects and special education status?

Method

In order to examine this issue, a logistic regression was performed using retention status to predict whether or not a student was suspended. For the elementary cohort, whether a student was enrolled in 6th grade or 7th grade during the 2001-2002 school year was used to predict whether or not the student was suspended during the 2001-2002 school year.

For the secondary cohort, whether a student was enrolled in 9th grade or 10th grade during the 2001-2002 school year was used to predict whether or not the student was suspended during the 2001-2002 school year.

Behavior-related Results

The results of the logistic regression analysis revealed that retention was a significant predictor of suspension and that the odds of being suspended were *2.09 times greater* for students in the elementary cohort who had been retained than students in the elementary cohort who had been promoted. The odds of being suspended were *6.33 times greater* for students in the secondary cohort who had been retained than students in the secondary cohort who had been promoted.

Special Education Results

Regarding special education placements before and after retention for the elementary cohort, 80% of the retained students were enrolled in regular education classes during both the 2000-2001 and 2001-2002 school years and 16% were enrolled in a special education class during both the 2000-2001 and 2001-2002 school years. Just over 4% of the retained students in the elementary cohort were in a special education class during only one of those school years mentioned above (six students went from a special education to a regular education classification and six students went from a regular education to a special education classification).

Concerning special education placements before and after retention for the secondary cohort, 73% of the retained students were enrolled in regular education classes during both the 2000-2001 and 2001-2002 school years and 25% were enrolled in a special education class during both the 2000-2001 and 2001-2002 school years. Just over two percent of the retained students in the secondary cohort were in a special education class during only one of those school years mentioned above (14 students went from a special education to a regular education classification and five students went from a regular education to a special education classification).

Consequently, these analyses revealed that placement into special education does NOT appear to be a prevalent practice in response to retention in grade.

SECTION 4: ATTRITION RATES

A question generated by the State Board of Education research subcommittee focused on the impact of Delaware's student accountability system on dropout rates. This year's longitudinal study only captures the progress of students through grade 10, i.e., the secondary cohort. Thus the question of dropout impact could not be fully addressed with the data available at the time of these analyses. In addition, our 2001 report alluded to a significant number of students who are identified in the state database as "missing." These students are not officially classified as dropouts yet they are not accounted for in the state's accountability system.

Method

Consequently, at the subcommittee's request, we identified students in both our elementary and secondary cohorts who are currently identified as "missing" and explored the demographic information for each of these groups. We identified the demographic information on this group by looking at their data from the beginning of our study. By comparing the percentages of students from each demographic grouping that are designated as missing to the overall percentages of that same group within the entire cohort, one can determine whether there is over-representation of any particular grouping in regards to attrition. We found the following:

<u>Elementary Cohort</u>

As of September 30, 2001, the entire elementary cohort that we have been tracking includes 8404 students; approximately 15% (n= 1233) were designated as "997" or "missing".
The profile of this group shows that:

Table 4.1: Demographic Characteristics of Students Identified as "Missing" within the Elementary Cohort

	"MISSING GROUP"		ENTIRE ELEMENTARY COHORT	
	n	%[7]	n	%
Special education	153	12.4%	1433	17.1%
Limited English Proficient	31	2.5%	173	2.1%
Title I reading	95	7.7%	1024	12.2%
Low Income	518	42%	3711	44.2%
African American	320	26%	2564	30.5%
Asian	34	2.8%	140	1.7%
Hispanic	80	6.5%	413	4.9%
Caucasian	796	64.6%	5260	62.6%

[7] NOTE: Percentages were calculated within each subgroup,; therefore, overall totals will exceed 100%. In addition, since students are counted in multiple categories, the overall total will exceed 1233.

Secondary Cohort

As of September 30, 2001, the entire secondary cohort that we have been tracking includes 8732 students; approximately 15% (n= 1336) were designated as "997" or "missing".
The profile of this group shows that:

Table 4.2: Demographic Characteristics of Students Identified as "Missing" within the Secondary Cohort

	"MISSING GROUP"		ENTIRE SECONDARY COHORT	
	n	%[8]	n	%
Special education	162	12.1%	1319	15.1%
Limited English Proficient	28	2.1%	114	1.3%
Title I (reading and/or math)	49	3.7%	377	4.4%
Low Income	652	48.8%	3691	42.3%
African American	354	26.5%	2670	30.6%
Asian	31	2.3%	145	1.7%
Hispanic	98	7.3%	398	4.6%
Caucasian	849	63.5%	5492	62.9%

Within the elementary cohort, there appears to be some over-representation of Hispanic students. Within the secondary cohort, over-representation as 'missing' appears with students classified as Limited English Proficient (LEP), students from low-income families, and Hispanic students.

[8] NOTE: Percentages were calculated within each subgroup,; therefore, overall totals will exceed 100%.

Policy Considerations

- In light of the new federal regulations under the No Child Left Behind legislation, what are the implications for the state in response to the overall decline in students' reading performance between 8th and 10th grade? Subsequently, what are the needs in regards to teacher professional development, certification, preparation, and accountability?

- The retention findings over each year of this study along with the decade of data provided by the Delaware Department of Education indicate that African American males are more likely to be retained than any other group of students. Considering the sanctions imposed by the NCLB legislation and the state's commitment to closing the achievement gap, what should districts/schools be doing to reverse this trend?

- Considering the finding that students enrolled in Title I programs at the secondary level have improved at least one performance level in reading and writing, what components of Title I programs might be expanded to other low-performing students?

- In spite of the proliferation of pre-school programs and readiness academies currently offered within the state, in the last 2 years there has been a considerable increase in the number of students retained at the kindergarten and 8th grade levels. What is being done to ascertain which of these programs are having positive effects on student performance and which are ineffective?

- In both 8th and 10th grade, very few special education students met the standards in any of the 3 content areas. Considering the implications of NCLB in this regard, should the state reconsider the issue of whether holding special education students to the same standards as regular education students is realistic or equitable?

APPENDIX A: PSYCHOMETRIC ISSUES

Technical Details Concerning the DSTP

To address the student achievement issue, a few technical details concerning the DSTP should be addressed. The DSTP is composed of multiple choice, short answer, and extended response items. Results are reported out in the form of national percentile ranks, standards-based scores, and performance levels.

Percentile Ranks

The national percentile rankings are based on abbreviated versions of the reading comprehension and the mathematical problem solving subsets of the Stanford Achievement Test series, 9th Edition (SAT9). The SAT9 is a norm-referenced test published by Harcourt Brace Educational Measurement.

Although percentile ranks can provide useful information by referencing student performance against set norms, percentile ranks cannot be manipulated mathematically because there are not equal intervals between them. For example, the difference between a percentile rank of 5 and 10 is not the same as the difference in achievement as the difference between a percentile rank of 50 and 55. This point is worth noting not only for the analyses that follow but also for the purpose of avoiding incorrect conclusions based on cursory examinations of data.

In order to be used in statistical analyses the national percentile rankings must be converted to another metric, in this case normal curve equivalents (NCEs). NCEs can range from 1 to 99 and provide an equal-interval scale which makes them amenable to mathematical manipulation. For the purposes of the current study, students' scores on the SAT9 portion of the DSTP will be reported in NCE units.

Standards-based Scores

The standards-based score reported for the DSTP ranges from 150 to 800 and is based on students' responses to items developed in Delaware and a subset of the SAT9 items that are considered to be aligned with the Delaware content standards. According to the 2000 DSTP Executive Summary, "students in the earlier grades should tend to score towards the lower part of the scale, while students in the upper grades should tend to score towards the higher part of the scale".[8]

The expectation that students in the early grades will have lower scores than students in the upper grades is a function of the manner in which the standards-based scores are scaled. A vertical scaling system has been applied to the scores which means that a score of 400 in 3rd grade is not equivalent to a score of 400 in 5th grade. This also means that if a student's score is the same in both 3rd grade and 5th grade, they have actually done worse, instead of holding steady the lack of change in scores would represent a decline.

Vertical scaling becomes an issue when attempts are made to track student improvement longitudinally. Since there is an expected amount of increase from year to year, an increase reflects the manner in which the scores are scaled, not true gains. Hypothetically, a score of 400 in 3rd grade may be equivalent to a score of 450 in 5th grade. In this case, if a student did receive these scores, it would appear to represent a 50 point increase, but because of the vertical scaling the students' achievement is actually unchanged from 3rd to 5th grade. For this reason, the questions related to student achievement over time can only be examined by using the SAT9 data.