

Measuring the Effectiveness of After-school Programs Via Participants' Pre and Posttest
Performance Levels on the Georgia Criterion Referenced Competency Test

A Dissertation

Presented to

the Faculty of the School of Education

Liberty University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

By

Cheri Ogden

March 2008

Measuring the Effectiveness of After-school Programs Via Participants' Pre and Posttest
Performance Levels on the Georgia Criterion Referenced Competency Test
by Cheri Ogden

APPROVED:
COMMITTEE CHAIR

Clarence Holland, Ed.D

COMMITTEE MEMBERS

Samuel Smith, Ed.D.
Marie Klofenstine, Ed.D.

CHAIR, GRADUATE STUDIES

Scott B. Watson, Ph.D.

Abstract

Schools failing to demonstrate Adequate Yearly Progress on annual state assessments are given federal and state funds to support supplemental educational services. This study was designed to measure the effectiveness of two after-school programs in one middle school in a medium-sized county in Georgia. The school has failed to demonstrate Adequate Yearly Progress for eight years and has been offering supplemental services for six years. The after-school programs are funded and students are offered opportunities to participate based on low performance on the annual assessment, the Georgia Criterion Referenced Competency Test. The effectiveness of the after-school programs was assessed based on the percentage of program participants who moved from Level 1 (not meeting standard) to Level 2 (meeting) or Level 3 (exceeding). Control groups were established for each after-school program. Consideration was also given to the participants' frequency of attendance in each program to determine if participants who attended frequently improved more than participants who attended infrequently. Program participants failed to demonstrate improved achievement greater than nonparticipants and increased attendance in the programs did not seem to positively impact student achievement.

Dedication

This dissertation is dedicated to the people in my life who served as the support system that is truly necessary to make this dream become a reality.

First, I thank my Heavenly Father, for it is He who put this dream in my heart. While I have ventured off on my own course from time to time, Jesus always guides me to the path He has chosen for my life.

My husband, Fred, has been more patient and supportive than I imagined possible when I began this journey. Many of my working hours he provided the wind for my wings, and I am forever grateful.

Stevi and Jona, my daughters, have shown love and support through the process. Only a parent can understand how much it means for your children to support your individual dream, and I'm proud that my daughters did just that.

My parents, Leonard (Bud) and Midge Ellyson, thought I could do this long before I did. They were my original inspiration. They never fail to tell me that if it can be done, I am one who can do it.

Acknowledgments

I want to acknowledge the support and advice I have received from Dr. Clarence “Chick” Holland during the process of writing my dissertation. Dr. Holland has been faithful in his role as my committee chairman in offering spiritual and academic support. I am deeply grateful for his guidance, kind heart, and keen eye.

I also want to express sincere appreciation for my other committee members: Dr. Samuel Smith and Dr. Marie Klofenstine. This work is not easy, but kind yet honest support does make it possible.

Many of my friends at Liberty served as constant support, but Mindy Kushniroff was especially helpful. She was there for the phone calls of panic and emails of questions. Her encouragement sustained me when I most needed it.

Dr. Lisa Schoer served as a faithful friend and colleague, giving advice and correction as needed.

God lead me to Liberty University, where I found wonderful guidance and instruction. I appreciate the leadership and love so many provided for my work on this graduate degree.

Tables

Table	Page
Table 1 Requirements for Schools Based on No. of Years Failing to Make AYP	20
Table 2 SES Participants Who Outperformed Nonparticipants.....	28
Table 3 Score Ranges for GA's CRCT by Subject and Grade	33
Table 4 Validation Items for Flagging Criteria Based on P-value	36
Table 5 Item Review Results for the 2006 GA CRCT	38
Table 6 Reliability and SEM for GA Middle School 2006 CRCT	39
Table 7 Difference in Pre and Posttest Results for After-School Academy	56
Table 8 Difference in Pre and Posttest Results for Saturday Scholars	61
Table 9 Difference in Pre and Posttest Results for After-School Academy by Attendance	63
Table 10 Difference in Pre and Posttest Results for Saturday Scholars by Attendance	65

Contents

	Page
Abstract.....	ii
Dedication.....	iii
Acknowledgements.....	iv
Tables	v
Chapter	Page
1 Introduction.....	1
Background of the Study	2
Problem Statement.....	4
Research Questions.....	4
Significance of the Study.....	6
Overview of the Methodology.....	7
Operational Definitions.....	9
2 Review of Related Literature.....	12
Issue of Time.....	12
Need for Extended Time.....	15
Funding Extended Time.....	15
Variety in After-school Programs.....	16
The No Child Left Behind Act.....	18
NCLB Supplemental Educational Services	19

	Programs Showing Improved Achievement	21
	Programs Failing to Show Improved Achievement.....	24
	Georgia After-school Programs	27
	Measuring Student Achievement.....	28
	Georgia Norm and Criterion Referenced Tests	30
	Georgia CRCT Score Interpretation	32
	Validity of the Georgia CRCT	34
	Reliability of the Georgia CRCT	37
	Summary.....	40
3	Research Methodology	43
	Research Context	43
	Research Participants.....	44
	Instruments.....	46
	Data Collection and Procedures.....	47
	Data Analysis	48
4	Results.....	51
	Establishing Comparison Groups	51
	After-School Academy and Matched Pairs’ Pretests.....	52
	After-School Academy and Matched Pairs’ Posttests	53
	Saturday Scholars and Matched Pairs’ Pretests	56
	Saturday Scholars and Matched Pairs’ Posttests	58
	After-School Academy Attendance Subgroups	62

	Saturday Scholars Attendance Subgroups	64
	Note Regarding Percentage of Performance Level Changes	66
5	Summary and Discussion.....	67
	Statement of the Problem.....	67
	Review of Methodology	67
	Summary of Results	68
	Interpretation of the Findings.....	72
	Related to Other Research	74
	Recommendations.....	76
	Limitations	79
	Recommendations for Further Research.....	79
	References.....	81

Chapter 1

Introduction

Secretary of Education Terrel Bell sent cries of concern echoing across our country in 1983 with his report to The National Commission on Excellence in Education: *A Nation at Risk* (Bell, 1983). According to this report, about 13% of American 17 year olds were functionally illiterate. Among minority youth, that illiteracy rate was thought to run as high as 40%. Average achievement of high school students was reported to be lower in 1983 than 26 years earlier when Sputnik was launched. Additionally, nearly 40% of the 17 year old students could not make inferences from written material, only 20% could write a persuasive essay, and only 33% could solve a math problem involving several steps. Sixteen years later *A Nation Still at Risk* (Bennett et al., 1999) indicated that some improvements were made, but too many American school children were still suffering from mediocrity. This second report stated that children of the poor and minorities were falling farther behind and were “stuck with” what the system tossed out to them. The No Child Left Behind Act of 2001 addressed the concern about the state of education in our country. This Act, unlike the previous reports, brought more than awareness. The Act brought new requirements that education associations prepare all students to meet rigorous standards of learning by 2014.

State and local boards of education were initially given few guidelines for accomplishing this task of helping all students achieve. Educational systems are now scrambling to find creative ways to help students reach the new level of standard. Efforts to improve student achievement frequently include extending learning time through

supplemental educational services (SES) as required for students and schools repeatedly failing to demonstrate adequate progress. Since the SES requirement is relatively new, the programs are often created without a model to follow, without a system for monitoring, and without clear guidance or regulation. This study examines the effectiveness of two such supplemental after-school programs funded and implemented to provide extra instructional support to students failing to meet the rigorous state standards.

Background of the Study

The No Child Left Behind Act (NCLB) of 2001 is a reauthorization of the Elementary and Secondary Education Act of 1965. The NCLB Act raises expectations for state and local education associations and students. Per this act, student achievement is measured annually by state adopted assessments. Students, schools, and systems are expected to show Adequate Yearly Progress (AYP) on these assessments, a key clause in the NCLB law. Schools and systems failing to demonstrate AYP must offer a variety of options to parents and students on an escalating continuum until a sustained demonstration of AYP is established.

A school or system failing to demonstrate Adequate Yearly Progress for three or more consecutive years moves into a category referred to as “Needs Improvement” or “NI” schools. Schools classified as “NI” schools must offer additional instructional programs to their students, which include some form of before or after-school tutoring or remedial classes. State education institutions, local education systems, and individual schools are responsible for determining the best ways to deliver these supplemental services.

The idea that extended school time improves student achievement seems to result initially from the report, *A Nation At Risk* (Bell, 1983). In his report, Secretary of Education Bell highlighted what he determined to be a key weakness in American education. He found that compared to children in other industrialized nations, American children spend less time on schoolwork and use the time they do have unwisely. Twenty-four years later schools and school systems are addressing the need for more time spent on schoolwork, though not all address the effective use of this extended time. One problem involved in determining the effectiveness of programs is a lack of consistency in what gauge is used to measure effectiveness.

A review of the literature on after-school instructional programs illustrates the variety of programs, delivery models, and program assessments traditionally offered and those recently employed by school systems to meet the SES requirement of the No Child Left Behind Act. Large scale before and after-school program implementation and research on their effectiveness as supported by empirical data are relatively recent research components. What seems most apparent in the literature is that programs are as varied as the districts and institutions creating and employing them (Kane, 2004). An equal amount of variance is found in how the effectiveness of a program is determined.

In Georgia, students are tested annually using the Georgia Criterion Referenced Competency Test (CRCT). This test is designed to assess mastery of knowledge and skills in the state curriculum (*Online Assessment, 2005*). Annual Measurable Objectives (AMOs) are established in the federal law and used to determine Adequate Yearly

Progress for schools and systems. Schools failing to meet the AMO, or at least show acceptable improvement, must offer supplemental educational services to select students in those schools, in addition to other required interventions. After-school programming is one of the more common delivery models for these services.

The supplemental programs are supported with federal and state funds, but most regulatory rules are made at the local level. Teacher selection and attendance requirements are two of those local decisions. As policy makers and tax payers are demanding more accountability for the expensive venture known as public education, it seems reasonable that leaders of the SES programs be required to provide empirical evidence of the programs' effectiveness. While some research provides evidence to support the use of extended learning time to improve achievement, other reports indicate that such evidence does not exist or is, at best, inconclusive.

Problem Statement

The purpose of this study was to determine if participants in two after-school programs in a Richmond County, Georgia "Needs Improvement" middle school demonstrated improved achievement on the CRCT in higher percentages than non-participants. Additionally, this study attempted to determine if participants' frequency of attendance in these programs impacted achievement.

Research Questions

1. What impact on achievement will voluntary attendance in the After-School Academy have for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

2. What impact on achievement will voluntary attendance in the Saturday Scholars program have for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

3. Will more frequent attendance in the After-School Academy have a greater impact on achievement than less frequent attendance for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

4. Will more frequent attendance in the Saturday Scholars program have a greater impact on achievement than less frequent attendance for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

Null Hypothesis

1. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who voluntarily participate in the school's After-School Academy and similar students who do not participate in this program.
2. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who voluntarily participate in the school's Saturday Scholars program and similar students who do not participate in this program.

3. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who frequently attend the school's After-School Academy and similar students who attend this program less frequently.
4. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who frequently attend the school's Saturday Scholars program and similar students who attend this program less frequently.

Significance of the Study

Implications

Much controversy exists over the impact of extended day/week instructional time, especially as it relates to improved academic achievement. The No Child Left Behind Act of 2001 requires that schools in "Needs Improvement" status offer supplemental educational service to their students. Federal and state funding supports these programs, which are implemented solely on the basis of annual assessment data showing that a school is not making "Adequate Yearly Progress." The programs are loosely regulated and evaluated at the federal level. It seems imperative that schools determine the effectiveness of such programs. Since the programs are created and continued because of low performance on the annual assessment, success of the programs should be measured by improvement on that same assessment. This study demonstrates the impact of two

different programs on the academic achievement of program participants and the effect of attendance patterns on the participants' improved achievement.

Applications

If the programs under study are found to have significant impact on the achievement levels of participants, the school and school system may investigate ways to include all low performing students in such programs. If the extended learning opportunities are shown to have little impact on the participants' achievement level, the school or system may want to explore other uses of those funds and implement programs found to improve student achievement. If attendance is shown to impact the degree of improved achievement, the program directors may consider adding attendance requirements and procedures for increasing attendance. Lastly, if one after-school program is found to have a greater impact than the other, the school or school system may initiate further study to determine what separates the two programs in terms of effectiveness.

Overview of the Methodology

The basis of this study was a comparison of improvement on the Georgia Criterion Referenced Competency Test for participants in two after-school programs and nonparticipants, with consideration given to the frequency of participants' attendance in those programs. The programs considered were the After-School Academy and Saturday Scholars. Participation was voluntary, though the programs did have qualification rules.

After-School Academy applications were offered initially to special education students only, with priority given to those who previously failed the CRCT. When space

was still available following the initial offering, applications were then given to any student who failed to meet standard on the 2006 CRCT. Students were recruited for the program in an effort to fill all 75 slots for which this program was funded.

Saturday Scholars applications were offered only to students who failed to meet standard on the 2006 CRCT and who were receiving free or reduced price school lunch. The free/reduced lunch requirement was nonnegotiable, as this was a Title 1 funded program that included such a qualifying clause. When some of the 125 slots for this program were still available, however, the offer was made to some free/reduced lunch students who met the standard on the 2006 CRCT. Some participants in this program also attended the After-School Academy, as they met requirements for both programs.

The 2006 and 2007 Georgia Criterion Referenced Competency Tests were the instruments used to measure achievement for pre and posttest comparisons. Statistical procedures have been followed to document validity and reliability for the purpose of measuring student achievement. The test follows a selected-response format and is used to assess achievement of the state's core content standards. Scaled scores and Performance Levels are reported. The percentage of students performing at Level 1 is the basis for determining a school's AYP status and determining which additional services a school must offer. Performance levels were, therefore, used to report gains in achievement for purposes of this study.

Participants scaled scores were sorted into groups based on the program(s) attended. Comparison groups were established by pairing each participant's scaled score with the closest possible nonparticipant's scaled score. Efforts were made to pair scores of similar students with consideration to race, gender, economic status, and status as a

“special needs” student. Scores were again divided into groups based on regularity of attendance in each program. Participants’ names were then replaced with numbers (1, 2, 3, ...) and scaled scores replaced with Performance Levels.

Comparisons were then made between participant and nonparticipant groups. For instance, the number and percentage of students regularly participating in After-School Academy that moved from Level 1 to Level 2 or 3 were compared to the number and percentage of paired nonparticipants moving to higher levels. Results are reported in four groups of paired scores, divided by program of participation and frequency of attendance. The four groups include frequent participants (present 30 or more of the days) in After-school Academy, infrequent participants (present fewer than 30 days) in After-school Academy, frequent participants (half or more days) in Saturday Scholars, and infrequent participants (fewer than half of the days) in Saturday Scholars. Performance Levels of participants in each group are compared with matched nonparticipants. The percentage of each group maintaining standard or moving from not meeting standard to meeting or exceeding standard is compared. Independent sample *t* tests of the participants’ and nonparticipants’ posttest scores were also conducted to assess the statistical reliability of any differences in scores.

Operational Definitions

Adequate Yearly Progress (AYP)- a federally defined term as part of the No Child Left

Behind Act. AYP is determined by the percentage of students in a school that meet or exceed the standard on the annual state assessment. The law also allows for schools to demonstrate *adequate progress* without meeting the established percentage of students meeting standard.

After-School Academy- an after-school program offered 3-5 days immediately following the school day. Participation is voluntary, but participants must meet guidelines established by the director to participate.

Annual Measurable Objective (AMO)- a given percentage, established by the federal NCLB law, of the school's student scores that must "meet" or "exceed" the standard for each content area indicated. Schools failing to demonstrate the established percentage of students meeting or exceeding in math and reading/language arts are deemed not making Adequate Yearly Progress.

Georgia Criterion References Competency Test (CRCT)- the annual assessment given to all Georgia public school students to determine the level of mastery of the state's curriculum.

Improved Achievement- improvement on the annual state assessment, measured in this study by Performance Levels on the Georgia Criterion Referenced Competency Test.

Needs Improvement- as applied in this paper means a school or system failing to demonstrate Adequate Yearly Progress for 3 or more years as defined by No Child Left Behind.

No Child Left Behind Act (NCLB)- a federal law designed to improve schools through more local control, more parental involvement and choice, and increased accountability for student achievement at the local level.

Performance Level- the score reporting method used to determine what students "pass" or meet the standard as measured by the Georgia CRCT. Level 1 students did not

pass, Level 2 passed, and Level 3 students exceeded the standard. Percentage of students at Levels 2 and 3 is used to determine Adequate Yearly Progress for a school and system.

Saturday Scholars- an after-school program offered on Saturday mornings. Participation in the program is voluntary. Students must meet established guidelines to participate in the program.

Supplemental Educational Service (SES)- additional instructional time in after-school programs offered to students in schools failing to demonstrate Adequate Yearly Progress four or more consecutive years as part of the No Child Left Behind Act.

Chapter 2

Review of Related Literature

The Issue of Time

In 1983 Secretary of Education Terrell H. Bell issued a report to the National Commission of Excellence in Education entitled *A Nation At Risk* (Bell, 1983). In the report Bell stated that compared to children in other countries, American children spend less time on schoolwork. He reported that not only do children spend less time in school, but also time spent in class and on homework is often used ineffectively. He cited this as a key weakness in American education and declared that the “rising tide of mediocrity” is threatening the foundations of our society. The report contained five principal recommendations, with increased instructional time being one. Increased time is the one recommendation that has received the least federal level action and funding (Farbman & Kaplan, 2005).

Data from Organization For Economic Co-Operation And Development (2005) illustrates that American children are spending less time involved in instruction than the six hours per day Bell reported in 1983. This report states that children in Australia spend about 24 hours a week in class instruction and have almost another 6 hours of homework. France is similar, with almost 25 hours in class instruction and another 7 hours of homework each week. Japanese children are reported to spend about 24 hours in class and have almost 4 hours of homework per week. In the United States children spend about 22 hours a week on classroom instruction with less than 6 hours of homework. Of the 12 countries in the report, the United States has the least amount of

weekly instructional time. Only 3 of the countries report less homework time. Secretary of Education Bell's efforts in 1983 were aimed, in part, at increasing instructional time. This report from Organization For Economic Co-Operation And Development some twenty years later indicates America has failed to meet that expectation.

In 1963, years before Bell's alarming report, educational psychologist John Carroll developed an equation to express the important relationship between time and learning:

$$\text{Degree of Learning} = \frac{\text{Time Spent}}{\text{Time Needed}}$$

Accordingly, the closer students come to having the time they need to learn, the more they should be able to learn (Farbman & Kaplan, 2005). Allotted time can be divided into three categories: (a) allocated time for the school day, (b) time allotted for academic subjects, known as engaged time, and (c) time students and teachers spend truly focusing on learning (Zimmerman, 1998). Following Carroll's equation, as educators increase the time spent focusing on learning, or engaged time, student achievement will increase (Rangel, 2007).

The debate over extending school time continues. While currently no organized federal plan for extending instructional time exists, many states and local school systems are attempting to increase time for learning. The plans fall into two basic categories: extended school day and extended school year. Kirsten Miller, David Snow, and Patricia Lauer (2004) found that *when* the extended time was offered was not important, but the programs should offer at least 45 hours of additional learning time in order to improve academic achievement. Proponents for both strategies, extended day and extended year,

tend to argue that more school time should specifically benefit disadvantaged children (Evans & Bechtel, 1997).

Schools, school systems, and communities eager to demonstrate improved student achievement are funding programs to extend student learning time. Blane Worthen and Stephen Zstray (1994) reviewed twenty years of research on extended learning time. They found little evidence to suggest that merely increasing school time leads to academic gains. Evans and Bechtel suggested that time is a *necessary* condition for student achievement, but simply adding more time is not a *sufficient* condition for improving achievement. Bill Metzker (2003) and Elena Silva (2007) reviewed extended day programs and found time, by itself, made little difference in student achievement. They suggested that improving how schools use the time they already have would do more to improve achievement. Metzker determined that professional development for teachers who need help with classroom and time management might be a better use of the after-school funds.

In *Critical Hours* (2003), Beth Miller seems to concur. Miller stated that the goal of the extended day programs is clear: increased student performance. She found that many of the available programs were merely an extension of the school day. For most disadvantaged students, “more school” is not showing gains in student achievement.

A Massachusetts 2020 report, *Time for a Change* (Farbman & Kaplan, 2005), argues that while time itself may not solve our educational crisis, increased instructional time is essential for some students. Modern academic expectations require that students know far more than in previous decades. The conventional school day and calendar, according to Farbman and Kaplan, are not adequate to meet the needs of learners. Jimmy

Byrd (2001) found that some students benefit from a shortened school year, providing additional specific, targeted instruction for struggling learners during intercessions.

Need for Extended Time

Meeting the needs of students is, after all, the most challenging but primary focus of schools. Researchers are increasingly investigating why some schools fail to meet those needs. *The Teacher Survey of Standards-Based Instruction: Addressing Time* (Florian, 1999) illustrates why more time for instruction may be absolutely necessary for student success. In this report teachers shared concern over the breadth of content addressed in many standards. The survey required that teachers determine the estimated time required to teach a given set of standards. Teachers from four school districts completed the survey, collectively reviewing hundreds of standards across four grade levels. Results suggested that over 1000 hours of instructional time is needed yearly to teach the designated standards. Since schools are reported to have 630-905 instructional hours available each year, the report found an obvious mismatch. In eighth grade, the teachers reported needing 1422 hours to teach the standards; about 500 more hours than they had. Farbman and Kaplan (2005) supported the idea of a time mismatch, stating that extended time would allow teachers to address the deeper issues involved in the content taught and reduce the rushed nature of the instructional class period.

Funding Extended Time

Without empirical evidence that extending learning time increases student achievement, some policy makers question the value of the additional expense. As essential as additional instructional time may be, extending instructional time is a costly venture. A thirty-day extension to the middle school year in California was estimated to

cost an additional \$100 million (Metzker, 2003). Metzker offered a possible solution. He suggested that when extended time is the strategy used to improve achievement, leaders may not find it feasible to fund such programs for all students. Services should be allocated to targeted schools or students with the greatest need. His report cites plans from Kentucky, Minnesota, and Texas where interventions are targeted to those students deemed disadvantaged and low achieving. Massachusetts is reported to offer extended time only to students needing assistance with state assessments. States and school systems that want to support extended instructional time may consider funding such programs for targeted populations.

Researchers tend to agree that even when funding is available, just adding more hours to the school day or more days to the school year will not bring desired increases in student achievement. What separates successful programs from unsuccessful programs seems to lie in what the programs offer, how those services are delivered, and what tools are used to determine success.

Variety In After-School Programs

A common delivery model for extending learning time is the after-school program. These after-school programs are frequently funded through grants sent directly to the school or community group organizing the programs. That means the program components are locally determined. Since research of instructional after-school opportunities is relatively new, definitive answers on the best way to run the programs can be difficult to find. The United States Conference of Mayors report, *After-School Programs* (2003), illustrates the variety of programming offered across the country. Ninety-four cities participated in a survey helping to establish baseline data on programs.

Over half of the programs (56%) were operated by public schools. The results reflect that about the same percentage (57%) offered academic assistance. About half of the programs (53%) offered arts and craft activities. Fewer than half (48%) provided homework assistance. Games and music combined (46%) were offered almost as often as help with homework. The most consistency (84%) was found in programs providing snacks. Over half (55%) of the responding cities had programs offering after-school and weekend activities. A few programs offered leadership development (11%), science education (12%), or a community service component (14%). There was no effort, as part of this particular survey, to determine the effectiveness of these programs. Karen Clark-Keys (2007) found more funding agencies beginning to ask how students are being served by these programs and what assessment tools are available to measure the effectiveness of the services provided.

Most research addresses what *is* offered in after-school programs. Conversely, Miller (2003) studied what *should* be included in successful programs based on research on the nature of early adolescents and what program components would best meet the needs of this “fragile” developmental age group. Miller suggested effective programs focus on arts education, tutoring, mentoring, project-based learning, and experiential education. Miller emphasized that research shows the *type* of program offered is less important than the environment in which the program is offered.

Current literature reports many successes of after-school programs. Extra instructional time allows teachers to differentiate lessons and meet individual student needs and allows for deepened adult-child relationships (Farbman & Kaplan, 2005). Priscilla Pardini (2001) found schools reporting better student work habits and

interpersonal skills. Additionally, improved school attendance, increased graduation rates, and improved school grades are also reported as benefits of after-school programs (Brown, Frates, Rudge, & Tradewell, 2002, Farbman & Kaplan, 2005, Pardini, 2001).

Many of the programs reviewed in the literature failed to establish a plan for monitoring student progress on assessments. Pardini (2001) found that research in this field is often wrought with flaws in methodology, providing inconsistent and inconclusive findings.

Schools implementing after-school programs as a support for struggling students would benefit greatly from longitudinal studies illustrating the precise components of programs that would bring the greatest benefits in student achievement. Since the interest in structured academic after-school programs is relatively modern, long-term effect studies with empirical evidence of success can be difficult to find. Waiting for more empirical data is not an option for many schools and systems. The No Child Left Behind Act of 2001 requires that schools failing to show adequate progress offer such extended day opportunities immediately.

The No Child Left Behind Act

The No Child Left Behind Act (NCLB) of 2001 sets a continuum of increasing expectations for student achievement (*No Child Left Behind*, 2005). This act requires that all students meet rigorous standards of learning by 2014. Student achievement is measured annually with state adopted assessments. The Act establishes benchmark levels of performance, stating what percentage of each school and school system must show mastery on the assessment of standards each year to demonstrate Adequate Yearly

Progress. The target is that by 2014 all American students will demonstrate mastery on those annual assessments.

Schools that meet or exceed the targeted percentage of students demonstrating mastery of the standards are said to be making Adequate Yearly Progress (AYP). Schools failing to make AYP by meeting established benchmarks are afforded other opportunities to show adequate improvement. After further evaluation, any school still failing to demonstrate acceptable improvement moves into a category of Needs Improvement (NI). A school failing to demonstrate AYP for one or two years is referred to as “not making AYP” but does not have to offer any additional services. Additionally, a school failing to demonstrate AYP for four or more consecutive years is required to offer a variety of other services. Table 1 shows the progression of services required when a school does not make AYP (United States Department of Education, Feb. 2007).

NCLB Supplemental Educational Services

Supplemental education services (SES) are opportunities for additional academic instruction designed to increase student achievement. These services are required offerings in schools that have not met state targets for more than 3 years. Supplemental services may include tutoring and/or after-school instructional programs. They may be offered by public or private schools and by nonprofit or for-profit organizations. The federal laws governing the No Child Left Behind Act require specific actions from the state and local education agencies.

State education responsibilities include determining which schools are failing to make AYP, monitoring effectiveness and quality of service from these providers, and

Table 1

Requirements for Failing to Make Adequate Yearly Progress

Years	School status	Required services
1	Does not make AYP	None required
2	Does not make AYP	None required
3	1st year of school improvement	Technical assistance Public school choice
4	2nd year of school improvement	Technical assistance Public school choice Supplemental educational services

removing from the list any SES provider failing to demonstrate improved student achievement for two consecutive years. Local education agencies are to determine which students are eligible for services, prioritizing that list of students based on prior poor achievement, notifying parents at least annually of services available, and providing information to the state agency as necessary for monitoring success of programs. States and local education agencies must cooperatively provide parents a list of outside providers and allow parents to choose the provider they wish for their child. Students from low-income families attending Title 1 schools that are not making AYP are eligible for these services. SES providers must report to schools and parents on the progress of participants (US Department of Education, Feb. 2007).

According to the guidelines, a school system is to “remove from the list any provider that fails for two consecutive years to contribute to increased student proficiency relative to state academic content and achievement standards.” Since most states have now adopted annual assessments to measure mastery of standards, the tool for measuring a provider’s effectiveness is clear. State and local education agencies are to determine what “contribution” to improved student achievement will be acceptable. Educational leaders must decide, for instance, if half of the attending students showing improvement on the year-end assessment is a sufficient contribution for a provider to continue offering supplemental services to students.

Schools may also offer programs at the school site by extending the school day or week. School or system leaders then become responsible for designing the programs. Program directors may choose to rely on evidence of other successful programs to determine what components are essential for improved student achievement.

Programs Showing Improved Achievement

Because of the No Child Left Behind Act, many school systems must now offer extended learning opportunities specifically to at-risk, low income, and underachieving students. Evidence of program effectiveness, according to NCLB guidelines, is improved student proficiency on state standards. Hence, some current studies available on the topic tend to address specifically the achievement gains of participants as demonstrated by annual state assessments. The programs in these particular studies were funded to support improved achievement in “Needs Improvement” schools. While some of the programs offered benefits beyond academics, the intent of each program was to improve

student performance on the annual state assessment. Some research exists to support the success of such programs.

In some cases a direct comparison between very similar schools was possible. In New York City, the Schools Under Registration Review (SURR) that offered extended learning time surpassed gains made by SURR schools not offering extended time (New York City Board of Education, 2000). University Park and Kipp Academy are two schools in the report that offered extended learning time. Students at these schools generally outperformed students of similar socioeconomic status at other schools in their districts. Kipp Academy exceeded the Bronx District and New York City Schools' average performance on the California Test Bureau/McGraw Hill New York State Test (CBT) in math and reading at all reported grades. University Park outperformed other high schools in the district on the Massachusetts Comprehensive Assessment System (MCAS) four years in a row. In 2002 all students at University Park in the Worcester School District achieved proficiency on the MCAS in English/language arts, while only 18% of the Worcester District demonstrated proficiency. In 2003 University Park students reached 100% proficiency in math when the district average was 28% proficient. University Park was no longer offering the extended instructional program when some of these scores were posted. Teachers argued that students were still benefiting from the additional assistance they received when they did participate in extended learning programs (Farbman & Kaplan, 2005).

A four-year study of LA's BEST, an extended learning program, shows increases in school attendance, as well as increased student achievement on standardized test scores (Huang, Gribbons, Kim, Lee, & Baker, 2000). Huang et al. also noted that those students

with the best attendance to the after-school program showed the most gains on the annual assessment. *Time For Change* (Farbman & Kaplan, 2005) reports on Boston schools where more than half the student body qualified for free or reduced lunch. The four schools in this report offering extended school days outperformed district averages in almost every area. Roxbury Prep, one of the schools studied, reached 89% proficiency in English/language arts when the district average was 42%. Murphy, another school in the study, achieved 48% proficiency in math compared to the district average of 15%. In the eight comparisons made in this study only one showed an extended learning time school falling below the district average, and that was by 1% in math.

Similarly, a meta-analysis of 56 studies shows that after-school programs can have statistically significant effects on student achievement. It should be noted this analysis considered 371 studies of out-of-school time strategies dating back to 1984 and found only 56 reporting documented positive impacts on student achievement. The study found the successful programs to have an average effect size of 0.26 standard deviations on elementary reading scores, which equals a gain of 10 percentile points. High school after-school participants demonstrated a gain of 17 percentile points, an effect size of 0.44. The largest gain was found in reading strategies that used one-on-one tutoring techniques, with an average gain of 19 points or an effect size of 0.50 (Lauer et al., 2003).

These studies seem to indicate that after-school programs can have a definite positive impact on student achievement. Evidence of even small gains on annual assessments can be promising for schools struggling to demonstrate Adequate Yearly Progress. Conversely, other research reveals that many after-school programs were unable to demonstrate such improvement in student performance.

Programs Failing to Show Improved Achievement

Many after-school programs reviewed discussed benefits for participants, some with empirical evidence to support the claims. Those benefits frequently included social and emotional gains, improved self-efficacy reports, and advantages for the community. It is a relatively new trend to require empirical evidence of improved student achievement as measured by assessments addressing state standards of learning. Using annual state assessments to measure improvement left some programs unable to support their claims of success.

Nancy Adelman (1996) reviewed 12 studies of after-school programs from 1991-1995. She found some evidence of an impact on student learning, in that a fraction of the sites were able to document improved student achievement. No single reform or initiative proved successful at all sites. Adelman reports that Boston added the equivalent of 36 days to the school year, and Boston school children demonstrated a dramatic increase in achievement. New Orleans, however, experimented with this same Japanese-length school year at two schools. The experiment proved unsuccessful, and schools were unable to show any gains in achievement. Adelman cautions that school leaders and program directors must consider how the time is used as seriously as they consider how much time to add. In a review of 32 studies from 1977-1992, Robert Worthen and Stephen Zstray (1994) found similarly mixed results. Some programs using the extended year concept were able to demonstrate statistically significant higher performance for participants. However, the compilation of results indicates that students in extended year programs performed about the same as students in schools following traditional school year schedules. Extending the school day or school year may bring non-instructional

benefits, but little research-based evidence exists to support the notion that increased time will consistently yield improved student achievement (Evans & Bechtel, 1997).

Some after-school programs are designed to follow the instructional program used during the regularly scheduled school day. According to Miller (2003), proponents of this school/after-school collaboration model assume the program will be more effective if the after-school program supports the regular school day's instructional plan. Miller found only limited evidence that such a link produced achievement gains. Some reports in Miller's study suggested that such a link might prove dangerous. After-school programs attempting to look too much like "regular school day" cause students to miss opportunities for the variety of essential learnings not offered during the school day. Robert Halpern (2000) reports similar findings in his study of after-school programs for low-income children. Proponents of after-school programs frequently claimed that these programs should purposefully avoid regular school day activities. Halpern was unable, however, to find evidence showing that programs diverging from traditional academic instruction produced the desired gains in student achievement with any degree of regularity.

Other research findings demonstrated equally disappointing achievement results from extended day programs. Thomas Kane (2004) reviewed the findings from four large supplemental programs: 21st Century Community Learning Centers (21st CCLC), The After-School Corporation (TASC), Extended-Service Schools Initiative (ESS), and San Francisco Beacons Initiative (SFBI). The TASC evaluation showed no impact on math or reading test scores following one year of participation, but did indicate achievement gains in math for second and third year participants. The SFBI program

seemed to have no impact on grades, test scores, or school attendance. Evaluation of results from the 21st CCLC program failed to demonstrate any improvement in reading for participants and did not use follow-up reporting in math. The ESS program focused only on participation and cost, so no data was available to study the impact on student achievement. Kane concluded that these programs did show relatively consistent growth in the areas of parental involvement, student engagement, and homework habits. He was unable to document a statistically significant impact on achievement test scores following one year in any of the programs. Other research on almost 2000 middle school students participating in 21st Century programs found no significant academic impact from program participation and found regular attendance in the programs produced no better academic result than poor attendance (*Dynarski et al., 2003*).

Rodney Roukema (2005) studied the North Carolina End of Grade scores for middle school students who participated in Support Our Students After-School Program for three years. He found that students who participated in the program throughout their middle grade years showed no significant difference in math and reading scores from those who did not participate. He sorted the groups to determine the difference for various influencing factors and found no significant impact from the Support Our Students After-School Program with regard to race, gender, or economic status. This study tends to indicate even three years in an after-school program may provide no true impact on student achievement.

After-school program supporters offer a variety of reasons to continue the programs. Reduced crime, improved parent/school communication, enhanced social and interpersonal skills, and age-appropriate engaging activities are central to the theme of

arguments favoring the programs. When empirical evidence of program effectiveness is measured by student gains on criterion-referenced tests, however, some researchers find a shortage of data to support such programs (Adelman, 1996, Evans & Bechtel, 1997, Miller, 2003, Worthen & Zstray, 1994).

Georgia After-School Programs

The Georgia Department of Education evaluates private providers of after-school programs and supplemental educational services annually to determine program effectiveness (Harnish, Thompson, Pollack, Cramer, & Alagoz, 2006). The evaluation consists of surveys and pre and posttest comparisons. To determine the effectiveness of the programs, Georgia Criterion Referenced Competency Test (CRCT) scores of students receiving supplemental educational services are matched with nonparticipants in the same school and grade who scored similarly on the previous CRCT. The majority of the service providers for the 2005-2006 school year report that about half of their students outperformed matched nonparticipants (GA Dept. of Ed. Title 1, 2006). Table 2 illustrates a comparison of SES participant and nonparticipant scores for the 111 private SES providers reporting results from the Spring 2006 Georgia state assessment. This report only includes those SES participants for whom the provider reported results and was able to provide the student identification number.

This table shows that less than half of the SES students exceeded the scores of matched students who did not participate in supplemental programs. More than half did only as well or worse than students who did not receive the additional instructional support. Seven percent of the SES providers had no students outscoring the nonparticipants. Seventy-three percent had fewer than half exceeding the reading score

Table 2

Percentage of Georgia's 9807 SES Participants Who Outperformed Nonparticipants

<u>Content</u>	<u>% Outperforming</u>
Math	46.3%
Reading	44.7%
ELA	44.7%

of the paired nonparticipants. The report also shows that size of the SES provider and grade level disaggregation yielded no noticeable differences. Most interestingly, second graders participating in supplemental services, as a group, did “much worse” than matched nonparticipants. While some providers certainly demonstrated gains for students who participated in the program, collectively the evidence did not show the desired improvement in student achievement as measured by gains on the CRCT.

Measuring Student Achievement

Much of the research concerning after-school programs adds little to the knowledge base of what makes an effective program as it relates to improved achievement. Many of the studies are flawed by poor measurement techniques, lack of control groups, or a failure to test for statistical significance (Weiss, 2005). Any effort to measure student achievement must include selection of the appropriate assessment tool. The No Child Left Behind Act requires that states establish standards and design tests to measure progress on those standards. Criterion-referenced and norm-referenced tests are the two assessment formats traditionally used by schools to track student achievement.

Criterion-referenced tests (CRT) are designed to measure how well a student has learned the material taught. CRTs do not compare one student to another and no bell curve is used to rank students (*Weaving Gender Equity into Math Reform*, 2001). Everyone could, theoretically, answer 100% of the test questions correctly and all would score a 100. Students are measured only against identified standards of achievement without regard to how others in the same group perform on the test (Dunn, Parry, & Morgan, 2002). Such a test is deemed a more accurate measure of student achievement than traditional norm-referenced tests (Fair Test, 2007).

Norm-referenced tests are known as traditional and “rather antiquated notions of academic rigor” (Dunn, Parry, & Morgan, 2002). Questions on norm-referenced tests may ask about information in the curriculum taught that year, but such criteria for the questions is not a requirement. CRT questions always address content knowledge and skills assigned to the particular grade level being assessed. Most students completing a course could pass the CRT if they were taught well, studied adequately, and the test was created to align to the expected curriculum (Fair Test, 2007). Therefore, many states use CRTs as pre and posttest measures for tracking student progress.

Scores for criterion-referenced tests are reported in a variety of ways. Traditionally, raw scores and scaled scores reveal how many questions were answered correctly (raw score) and where that score fell within a range of scores in that specific content area and grade level (scaled score). A recent variation to score reporting also includes *performance levels*. Criterion-referenced tests in many states now test what students know and are able to do in different subjects based on content standards, and performance standards address how much of the content standard students must be able

to demonstrate to show mastery (Fair Test, 2007). These performance standards lend themselves to *levels of performance*. Score reports may define the performance levels as basic, proficient, and advanced.

Norm-referenced and criterion-referenced tests each have an important, though different, role in assessment. High-stakes tests are playing an increasingly important role in education. When standardized tests must be used, criterion-referenced tests may be far better tools for determining achievement gains than norm-referenced (Fair Test, 2007). The No Child Left Behind Act requires that states create and use such tests annually to measure academic progress of students and schools.

Georgia Norm and Criterion Referenced Tests

In Georgia, norm-referenced testing is mandated by state law for grades three, five, and eight. The purpose of this testing, according to the Georgia Department of Education, is to compare performance of Georgia students with a national sample (GA Dept. of Ed. NRT Guide, 2007). Additionally, all students, kindergarten through eighth grade, take a criterion-referenced test each year. This annual assessment is designed to measure individual progress.

Georgia is moving from its Quality Core Curriculum (QCC) to Georgia Performance Standards (GPS). The “rollout” of standards began with English/language arts in all tested grades, 1-8. The implementation of new performance standards begins with a year of teacher training. As the second year of a rollout is ending, students are tested on the standards.

Performance standards for core content areas in grades 1-8 are now in place. Mastery of the standards is assessed each year with the Georgia Criterion Referenced

Competency Test (CRCT). The CRCT was implemented in the spring of 2000, with end-of-year testing in reading, English/language arts, and math in grades four, six, and eight. Science and social studies assessments in grades three through eight were added in the spring of 2002. The spring 2006 CRCT was based on the Georgia Performance Standards for certain grades and subjects, as determined by the GPS rollout schedule.

Georgia law now requires students be assessed annually with the CRCT in grades one through eight in reading, English/language arts, and math. In grades three through eight, science and social studies are also required. Performance on this assessment is used to help determine the appropriateness of promotion from 3rd, 5th, and 8th grades. This assessment is also used to determine if a school is making Adequate Yearly Progress under the rules of the No Child Left Behind Act.

The Georgia CRCT is used to identify instructional strengths and weaknesses. The intended result is to improve instruction, increase student achievement, and to gauge the quality of education throughout Georgia (Hunt & Millicans, 2005, Millicans, 2004). The test is also used to determine promotion to key grades and assess an individual's progress from grade to grade.

Each content area test has 2 sections, with a short break given in between. Multiple forms of each test are given, with field items included for all grades and contents. Scripted examiners' manuals are used during test administration.

Systems are allowed to select an 8-day test window within an established time frame. Students who miss a section of the test are only allowed to make it up during the designated test window. Some scores, for that reason, may be absent from the test report.

Retests are given only for students who perform at Level 1 (Does Not Meet Standard) in a promotion/retention content area and grade level. Promotion/retention grade levels are those grades in which the CRCT is used as a consideration for promotion to the next grade. Those are currently 3rd, 5th, and 8th grades. Promotion/retention content areas are currently math and a combination of reading/English language arts. Students not meeting the standard in other grade levels or content areas are not retested that same school year.

Individual student reports are sent with a variety of scores, as are school-wide reports. Interpretation of the results is important, so that current strengths and areas in need of improvement can be determined.

Georgia CRCT Score Interpretation

According to the Georgia Department of Education's 2007 CRCT Interpretation Guide, score reports include raw scores, scaled scores, and Performance Levels. A raw score indicates how many questions the student answered correctly. Scaled scores are developed from the raw scores using mathematical procedures. The number of correct responses (raw score) is converted to the CRCT scale. Scales run from 150-450 for each content area still following the Georgia Quality Core Curriculum (QCC). Most middle grades subjects have moved to the Georgia Performance Standards (GPS) and have scaled scores in the range of 650-900. Table 3 shows which subjects/grades have performance standards. For QCC courses, scores below 300 indicate performance that Does Not Meet The Standard. Scores in the 300-349 range are said to Meet The Standard. Scaled scores at or above 350 represent performance that Exceeds The Standard. For GPS courses, scores below 800 Do Not Meet the Standard, scores 800-849

Meet the Standard, and scores 850 and above Exceed the Standard (GA Dept. of Ed. CRCT Score Interpretation Guide, 2007). Since some courses are QCC and others follow the GPS, scaled scores cannot be fairly compared across content areas. Additionally, since standards vary in difficulty across grades and content areas, the scaled scores cannot be accurately compared to other grades or subjects. Scaled scores, like domain scores, should only be used to help identify a student’s areas of strength and those areas in need of more instructional support.

Table 3

<u>Score Ranges for Georgia's Criterion Referenced Competency Test by Subjects and Grades</u>	
<u>QCC Courses</u> 150-450	<u>GPS Courses</u> 650-900
Mathematics grades 3, 4, 5, 8	Mathematics grades 1, 2, 6, 7
Science grade 8	Science grades 3, 4, 5, 6, 7
Social Studies grades 3, 4, 5, 6, 7, 8	Reading grades 1, 2, 3, 4, 5, 6, 7, 8
	English/Language Arts grades 1, 2, 3, 4, 5, 6, 7, 8

Performance Levels, unlike raw or scaled scores, are comparable across all content areas and grade levels. Performance Levels are, therefore, used to compare student achievement with students in other grades, content areas, and schools across the system and state. Because of the comparability from grade to grade, schools and systems also use the Performance Level reports to demonstrate Adequate Yearly Progress as required by the No Child Left Behind Act.

Some researchers find inherent problems with all high stakes testing designed to measure a student's mastery of standards (Fair Test, 2007). Law, nonetheless, requires states to design and implement an annual assessment to measure student achievement of rigorous standards. Georgia, like many other states, has chosen a criterion-referenced test to measure mastery of newly developed performance standards. In Georgia, the CRCT has become the yardstick by which academic achievement is measured. The validity and reliability of the test is essential to its success as an appropriate tool.

Validity of Georgia CRCT

According to the APA OnLine report, *Psychology Matters*, validity is “the extent to which a test measures what it is intended to measure (Gerrig & Zimbardo, 2002). Shirley Millicans, Georgia DOE Assessment Specialists (2007), provided information pertaining to the validity and reliability of the CRCT. This report states that the primary purpose of the CRCT is to evaluate school Adequate Yearly Progress. Validity of the Georgia CRCT must be defined based on this purpose.

Most of the 2006 CRCT was constructed using field-tested questions from the 2005 assessment. Statistical characteristics were monitored to ensure that the statistical properties of the 2006 CRCT were similar to previous tests. Since Georgia is moving from the Quality Core Curriculum to Georgia Performance Standards, special attention was given to the questions' applicability to the new standards.

Riverside Publishing Company was selected to write the CRCT. Company specialists first had to determine the comparability of QCC and GPS assessment items. Each QCC standard was matched with a similar GPS. When a QCC or GPS was found to have no comparable match, the content item was labeled “No Match.” Riverside Test

Development Specialists then reviewed over 10,000 former test items and aligned them with GPS in the appropriate grade and content area. The reviewed items were from previous tests and from the Online Assessment System used as classroom practice assessments. After reviewing all QCC test items, Riverside Test Development Specialists were able to successfully match 82% of those with a GPS. The remaining 18% were rejected as possible test items.

Following the review of previous test items, the Test Development Specialists (TDS) devised a plan for creating the newest CRCT. Item writers were screened to determine academic background, experience in education, and ability to write quality test items. Those selected were both newly trained and experienced test writers. Selected writers were then divided into five groups based on content expertise. The TDS then conducted training sessions on content and avoiding bias. Items selected by the writing committee were continuously checked by TDS for clarity, grade appropriateness, artwork as applicable, and how well the test item measured the standard. At least two specialists reviewed each selected test item. A sufficient number of acceptable items were created to form a comprehensive CRCT.

Riverside Publishing then hosted an item review. Georgia educators were recommended by System Test Coordinators and selected by Riverside to participate in the review. Riverside TDS facilitated training for review committee members. Test items were to be evaluated for:

- Overall quality and syntactical clarity
- Content coverage and content appropriateness
- Alignment to the specified GPS

- Grade-appropriate stimuli, with an emphasis on higher order thinking skills
- One clearly correct answer and the appropriate number of relevant and reasonable distractors
- Freedom from bias toward or against any particular group

Difficulty of each test item was determined via analysis of how many test participants selected the correct and each incorrect answer. P-values close to 1.00 indicate many students selected the correct response and the item is very easy. As test items increase in difficulty the p-value moves closer to 0.00. A point-biserial correlation considers how many of the top performers selected the correct answer compared to how many low performers selected that same correct answer. The review committee was charged with reviewing field-tested items previously “flagged” due to their p-values and the point biserial value. Table 4 shows the criteria used for determining those flagged items.

Table 4

Validation Items for Flagging Criteria Based on P-value of Keyed Response

<u>P-value:</u>	<u>Indicates</u>
Keyed response <0.35	Difficult item
Keyed response <0.05 or > 0.95	Extreme item
Keyed response < p-value of distractor	Possible mis-key
Keyed distractor > 0.35	Possible second correct option
If point biserial of coded response < 0.20	Poorly discriminating item

Riverside TDS held a training session for committee members on the information provided by the statistics. Reviewers were also reminded that questions were aligned to the GPS, but some were field-tested on the QCC operational form. The test items may have been valid under GPS, but the students had not yet had instruction in that curriculum. The review committee then considered the following information for all flagged test items:

- Form
- Position
- Item as it appeared in the printed books
- Passage for reading sections
- P-value of correct answers
- Percentage of students choosing each response option
- Point biserial
- Differential item functioning using the Mantel-Haenszel procedure

Each field-test item was reviewed and either rejected, accepted, or set aside to be revised. Accepted items were made available for the pool of GPS items for the 2006 CRCT operational form. Rejected items were removed from the pool. Items for revision were those questions the educators felt could be made appropriate with minor reworking. They were put in the group for later revision and more field-testing. Table 5 shows the results of the data review. Of the 810 test items reviewed, 85% were accepted.

Reliability of the Georgia CRCT

Test reliability is the portion of variance in a score that is not due to error. A test is said to be reliable to the degree that an individual makes the same grade on repeated

Table 5

Item Review Results for the 2006 Georgia CRCT

Content Area	Grades	# Of Test Items			
		Reviewed	Accepted	Rejected	Revised
Reading	1-5	156	141	7	8
Reading	6-8	123	91	16	16
ELA	1-5	189	163	23	3
ELA	6-8	135	122	11	2
Math	6	73	60	8	5
Science	6-7	134	117	3	14
<i>TOTAL</i>		810	694	68	48

measurements (Ary, Cheser Jacobs, Razavieh, & Sorensen, 2006). The more likely a person is to get the same or a similar score on repeated tests, the more reliable the test is said to be.

Reliability is reported using two statistical measures for the CRCT. One method uses a reliability coefficient, expressing the ratio of the true score variance to the true score plus error variance or total score variance. A reliability coefficient 1.0 would indicate that all test score variance is true variance and there is no error in the measurement. A coefficient of 0.0 would show no true variance, indicating the measurement is all error (Ary et al., 2006). The closer the reliability coefficient is to 1.0 the less likely it is that random error has influenced the score and the more reliable the measurement is. This reliability coefficient is independent of the measurement scale and is comparable from test to test.

Standard Error of Measurement (SEM) is also determined and reported for the CRCT. This statistical index reports random variability in test score units. A SEM of 10, for instance, indicates a true score probably lies within 10 points of the reported score. A smaller SEM indicates a more reliable score. Table 6 shows the reliability and SEM for CRCT scores in middle school reading, English/language arts, and math. The table shows that reliability ranges from a low of 0.81 (in reading) to a high of 0.91 (in math). The standard error of measurement is fairly consistent, ranging from 9-12.

Table 6

Reliability and Standard Error of Measurement for the Georgia Middle School 2006 CRCT

<u>6th Grade</u>	Reliability	SEM
Reading	0.83	11
English/Language Arts	0.86	9
Math	0.90	9
<u>7th Grade</u>		
Reading	0.85	9
English/Language Arts	0.87	11
Math	0.91	9
<u>8th Grade</u>		
Reading	0.81	9
English/Language Arts	0.86	12
Math	0.89	11

While no set level of acceptable reliability has been established, there are guidelines in the literature among professionals for what is acceptable with given circumstances. Ary, Jacobs, Razavieh, and Sorensen (2006) state that when important

and irreversible decisions are being made based on an assessment, only instruments with the “highest reliability are acceptable.” They say measurement experts believe a reliability of 0.90 is the least that should be considered and a 0.95 should be the target. In *Test Item Analysis & Decision Making* (University of Texas, 2003) a reliability of .90 or above is cited as excellent and “...the level of the best standardized tests.” Reliability .80-.90 is listed as “very good for classroom test.” That being the case, only two of the nine assessments in Table 6 meet the lowest acceptable level for standardized tests and none of the nine meet the ideal reliability of 0.95. Most of the CRCT test reliability ranges fall in the category listed as “good for classroom tests.”

Summary

While the typical school day start and end times vary by state and county guidelines and within grade structures, most students attend school about seven hours a day. Of those seven hours, some time is reserved for lunch, fine arts, recess, homeroom, etc. Students have roughly five or six hours of instructional time each school day. Research indicates only a fraction of that time is spent engaged in true learning.

Educational leaders are now faced with demanding and rigorous standards of learning and laws requiring that all children succeed. It seems that for some students, the time allotted for learning is not sufficient. A school failing to demonstrate Adequate Yearly Progress is required to offer additional learning opportunities to those students labeled disadvantaged and/or Not Meeting the Standard. Since strictly academic after-school programs are relatively new in education, the effectiveness of those extended opportunities has only recently come under scrutiny.

States are required, under the No Child Left Behind Act, to devise an assessment tool to annually measure student achievement. After-school programs and supplemental educational services are sometimes afforded students who need additional instruction to show necessary improvement on the annual assessment. Program directors are to use the assessment results to evaluate the program's impact on achievement with pre and posttest scores.

Some research of such programs exists to support the idea that after-school programming does have a positive impact on achievement. Other reports indicate such programs have no impact on academic gains for the participants. In Georgia, the Criterion Referenced Competency Test (CRCT) is used to determine Adequate Yearly Progress (AYP) for schools and systems and to measure individual student progress. Students are tested each spring in grades 1-2 in reading, English/language arts and math. Grades 3-8 are tested in all content areas. As Georgia moves from its Quality Core Curriculum to Georgia Performance Standards, the CRCT is revised. Scores are reported in a variety of formats, with Performance Levels being the format used to determine what percentage of a school "passes" or meets the standard. Schools not demonstrating AYP are required to offer extended learning opportunities to under-performing students.

The extended learning programs are offered by school systems, nonprofit agencies, and for-profit companies. State and local systems are responsible for determining the effectiveness of the programs as measured by student performance on the CRCT. Data at the state level measuring the effectiveness of public programs indicates inconsistent impact on participants' performance on the CRCT.

Little longitudinal data to support effective program components is available, as the expectation of improved student achievement due to program participation is relatively new. The challenge for program directors seems to be to find components of similar programs helping students improve, to avoid situations and strategies proving to have no effect on student achievement, and to determine how their own programs are impacting participants. Close scrutiny and empirical studies of after-school programs is necessary for improvement in program design and implementation.

Chapter 3

Research Methodology

This study was designed to evaluate the effectiveness of two middle school after-school programs as measured by improved student achievement on an annual assessment. This chapter describes the research context, participants in the study, instruments and procedures used, and data analysis.

Research Context

This study examined the impact of after-school participation in a suburban Title 1 middle school in Richmond County (Augusta), Georgia. At the time of the study, this was the largest middle school in the county. The school offered grades 6-8 and served about 1000 students. Over 95% of the students were minorities, with about 90% being African American students. Eighty-two percent of the students were deemed “economically disadvantaged” and qualified for free or reduced price school lunch. About 145 students were considered “special needs” students with an active IEP, comprising 14% of the student body. About 10% of the students missed more than 15 days of school per year, down from 15% just two years ago. This was an NI8 school, indicating the school had failed to demonstrate Adequate Yearly Progress for the eight years the Georgia Department of Education had tracked such information.

The school had a total of 71 teachers, 12 of whom had positions that were funded for the special needs students. The school was in its third year of inclusion classes, and all but 7 of the special needs students participated in regular education instruction. All

but one of the teachers was considered Highly Qualified as defined by state law. All the core content teachers and special education teachers were Highly Qualified.

Math, science, and English/language arts teachers who volunteered to teach in the after-school programs were paid an hourly rate based on their individual salaries. Teachers were not required to demonstrate their effectiveness as a classroom teacher before being invited to serve as an after-school teacher, though principal discretion was used in offering the positions to teachers. After-School Academy was offered Monday through Wednesday for two hours immediately following the regular school day. This program began in October and continued through March. Saturday Scholars was a weekend program from 9:00 to 1:00 each Saturday, running from October through April. Teachers were allowed and encouraged to follow the curriculum of the regular school day and designed lessons independently. Programs were interrupted for school holidays and special activities from time to time as deemed necessary by program directors or the principal.

Research Participants

Participants in the study were those middle school students who voluntarily participated in either after-school program. Students were offered applications to each program based on specific program criteria.

After-School Academy was designated by the school's principal as a program to serve those students in special education. The program was funded for 75 students and staffed at a ratio of one teacher for every 15 students. When the program failed to attract 75 special education students, the principal allowed other students who had failed the

math and/or reading portions of the 2006 CRCT to participate and finally enrolled students who met the standards on the 2006 CRCT.

Saturday Scholars was a program funded by Title 1 solely for economically disadvantaged students. Initially applications were given to only free or reduced lunch students who failed the math and/or reading portions of the 2006 CRCT. When the program failed to attract the 125 students for which the program was funded and staffed, free/reduced lunch students who had *not* failed the CRCT were allowed to participate.

Students were given the applications for both programs at school during assemblies where only qualifying students were present. A deadline was set for applications to be returned. No data was maintained on how many students returned the application by the deadline. However, teachers later made phone calls to parents and guardians of qualifying students to make sure they were aware of the programs and had seen the applications.

After-School Academy was unable to attract the number of students for which the program was funded. Throughout the year, upon teacher recommendation or as parents requested their children be added to the After-School Academy program, students were allowed to join.

The Saturday Scholars program eventually enrolled the 125 students for which this program was funded. Other students wishing to join were placed on a waiting list. As students were dropped from the program for lack of attendance, students from the waiting list were allowed to join the program.

Attendance records for Saturday Scholars were maintained by the school-level program director, but only first semester data was available at the time of the request for

such information. There was no attendance requirement in place for this program. Some students were dropped from rosters due to lack of participation and later re-enrolled when they wanted to resume participation. Some fun activities were included in each program, such as field trips and parties. Students were only allowed to participate in such activities if they had attended a given number of program days prior to each activity. Rules were not clear about participation until a week or two prior to the activities and announced attendance requirements seemed to be rather arbitrary. Both program directors reported that attendance improved for those periods prior to an announced activity, though no record of activity dates was maintained to support the claim.

Instruments

The Georgia Criterion Referenced Competency Test is the annual measurement of student achievement. This is a selected-response test assessing the mastery of standards across the state. State law requires that all students in grades 1-8 take the CRCT in math, reading, and English/language arts each year. Additionally, students in grades 3-8 are assessed in science and social studies. Scores are reported in terms of raw scores, scaled scores, and Performance Levels. Components of the test assessing Georgia's Quality Core Curriculum have scores ranging from 150-400. Sections of the test assessing the newer Georgia Performance Standards have a range of 650-900. Riverside Publishing, the test publisher, has data to support the reliability and validity of the test as used for a measure of student achievement. Reading is reported for purposes of AYP determination as a combination of the reading and English/language arts scores. Math and reading are the critical content areas, in that these areas are used for student and school achievement decisions. Math and reading performance, therefore, is considered in this study.

Data Collection and Procedures

Permission to obtain data necessary for this study was granted by the school system's Superintendent. The county's statistical analyst provided CRCT scores for the 2006 and 2007 school years. The After-School Academy director provided the attendance data for program participants. The Director of Guidance and Testing provided attendance data for the Saturday Scholars participants. Liberty University's Institutional Review Board approved the methodology used for this study.

Participants were separated by program of participation forming two groups, After-School Academy and Saturday Scholars. Program participants were then coded to illustrate attendance habits in each program. The After-School Academy group with available pre and posttest data only presented two students attending fewer than half of the program days. A group of two would not make a fair comparison, so a cutoff of 30 days attendance was used. That cutoff created more comparable comparison groups. Saturday Scholars had more participants with pre and posttest data available, so the half-time participation served an adequate division for the attendance groups. Those attending more of the program days were coded A and those attending less of the time were coded with a B. Four groups were created: 30 or more attendance days in After-school Academy, less than 30 days attendance in After-school Academy, half or more attendance days in Saturday Scholars, and less than half attendance days in Saturday Scholars.

Program participants' scores in the lowest scored content area (math or reading) were then paired with similar scores from students who qualified for the programs but chose not to participate. A student who scored 850 in reading and 799 in math, for

instance, was paired with a student who scored similarly to the 799 math score. The higher 850 reading score was considered and used in the pairing when possible.

Effort was made to ensure that attendance in a program was the key variable impacting the posttest scores. The researcher controlled for other variables that might impact student achievement. Economically disadvantaged students' scores were paired with other economically disadvantaged students' scores. Program participants who were coded as "students with disabilities" were paired with nonparticipant "students with disabilities." As often as possible, race and gender were controlled when pairing scores.

Scaled scores were converted to Performance Levels as used by the Georgia Department of Education to determine a school's AYP status. QCC scores below 300 and GPS scores below 800 are Performance Level 1. QCC scores 300-349 and GPS scores 800-849 are Performance Level 2. QCC scores 350 and above and GPS scores 850 and above are Performance Level 3.

All performance level scores analyzed were placed into charts, showing participants in each program and the paired scores. The charts were divided to show those attending at least half of the program days, those attending less than half of the days, and paired scores for each group. Independent sample *t* tests of the posttest scores were included for each of these groups.

Data Analysis

Eight groups of performance level scores were considered. Two groups for After-School Academy (those attending 30 or more of the program days and those attending fewer than 30 days), two groups for Saturday Scholars (also divided based on attendance data), and matched scores for each of the participant groups. The number and percentage

of members in each group who moved from one performance level to a higher level were reported. The number of participants and percentage of each group were compared in order to determine a statistically significant difference in pre and posttest scores from the 2006 and 2007 math and/or reading portions of the CRCT.

Georgia has established a level of adequate performance for schools known as Annual Measurable Objectives (AMO) based on requirements of the No Child Left Behind Act. The AMOs set the percentage of students in a school that must meet or exceed the standards on the CRCT for AYP status. The AMOs increase every three years until 100% mastery is attained in 2014. The school in this study had failed to demonstrate AYP for eight consecutive years. Each time a school fails to meet the AMO, other formulas are employed (second looks) to determine if the school can be deemed sufficiently progressing. The least rigorous “second look” is safe harbor. Safe harbor requires that a school decrease by 10 percent the number of students not meeting the standard. This particular school had a school improvement plan written with goals using safe harbor formulas.

The school had 55% of its students meeting or exceeding standards in math on the 2006 CRCT. Safe harbor required that the school increase that number by 4.5% or a 10% reduction of the 45% not meeting standard. The students had 90 minutes of math each school day, for a total of 270 hours of math for the school year.

After-School Academy (ASA) provided an additional 90 hours of instruction for participants. Half of that time was spent in math instruction, for an addition 45 hours of math offered. Since the typical math class during the regular school day was expected to increase the number of students passing the test by 4.5%, ASA participants should have

an increase of 5.3% based solely on additional instructional time. For the purposes of this study, 0.8% (5.3% - 4.5%) more of the ASA participants should move from one Performance Level to a higher Performance Level than nonparticipants for the difference to be significant.

The Saturday Scholars program offered an additional 104 hours of instructional time, half of which was spent on math instruction. That resulted in an additional 52 hours of math instruction. Using the 4.5% improvement expectation for all students, Saturday Scholars should have 5.4% of the participants moving to a higher Performance Level. Participants in this program should demonstrate improved achievement by moving to a higher Performance Level at a rate 0.9% (5.4%- 4.5%) larger than nonparticipants for the difference to be significant.

Independent sample *t* tests on the posttest scores were also used to determine the significance of any score differences. Eight such tests were conducted to include a math and reading posttest score *t* test for frequent attendees and less frequent attendees for each of the two after-school programs in the study.

Chapter 4

Results

As stated in Chapter 1, this study examined pre and posttest scores of participants in two after-school programs to determine the impact of those programs on student achievement. Additionally, this study examined the impact of regular attendance in those programs as compared to less frequent attendance by program participants. This chapter is organized in terms of four research questions presented in Chapter 1. It reports (1) the impact of participation in After-School Academy on student achievement and (2) the impact of participation in the Saturday Scholars program on student achievement as measured by improved Performance Levels on the Georgia Criterion Referenced Competency Test. This report then shows achievement gains made by (3) After-School Academy participants and (4) Saturday Scholar participants divided into subgroups based on frequency of attendance in those programs.

Establishing Comparison Groups

Matched pairs for a comparison group were established based on scaled scores from the 2006 Georgia Criterion Referenced Competency Test to more closely align pretest performance than using only Performance Levels would allow. Performance Level 1 for the student scores in this study represents scaled scores ranging from 745 to 797 on the 2006 CRCT in 5th grade reading. Similar scaled score spreads were presented for each grade level and content area in this study. Matching scaled scores allowed the pairs to be more homogenous than matching only Performance Levels would allow. In matching pairs of participants' and nonparticipants' pretest scores an effort was made to

align both math and reading performance, since ability in one area may impact one's ability to improve in the other area. Further, matching was completed with consideration to variables that could impact the participants' scores to include gender, race, economic status, and status as a "student with disabilities."

After-School Academy Participants' and Matched Pairs' Pretests

The After-School Academy program began in October 2006 with 47 students enrolled in the program. By December 2006, the program had 58 students on the roster, though 11 of those participants did not attend the program any days in December. In March of 2007, the last month of the program, 67 students were on the roster. Of that 67 total, 34 participants did not attend After-School Academy in March. From the original 47 program participants enrolled in October 2006, only 19 students were still active participants in March 2007. The pre and posttest scores of those 19 participants were considered for data analysis in this study.

The 2006 Georgia Criterion Referenced Competency Test (CRCT) scores were used as pretest scores. The participants' pretest scores show most of the After-School Academy group performing at Performance Level 1, or Not Meeting the Standard, in both math and reading. Analysis of the group's scores shows:

- 5 Students at Level 2 in reading and math
- 2 students at Level 2 in reading and Level 1 in math
- 0 students at Level 1 in reading and Level 2 in math
- 12 students at Level 1 in reading and math

A comparison group of nonparticipants who qualified but chose not to participate was established to reflect similar pretest scores.

For 3 After-School Academy participants, a different Performance Level was used as a “matched pair.” When considering all impacting variables, a Performance Level match was not available. A scaled score of 797 (Performance Level 1) was deemed a better match for a scaled score of 800 (Performance Level 2) than a scaled score of 825 (Performance Level 2). Three characteristics were considered nonnegotiable in matching the pairs: race, economic status, and status as a “student with disabilities.” Homogeneity was compromised with respect to gender when necessary to establish a “best” match in scaled scores. The pretest scaled scores for the comparison group, then, were very similar to the participants’ scores.

The comparison group for After-School Academy consisted of 19 students who were eligible for the program but chose not to participate. The pretest scores for those students show:

- 2 students at Level 2 in reading and math
- 2 students at Level 2 in reading and Level 1 in math
- 2 students at Level 1 in reading and Level 2 in math
- 13 students at Level 1 in reading and math

Of the 19 participants, 4 were paired with nonparticipants performing at a different Performance Level in either reading or math. Any difference in Performance Level between matched pairs is a result of the effort to closely match scaled scores while controlling for race, economic status, and “student with disabilities” status.

After-School Academy Matched Pairs’ Posttests

Once matched pairs were established based on pretest scores, posttest scores were added for analysis. Some of the comparison group had to be replaced at this point.

Students who did not participate in both reading and math portions of the 2007 CRCT did not have available posttest scores. Of the original 19 comparison group's scores 7 were replaced. Details given above concerning the scores from the comparison group reflect the scores of the final matched pairs.

The After-School Academy participants' 2007 CRCT posttest scores were similar to the pretest scores. Many After-School Academy participants remained at Performance Level 1. Specifically, posttest results for the participant group shows:

- 4 students at Level 2 in reading and math
- 2 students at Level 2 in reading and Level 1 in math
- 1 student at Level 1 in reading and Level 2 in math
- 12 students at Level 1 in reading and math

The comparison group for After-School Academy also had posttest scores similar to the pretest scores. Most scores remained at Performance Level 1. Posttest results for this comparison group show:

- 2 students at Performance Level 2 in reading and math
- 2 students at Performance Level 2 in reading and Level 1 in math
- 2 students at Performance Level 1 in reading and Level 2 in math
- 13 students at Performance Level 1 in reading and math

For both the After-School Academy participant group and the comparison group most scores remained at Performance Level 1.

The After-School Academy participant group had 4 of its 19 participants with score changes. The pre-posttest comparison showed 2 program participants with improved achievement demonstrated by moving from Performance Level 1 to Level 2 in

either math or reading. One student moved from Performance Level 1 to Performance Level 2 in math. Another student moved from Performance Level 1 to Level 2 in reading. The other 2 participants with score changes dropped from Performance Level 2 to Level 1: 1 in math and 1 in reading and math. A total of 11 students maintained Level 1 in reading, and 13 remained at Level 1 in math. In all, 38 program participants' scores were considered: 1 math and 1 reading score for each of the 19 participants. The After-School Academy participant group had 2 scores that improved from Performance Level 1 to Level 2, 33 scores remained at Level 1 or Level 2, and 3 scores dropped from Performance Level 2 to Level 1.

The comparison group had 6 of its 19 students with score changes. Of those 6 students with score changes, 5 students showed improved achievement by moving from Performance Level 1 to Level 2. The results reflect that 2 students moved from Performance Level 1 to Performance Level 2 in math, and 2 students moved up to Performance Level 2 in reading. Other scores changes include 3 students who dropped from Performance Level 2 to Level 1: 1 in reading, 1 in math, and 1 in reading and math. A total of 13 students maintained Level 1 in reading, and 13 remained at Level 1 in math. Of the 38 scores in the comparison group considered, 4 scores improved from Performance Level 1 to Level 2, 30 scores remained at either Level 1 or Level 2, and 4 scores dropped from Level 2 to Level 1. Table 7 shows the percentage of After-School Academy participants and the comparison group's scores that changed or maintained the pretest Performance Level. Columns may not total to 100% due to rounding.

1. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a

suburban middle school in Augusta, Georgia who voluntarily participate in the school's After-school Academy and similar students who do not participate in this program.

Table 7

Difference in Pre and Posttest Results for After-School Academy

	<u>Participants</u>		<u>Comparison Group</u>	
	Reading	Math	Reading	Math
Improved a Performance Level	5%	5%	11%	11%
Maintained Performance Level	84%	89%	79%	79%
Dropped a Performance Level	11%	5%	11%	11%

After-School Academy participants, for this study, were to show a 0.8% greater improvement than nonparticipants for the program to demonstrate a statistically significant impact on student achievement. After-School Academy participants did not demonstrate improved achievement at a rate higher than nonparticipants.

An independent sample *t* test failed to reveal a statistically significant difference between the participants' and nonparticipants' posttest scores in reading $t(36) = 1.408$, $p = .244$, $\alpha = .05$, and in math $t(36) = 1.355$, $p = .184$, $\alpha = .05$. The null hypothesis was retained.

Saturday Scholars and Matched Pairs Pretests

The Saturday Scholars program began in September 2006 with approximately 100 students enrolled. By December 2006, only 72 students were recorded as actively participating in the program. No attendance data beyond December 2006 is available for this program. Of those 72 listed participants, pre and posttest scores were available for

only 57. Those 57 participants comprise the Saturday Scholars participant group for this study and the group for which matched pairs were established.

The 2006 Georgia Criterion Referenced Competency Test (CRCT) was used as the pretest, showing 45 of the 57 participants performing at Level 1 (Does Not Meet the Standard) in either reading or math or both. Overall, the Saturday Scholars' pretest scores show:

- 11 students at Performance Level 2 in reading and math
- 16 students at Performance Level 2 in reading and Level 1 in math
- 6 students at Performance Level 1 in reading and Level 2 in math
- 23 students at Performance Level 1 in reading and math

This group had 1 participant with a pretest score at Level 3 (Exceeds the Standard) in reading. Matched pairs were established to resemble as closely as possible both reading and math scores of participants.

As with the pairings for After-School Academy participants, matched pairs were established based on scaled scores rather than Performance Levels. The scaled scores allowed for a closer alignment of performance on the pretest than the broader Performance Level score. Matching of pairs was completed with careful consideration to variables that could possibly influence achievement gains. Gender, race, economic status, and status as a "student with disabilities" were considered as scores were paired to create the comparison group.

For 31 Saturday Scholar participants a different Performance Level score was used as a "matched" pair. A close scaled score, within 4 to 6 points, was considered a better match than a scaled score 10 to 15 points away but at the same Performance Level.

Three student characteristics were matched pervasively: race, economic status, and “student with disabilities” status. A gender match was compromised when necessary to establish a “best” match in scaled scores. The participant group is closely paired according to scaled scores and student characteristics listed above.

The comparison group for the Saturday Scholars program posted scores similar to the Saturday Scholars. The scores from this group show:

- 15 students at Performance Level 2 in reading and math
- 17 students at Performance Level 2 in reading and Level 1 in math
- 0 students at Performance Level 1 in reading and Level 2 in math
- 24 students at Performance Level 1 in reading and math

The comparison group also had one student who performed at Level 3 in reading on the pretest. Some scores, 26 Saturday Scholar participants’ scores, were paired with nonparticipant scores at the same Performance Level in both reading and math. For the other matched pairs, the scaled scores were more closely matched than Performance Level pairing allowed. Any difference in Performance Levels between matched pairs is a result of the effort to maintain race, economic status, and disability matches while closely aligning scaled scores.

Saturday Scholars and Matched Pairs Posttests

Posttest scores were added once all matched pairs were established. Many scores from the comparison group had to be replaced during the addition of the posttest scores. Students who did not participate in both the math and reading testing sessions of the 2007 CRCT had to be replaced so that all pre and posttest data were available. A total of 19 of

the original 57 scores in the comparison group were replaced. The comparison group's scores mentioned above are from this final group of scores used for analysis.

The data shows many of the Saturday Scholar participants performed at the same Performance Level on the pre and posttest. Posttest results reflect 23 participants remained at Level 1 in reading and 40 remained at Level 1 in math. Across all Performance Levels, 38 maintained the same level from pre to posttest in reading and 43 maintained their pretest Performance Level in math. Posttest results reflect:

- 15 students at Performance Level 2 in reading and math
- 17 students at Performance Level 2 in reading and Level 1 in math
- 0 students at Performance Level 1 in reading and Level 2 in math
- 23 students at Performance Level 1 in reading and math

Only 2 students from the Saturday Scholars program scored at Performance Level 3, 1 student in reading and 1 in both reading and math.

The comparison group had posttest scores similar to their pretest scores. Posttest results reflect:

- 18 students at Performance Level 2 in reading and math
- 20 students at Performance Level 2 in reading and Level 1 in math
- 1 students at Performance Level 1 in reading and Level 2 in math
- 15 students at Performance Level 1 in reading and math

The comparison group posted 3 scores at Performance Level 3, 2 in reading and 1 in math. For both the Saturday Scholars participants and the comparison group, many scores remained at the pretest Performance Level.

The Saturday Scholars group had 29 of the 57 participants with Performance Level changes. A total of 18 program participants showed improved achievement by moving from Level 1 to Level 2 in either reading or math. Of the 18 participants with improved Performance Level scores, 6 students moved from Level 1 to Level 2 in math, and 12 moved from Level 1 to 2 in reading. Improvement was also demonstrated by 2 participants' moving from Level 2 to Level 3. Another 12 participants dropped from Performance Level 2 to Level 1; 6 in math, 5 in reading, and 1 in both math and reading. Posttest scores show that 15 students maintained Level 1 in reading, and 33 remained at Level 1 in math. In all, 114 scores were considered for the Saturday Scholars program participants: 1 math and 1 reading score for each of the 57 participants. Of those 114 scores, 20 scores improved from one Performance Level to a higher Level, 81 scores remained at the pretest Performance Level, and 13 scores dropped to a lower Performance Level.

The comparison group also consisted of 114 scores for consideration. Of the 114 scores for the comparison group, 20 scores indicated improved achievement by moving to a higher Performance Level, 90 scores maintained the pretest Performance Level, and 4 scores dropped a Performance Level from the pretest. Table 8 illustrates the percentage of Saturday Scholars participants' scores and comparison group scores that either changed from or maintained the pretest Performance Level. Columns may not total to 100% due to rounding.

2. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a

suburban middle school in Augusta, Georgia who voluntarily participate in the school's Saturday Scholars program and similar students who do not participate in this program.

Saturday Scholars participants, for this study, were to show a 0.9% greater improvement than nonparticipants for the program to demonstrate a statistically

Table 8

Difference in Pre and Posttest Results for Saturday Scholars

	Participants		Comparison Group	
	Reading	Math	Reading	Math
Improved a Performance Level	23%	12%	21%	14%
Maintained Performance Level	67%	75%	74%	84%
Dropped a Performance Level	11%	12%	5%	2%

significant impact on student achievement. Saturday Scholars participants improved a Performance Level in reading 2% greater than the improvement demonstrated by nonparticipants. The participant group, however, did not demonstrate greater improvement in math and had a larger percentage drop a Performance Level in both math and reading.

An independent sample t test failed to reveal a statistically significant difference between the participants' and nonparticipants' posttest scores in reading $t(112) = 1.272$, $p = .206$, $\alpha = .05$, and in math $t(112) = 1.938$, $p = .055$, $\alpha = .05$. The null hypothesis was retained.

After-School Academy Attendance Subgroups

A second consideration was given to the impact of the After-School Academy program on student achievement with respect to the number of days students participated in the instructional program. The program began in October 2006 and lasted until March 2007, meeting 3 days per week for additional after-school instruction. Students were not required to participate or to attend any specific number of days to be considered active in the program. Attendance records indicate that some students rarely attended and only a few students attended consistently. Of the 44 days the program offered instruction, 17 students attend half or more of those days. Only 2 of the participants with pre and posttest scores available attended fewer than half of the program days. Participants were divided into two groups: those attending 30 or more days and those attending fewer than 30 days. The 30-day cutoff was selected because it split the group into more comparable subgroups than using half of the attendance days, 22, as the cutoff. Both groups, with respect to days in attendance, show that most participants maintained the pretest Performance Level.

One participant in the 30+ days attendance group dropped a Performance Level in math, and one participant in the fewer than 30 days group dropped a Performance Level in both math and reading. Both subgroups have one participant showing an improved Performance Level. A student with 30+ days in attendance improved a Performance Level in reading, while a student with fewer than 30 days in attendance showed improvement in math. Table 9 illustrates the percentage of After-School Academy participants' scores that either changed from or maintained the pretest Performance Level

divided into subgroups based on attendance. Columns may not total to 100% due to rounding.

Table 9

Difference in Pre and Posttest Results for After-School Academy by Attendance

	30 or more days		Fewer than 30 Days	
	Reading	Math	Reading	Math
Improved a Performance Level	8%	0%	0%	14%
Maintained Performance Level	83%	100%	86%	71%
Dropped a Performance Level	8%	0%	14%	14%

3. There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who frequently attend the school's After-School Academy and similar students who attend this program less frequently.

The participants who attended After-School Academy 30 or more days demonstrated 8% greater improved achievement in reading than participants attending fewer than 30 days. Those participants attending 30 or more days showed no improvement in math Performance Level, while the participants attending fewer than 30 days showed a 14% improvement in math. Using the same level of significance established for all participant scores in this program (0.8% greater improvement), participants who attended 30 or more days demonstrated improved achievement in reading at a greater rate than participants attending fewer than 30 days, but they were not able to demonstrate the same improvement in math.

An independent sample t test failed to reveal a statistically significant difference between the more frequent and less frequent participants' posttest scores in reading $t(17) = 1.246, p = .230, \alpha = .05$, and in math $t(17) = .870, p = .397, \alpha = .05$. The null hypothesis was retained.

Saturday Scholars Attendance Subgroups

Participation and attendance in the Saturday Scholar's Program was also voluntary. Even through the first half of the program, the time for which attendance records are available, participation was sporadic for most students. Only 17 participants attended half of the sessions available. More than half of the participants, 34 of the 57 participants considered in this study, attended less than one third of the program sessions the first semester. Subgroups for this program were divided between those attending half of the sessions first semester and those attending fewer than half of the program sessions.

As with After-School Academy participants, most Saturday Scholars participants in each subgroup performed at the same Performance Level for both the pre and posttest on the CRCT. A total of 34 scores were analyzed for the subgroup attending half or more of the sessions, and those attending fewer than half had a total of 80 scores.

For the 17 participants attending half or more of the sessions, 3 scores showed an increase in Performance Level, 2 in reading and 1 in math. For 4 participants the scores actually dropped from the pretest Performance Level. A total of 27 participants' scores maintained the pretest Performance Level. The subgroup attending fewer than half of the program days showed an increase in 17 scores, a decrease in 9 scores, and maintained pretest Performance Levels for 54 scores. Table 10 shows the percentage of changed

Performance Level scores by attendance subgroups, based on first semester attendance, for reading and math. Columns may not total to 100% due to rounding.

Table 10

Difference in Pre and Posttest Results for Saturday Scholars by Attendance

	12+ Days		Fewer than 12 Days	
	Reading	Math	Reading	Math
Improved A Performance Level	12%	6%	28%	15%
Maintained Performance Level	76%	82%	63%	73%
Dropped A Performance Level	12%	12%	10%	13%

4. *There will be no significant difference in improved achievement, as measured by the Georgia Criterion Referenced Competency Test, for low performing students in a suburban middle school in Augusta, Georgia who frequently attend the school's Saturday Scholars program and similar students who attend this program less frequently.*

The participants who attended the Saturday Scholars program 12 or more days (half of the first semester) did not demonstrate improved achievement at a rate greater than participants who attended fewer than 12 days.

An independent sample *t* test failed to reveal a statistically significant difference between the more frequent and less frequent participants' posttest scores in reading $t(55) = .032, p = .975, \alpha = .05$, and in math $t(55) = .714, p = .478, \alpha = .05$. The null hypothesis was retained

Note Regarding Percentage of Performance Level Changes

In some cases, small numbers are represented by deceptively larger percentages. One should consider the number of improved scores with regard to the number of scores in the subgroup. In the final chapter, Performance Level changes within each group will be analyzed with discussion as to the possible impact of these programs on student achievement.

Chapter 5

Summary and Discussion

For the benefit of the reader, this final chapter reviews the research problem and the methods used in this study. That review is followed by a summary of the results and a discussion of their implications.

Statement of the Problem

The purpose of this study was to determine if participants in two after-school programs in a Richmond County, Georgia “Needs Improvement” middle school demonstrated improved achievement on the Georgia Criterion Referenced Competency Test (CRCT) in higher percentages than nonparticipants. Additionally, this study attempted to determine if participants’ frequency of attendance in these programs impacted achievement.

Review of Methodology

As reported in Chapter 3, this was a study of two after-school programs offered at a middle school that was in their 8th year of Needs Improvement status. Participation in each program was voluntary and there were no attendance requirements for staying in or returning to either program. One program, After-School Academy, was immediately after school three days a week. The other, Saturday Scholars, was offered on Saturday mornings. Applications were initially given only to students with certain qualifying characteristics, such as failing the 2006 CRCT, status as a disabled student, or status as a free/reduced lunch (economically disadvantaged) student. Most of those restrictions were removed when the programs were unable to attract the number of students for which

the programs were funded. Saturday Scholars was a Title I funded program and had to retain the free/reduced lunch qualification rule, but the 2006 CRCT scores were not considered when efforts were made to increase enrollment.

The CRCT scores from 2006 and 2007 were used as pre and posttest scores to measure student achievement. Score reports offer a variety of reporting formats, to include scaled scores and performance levels. Participants in the after-school programs were paired with nonparticipants based on their scaled scores in an effort to closely align pretest achievement levels. The analysis of improved achievement was based on improvement in Performance Level as reported on the CRCT. The Performance Level is the score used to determine pass/fail for this particular test and the score used to determine school funding for these programs. Results are then reported in terms of the percentage of students in each group who improved a Performance Level from pre to posttest, who maintained their pretest Performance Level, and who dropped a Performance Level on the posttest. Independent sample *t* tests were then conducted on posttest scores.

Summary of Results

This study examined the impact of two after-school programs with additional consideration given to the frequency of attendance in those programs. The summary of the results is divided into sections addressing the four research questions set forth in Chapter 1.

1. What impact on achievement will voluntary attendance in the After-school Academy have for low performing students in a suburban middle school in

Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

Nineteen students' scores were considered in the After-School Academy participant group. The group began with a much larger number of students enrolled in the program, but only nineteen of the original group remained active by the end of the program.

Most students in the study, both participants in After-School Academy and nonparticipants, remained at their pretest Performance Level. Only 5% of the participants and 11% of the nonparticipant group improved a Performance Level. Similar percentages dropped a Performance Level. In the nonparticipant group, 11% dropped in reading and math. For participants, 11% dropped in reading and 5% in math. After-School Academy participants did not improve as much as nonparticipants in the comparison group. A *t* test failed to reveal a statistically significant difference between the posttest scores of participants and nonparticipants.

2. What impact on achievement will voluntary attendance in the Saturday Scholars program have for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

Saturday Scholars had 57 active participants with pre and posttest scores available at the end of the first semester. Attendance records kept at the county's central office only maintained attendance data through December. This study, therefore, used the first semester attendance data for determining active participation.

The Saturday Scholar participant group did have a slightly higher percentage improve a Performance Level in reading than the comparison group, 23%

compared to 21%. The Saturday Scholars participants had a slightly smaller percentage improve in math than nonparticipants, 12% compared to 14%. In both reading and math the Saturday Scholars group had a larger percentage drop a Performance Level than the comparison group. Saturday Scholars had 11% of their participants drop a Performance Level in reading and 12% drop in math, compared to 5% and 2% respectively in the comparison group. Most students in each group remained at their pretest Performance Level. The *t* test failed to reveal a statistically significant difference between the posttest scores of participants and nonparticipants.

3. Will more frequent attendance in the After-school Program have a greater impact on achievement than less frequent attendance for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

The After-School Academy met for a total of 44 days. Using half of those days to divide the group would have created one subgroup of two students. That seemed too small of a group for any reliable comparison. Therefore, the group was divided into a subgroup of participants who attended 30+ days and a subgroup that attended fewer than 30 days. This created a group of 12 participants with 30+ days in attendance and a group of 7 participants attending fewer than 30 days.

Students attending After-School Academy 30+ days showed a higher percentage of improvement in reading than those attending fewer than 30 days, 8% compared to 0%. The 30+ days group, however, had no students improving a Performance Level in math, while the fewer than 30 days group had a 14% improvement. The 30+ days group had 8% drop a Performance Level in reading and no one drop in math. The fewer than 30

days group had 14% drop in both reading and math. Most students in each group maintained their pretest Performance Level. A *t* test failed to reveal a statistically significant difference between the posttest scores of participants who attended frequently and those attending less frequently.

4. Will more frequent attendance in the Saturday Scholars program have a greater impact on achievement than less frequent attendance for low performing students in a suburban middle school in Augusta, Georgia, as measured by the Georgia Criterion Referenced Competency Test?

The Saturday Scholars participant group was divided into subgroups based on half of the 24 available program days. One subgroup, those attending 12 or more days, had 17 participants. The other subgroup, those attending fewer than 12 days, had 40 participants.

The Saturday Scholars group showed a greater difference in student achievement when divided into these attendance subgroups. Those participants attending fewer than 12 days outperformed those with more regular attendance in most areas. The 12+ days attendance subgroup showed a gain of 12% in reading and 6% in math. The subgroup attending fewer than 12 days had 28% improve a Performance Level in reading and 15% in math. The subgroup attending more days had 12% drop a Performance Level in both reading and math, while those attending fewer days showed a 10% drop in reading and 13% in math. Again, most participants remained at their pretest Performance Level. A *t* test failed to reveal a statistically significant difference between the posttest scores of participants who attended frequently and those attending less frequently.

Discussion of the Results

Interpretation of the Findings

This report includes analysis of eight score comparisons when reading and math were considered separately. There were four groups of scores compared:

1. After-School Academy participants compared to a group of nonparticipants
2. Saturday Scholars participants compared to a group of nonparticipants
3. After-School Academy participants attending 30+ days compared to participants attending fewer than 30 days
4. Saturday Scholars participants attending 12+ days compared to participants attending fewer than 12 days.

Each group posted reading and math scores on the pre and posttest. Nonparticipants outscored after-school program participants in six of the eight comparisons. In only two comparisons did program participants demonstrate improved achievement greater than the nonparticipants or participants with fewer days attendance. Saturday Scholars participants improved in reading by 2% more than nonparticipants, and the participants attending 30+ days in After-School Academy improved in reading by 8% more than those attending fewer than 30 days. The nonparticipant group demonstrated improved student achievement greater than the program participants in three of four comparisons. Those attending the after-school programs less frequently showed greater improvement than the participants with better program attendance in three of four comparisons. In several comparisons the program participants or those with better attendance also posted a greater percentage of posttest scores dropping a Performance Level than the comparison

group. A positive impact of participation in an after-school program on student achievement is not evident in the posttest scores of the participants in this study.

The null hypothesis, as set forth in Chapter 1, anticipated no difference in performance between program participants and nonparticipants or between those who attended the after-school programs regularly and those who attended less frequently. The results seem to indicate program participation and good attendance are related to reduced student achievement. Leon Botstein wrote in *The New York Times* (2001) that additional school time "... as it is now utilized might even lower achievement." While the study at hand is too small to make any broad statements about after-school programs, it is interesting that the students participating in the programs and those attending more regularly performed poorly compared to the nonparticipant control groups or less frequent attendees. A broader look helps to show how this particular school and the findings compare to neighboring schools and the county at large.

The results are not unique to the middle school in this study. An article in the *Augusta Chronicle* (Gelpi, 2007) reports similar results were found across the county. According to this article, the county's Saturday Scholars scored slightly better than the control group in reading, but they were outperformed in math by the control group. The opposite was true for the After-School Academy participants. After-School Academy students performed slightly better than nonparticipants in math, but showed less improvement than nonparticipants in reading.

Richmond County, home to the middle school in this study, funded research by The Edgewater Institute (2007). Their report considered data from 25 schools in Richmond County with after-school programs. In addition to the achievement data

reported in the Augusta Chronicle, this report considered the impact of regular attendance in the after-school programs. An analysis of variance showed that those participants attending the programs 80% of the time scored better than those attending more or fewer days. Interestingly, this study found it did not matter when students enrolled in the programs. Those participants who began in January did as well, as a group, as those who began the program in September.

The middle school programs in this study seem to demonstrate results similar to the rest of the county. Program participation and regular attendance showed no consistent positive impact on student achievement. Other research exists to indicate these results were to be expected.

Related to Other Research

A fair amount of research on after-school programs is available and much of the data is relatively current. Finding agreement in the studies' results and conclusions is the challenge. That is to be expected, according to Thomas Kane (2004). Kane states that the problem with most after-school program studies is in research design, citing research that seeks to find too great of an impact in too little time. Statistical significance for this After-School Academy and Saturday Scholars study was derived from a procedure suggested by Kane, basing expectations of achievement gains on clock hours in the program and anticipated achievement gains from a regular school year. In agreement with Kane is Darcy Olsen (2000), Director of Education and Child Policy at Cato Institute, who says the after-school program research is so wrought with flaws in methodology that the findings reveal little about the impact of the programs on student achievement. One should consider the findings of a study in light of these concerns.

A Mid-continent Research for Education and Learning (McREL, 2003) study looked at 371 studies of after-school programs and found only 56 that met high standards of rigor necessary for serious research consideration. Those studies showed the most gains in student achievement resulted from one-on-one tutoring. That delivery model was never used in the After-School Academy or Saturday Scholars program in the current study.

A study by the AfterSchool Alliance (2006) reports of several after-school programs demonstrating positive impacts on student achievement, including The After-School Corporation and North Carolina's Young Scholars Program. No statistical evidence is given, but the report does indicate that attendance was a critical factor in at least one study. Other researchers also find improved student achievement resulting from after-school programs with a focus on regular attendance (Huang, 2000, National Institute of Child Health, 2004, AfterSchool Alliance, 2006). Attendance was not found to be a critical influence on improved achievement for participants in the After-School Academy or Saturday Scholars programs.

Wanda Washington (2000) found that extended learning programs are generally more effective with elementary participants than other grade levels. Extended week programs, such as Saturday Scholars, were more effective with middle school students than extended day programs in the schools Washington studied. The present research of Saturday Scholars and After-School Academy did not make a side-by-side comparison to make such a determination. However, the Saturday Scholar participants fared no better when compared to the nonparticipant control group than After-School Academy participants did in comparison to a similar control group. It seems unlikely that one

program would demonstrate a greater positive impact on student achievement than another, but that has not been determined.

Much of the current research on after-school programs addresses the impact on participants other than improved academic achievement. Self-esteem, self-help skills, social skills, and independence are a few of the skills reportedly improved through program participation. No such data was collected on participants in the After-School Academy or Saturday Scholars programs. These programs were funded solely to improve student achievement, and the program directors implemented programs of academic instruction.

The No Child Left Behind rules do not allow schools the luxury of ignoring the impact of such programs on student achievement. An interim report of accountability (US Dept. of Ed. Office of Planning, 2007) states that 25% of the nation's school failed to demonstrate Adequate Yearly Progress, AYP, in 2004; about a third of those failed to demonstrate AYP due to students with disabilities or limited English proficiency. The available research, of which this current study is a small piece, indicates after-school programs are often failing to meet the academic needs of struggling learners in the schools least able to accomplish AYP. As 2014 is quickly approaching, school leaders and program directors must find and implement those strategies most likely to impact student achievement and provide sustainable improvement.

Recommendations

This section attempts to address possible barriers to student success in After-School Academy and Saturday School, suggesting specific areas program leaders might address. Anne Turnbaugh Lockwood (2003) found that after-school program directors

must look closely at such barriers and remove them before it is possible to find improvements in student achievement directly resulting from after-school programs.

Lorna Idol (1998) found that professional learning for after-school instructors was a key piece missing from many programs. She suggested that professional learning could help the instructors bridge the gaps between school day instruction and after-school support. Julie Aronson (1999) suggested that teachers could benefit from professional development that focuses specifically on time management. Bill Metzker (2003) cautions that in a school where time is not already well used, adding more time is unlikely to lead to improved academic achievement.

- Program directors should provide professional development for instruction and time management for all teachers in the program and conduct follow-up visits to ensure the strategies are being implemented and all instructional time is used wisely.

The Massachusetts 2020 study, *Time for a Change* (Farbman & Kaplan, 2005), finds that strong leadership and positive school culture are essential to the success of after-school programs. Successful principals and leaders, according to this report, are creative, supportive, and convey a compelling vision. They create safe environments that focus on education at every level of the institution. Leaders for programs must be carefully selected and well equipped for the challenges before them.

- After-school program leaders should be carefully selected based on personal and professional qualities shown to inspire high performance among teachers and students. Teacher selection for the programs should

also receive careful consideration and be based, in large part, on data indicating the ability to positively impact student achievement.

Beth Miller (2003) found one of the biggest barriers to success of middle school programs was an inability to maintain student participation. Middle School students often choose what programs they want to attend, and they opt out of any activity they find boring. Miller says this age group “votes with their feet.” Since many struggling learners already dislike school, crafting after-school programs to attract middle school students is difficult. Miller found activity-based programs had an easier job of recruiting and retaining participants.

- Middle school after-school programs should be highly creative. The instructional blocks should be aligned with the regular school day standards, but the activities should look very engaging and enticing to middle school students. The participants should have choice of activities and a variety in programming options.

As previously stated, research indicates a need for clear data and data-driven decisions in the after-school programs. The *Massachusetts 2020* report (Farbman & Kaplan, 2005) lists examples of several programs for which data was the constant driving force. One school, Community Day Charter School, hired a full time data analyst to keep teachers abreast of how each student did on each type of question the previous year. While that type of exuberance is not common, data driven instruction is a common thread in many of the successful programs. Metzker (2003) says teachers must make sure instruction is a “good fit” with the student’s ability and readiness to learn.

- Data should be used to craft individual learning goals in the after-school program. Allow the middle school students to participate in decisions concerning the instructional areas of focus, target goals, and measures of success. Teachers may need additional professional learning for making such data-driven decisions.

Limitations

The participants in this study were those students who were invited and volunteered to attend either of the two after-school programs from the first month of program operations and were still attending in the last month of operations for which attendance data was available. Invitations to participate were given initially only to students who “failed” the 2006 Georgia Criterion Referenced Competency Test and were deemed a “student with disabilities” or an “economically disadvantaged student.” Both programs had far fewer participants most days than funding allowed, and a very small number of participants were considered for the After-School Academy participant group in this study. The qualifying factors used to determine who initially qualified to participate limits the researcher’s ability to apply these findings to other participants in the same after-school programs or other similar programs.

The reliability of the Georgia Criterion Referenced Competency Test used as the tool to measure achievement also limited the researchers ability to determine with confidence the degree of improved achievement.

Recommendations for Further Research

Both after-school program participant groups in this study were considerably smaller than the size necessary to draw sweeping conclusions about after-school

programs' impact on student achievement. Continued research in after-school programs is necessary before a conclusive determination can be made concerning the impact on student achievement.

The anticipated impact on student achievement from a program that lasts only a few hours a week would be small. Some research indicates the impact of a single year of after-school participation is undetectable with a standardized test. A more comprehensive understanding of the impact from after-school programs would be gleaned from a longer study, assessing the improved achievement from several years of consistent program participation.

Much work remains in the area of improving student achievement. After-school programming is one attempt by many schools and districts to meet the expectations set forth in the No Child Left Behind Act. Rigorous standards have been established, and standardized tests have been created to measure student progress on the standards. The current goal of many after-school programs is to do that which is necessary to move struggling learners to areas of proficiency on those tests. That being the goal, this study and other research reported herein seems to indicate the after-school programs are not effective. Should the programs continue to be funded to support academic achievement of those most in need of additional instruction? Program directors will continue to look to research and data for answers to the many questions the programs create. Further research to help determine what constitutes quality instruction and the type of curriculum effective in after-school programs will prove valuable to program directors and participants. While much has been learned, there remain, unfortunately, far more questions than the research has yet answered.

References

- Adelman, N. (1996). *The uses of time for teaching and learning*. Washington, DC: Policy Studies Associates, Inc.
- Afterschool Alliance. (2006, September). *Evaluations backgrounder: A summary of formal evaluations of the academic impact of afterschool programs*. Washington, DC: Author.
- Aronson, J., Zimmerman, J., & Carlos, L. (1999). *Improving student achievement by extending school: Is it just a matter of time?* San Francisco, CA: WestEd. (ERIC Document Reproduction Service No. ED435127).
- Ary, D., Cheser Jacobs, L., Razavieh, A., & Sorensen, C. ((2006). *Introduction to Research in Education*. Belmont, CA: Thomson Higher Education.
- Bell, T. H. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: The National Commission on Excellence in Education.
- Bennett, W. J., Fair, W., Finn, C. E., Flake, F. H., Hirsch, E. D., Marshall, W. & Ravitch, D. (1999). *A nation still at risk*. (Report No. ED 429988). College Park, MD: Clearinghouse on Assessment and Evaluation.
- Botstein, L. (2001, January 25). We waste our children's time. *The New York Times*. Opinion. Retrieved December 13, 2007 from <http://query.nytimes.com/gst/fullpage.html?res=9DO1E3D61F3CF936A15752COA9679C8B63>

- Brown, W. O., Frates, S. B., Rudge, J. S., & Tradewell, R. L. (2002). *The costs and benefits of after school programs: The estimated effects of the After School Education and Safety Program Act of 2002*. Claremont, CA: Rose Institute.
- Byrd, J. K. (2001, April). *Student achievement: Is equality really necessary?* Paper presented at the Annual Meeting of the American Education Research Association, Seattle, WA.
- Clark-Keys, Karen M. (2007) *A participants' Alignment of Goals Assessment (PAGE) of after-school/expanded learning opportunities art education programming*. Unpublished doctoral dissertation, Ohio State University.
- Dunn, L, Parry, S. & Morgan, C. (2002). *Seeking quality in criterion referenced assessment*. Paper presented at the Learning Communities and Assessment Cultures Conference, Southern Cross University, Australia.
- Dynarski, M., Moore, M., Mullens, J., Gleason, P. James-Burdumy, S., Rosenberg, L., Pistorino, C., Silva, T., Deke, J., Mansfield, W., Heaviside, S., & Levy, D. (2003). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers program. First Year Findings*. Washington, D.C.: Education Publication Center.
- Evans, W. & Bechtal, D. (1997). *Extended school day/year programs: A research synthesis* (Report No. LSS-Ser-212). Mid-Atlantic Lab for Student Success. Philadelphia, PA: Office of Educational Research and Improvement.
- FairTest: The National Center for Fair & Open Testing. Retrieved September 30, 2007, from <http://www.fairtest.org/facts/srctests.html>

- Farbman, D. & Kaplan, C. (2005). *Time for change: The promise of extended-time schools for promoting student achievement*. Research Report. Boston, MA: Massachusetts 2020.
- Florian, J. (1999) *Teacher survey of standards-based instruction: Addressing time*. Aurora, CO: Mid-Continent Research for Education and Learning. (ERIC Document Reproduction Service No. ED471246)
- Gelpi, G. (2007, July 8). Programs show little benefit. *The Augusta Chronicle*, pp. A1, A11.
- Georgia Department of Education- *2007 CRCT Score Interpretation Guide*. [Data file]. Atlanta, GA.
- Georgia Department of Education- *NRT System Test Coordinator Guide, 2007* [Data file]. Atlanta, GA.
- Georgia Department of Education- Title 1 Office: August 30, 2006- *Provider contributions to student academic proficiency: Individual SES providers' students achievement test results in 2005-2006* [Data file]. Atlanta, GA.
- Gerrig, R., & Zimbardo, P. (2002) Psychology and life. *APA OnLine: Psychology Matters Glossary*. 16/e. Retrieved October 1, 2007, from <http://www.psychologymatters.org/glossary.html>
- Halpern, R. (2000). The promise of after-school programs for low-income children, *Early Childhood Quarterly*, 15,185-214.

- Harnish, D., Thompson, J., Pollack, S., Cramer, S., & Alagoz, C. (2006). *Provider contributions to student academic proficiency: Individual SES providers' student achievement test results in 2005-2006*. University of Georgia: College of Education
- Huang, D., Gribbons, B., Kim, K. S., Lee, C., & Baker, E. L. (2000). *A decade of results: The impact of the LA's Best After School Enrichment Program on subsequent student achievement and performance*. UCLA Center for the Study of Evaluation.
- Hunt, Sharon, & Millicans, S. (2005). *Georgia CRCT pre-administration workshop* [Powerpoint]. Georgia Department of Education, Atlanta, GA.
- Idol, L. (1998). *Optional extended year program*. Austin Independent School District. Austin, TX: Office of Program Evaluation.
- Kane, T. J. (2004, January). *The impact of after-school programs: Interpreting the results of four recent evaluations*. Los Angeles, CA: William T. Grant Foundation.
- Lauer, P., Akiba, M., Wilkerson, S., Athrop, H., Snow, D., Martin-Glenn, M. (2003). *The effectiveness of out-of-school time strategies in assisting low-achieving students in reading and mathematics: A research synthesis*. Aurora, CO: Mid-Continent Research for Education and Learning. (ERIC Document Reproduction Service No. ED483021)
- Lockwood, A. T. (2003, August). *Barriers to afterschool programs* [Electronic version]. The School Administrator.
- Metzker, B. (2003). *Time and learning*. Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Report No. ED474260).

- Mid-Continent Research for Education and Learning. (2003, December). *The effectiveness of out-of-school time strategies in assisting low-achieving students in reading and mathematics: A research synthesis*. Aurora, CO: Author.
- Miller, B. M. (2003). *Critical hours: Afterschool programs and educational success*. Brookline, MA: Nellie Mae Education Foundation.
- Miller, K., Snow, D., & Lauer, P. (2004). Out-of-school time programs for at-risk students. *Noteworthy Perspectives*. Aurora, CO: Mid-Continent Research for Education and Learning.
- Millicans, S. (2004). *Georgia criterion-referenced competency tests: Fall assessment conference* [Powerpoint]. Georgia Department of Education, Atlanta, GA.
- Millicans, S. (2007). *2006 GA CRCT Technical Report*. Personal communication, September 18, 2007.
- National Institute of Child Health and Human Development Early Child Care Research Network. (2004). *Are child development outcomes related to before- and after-school care arrangements? Results from the NICHD study of early child care*. *Child Development*, 75 (1), 280-295.
- New York City Board of Education. (2000). *Analyses of performance of extended-time and non-extended time SURR schools*. (UD 034096). Brooklyn, NY: NYC BOE Division of Assessment and Accountability.
- No Child Left Behind Act of 2001 (2005). Retrieved January 11, 2007, from <http://www.doe.k12.ga.us/support/plan/nclb/asp>

- Olsen, Darcy. (2000). *12-hour school days? Why government should leave afterschool arrangements to parents*. Policy Analysis, No. 372. Washington, DC: Cato Institute. Retrieved December 2007 from <http://www.cato.org/pubs/pas/pa-372es.html>
- Online Assessment System, Criterion-Referenced Competency Tests (2005). Retrieved January 11, 2007, from <http://www.doe.k12.ga.us/curriculum/testing/crcr/asp>
- Organization For Economic Co-operation and Development. (2005). *Education at a glance*. Retrieved September 10, 2007 from, <http://www.oecd.org/edu/eag2005>
- Pardini, P. (2001). Extended school days. *The School Administrator*. August 2001.
- Rangel, E. S. (2007). Essential information for education policy: Time to learn. Research Points. American Educational Research Association, Winter 2007, 5, 2.
- Roukema, R. A. (2005). *The impact of the Support Our Students (SOS) after-school program on the achievement of middle-grade students at risk of academic failure*. Unpublished doctoral dissertation, North Carolina State University.
- Silva, E. (2007). On the clock: Rethinking the way schools use time. *EducationSector Reports*. Retrieved January 15, 2007 from http://www.educationsector.org/index_show.htm?doc_id=548916
- The Edgewater Institute. (2007, June). *Analysis of Richmond County Saturday Scholars and After School Academy Programs*. Okatie, South Carolina: Author.
- United States Conference of Mayors. (2003, June). *After school programs in cities across the United States*. Washington, DC: Charles Scott Moss Foundation.

- United States Department of Education. (February 26, 2007). *Choices for parents: Description of supplemental educational services* [Data file]. Washington, DC. Retrieved September 3, 2007, from <http://www.ed.gov/nclb/choice/help/ses/description.html>
- United States Department of Education, Office of Planning, Evaluation, and Policy Development. (2007). *State and local implementation of the No Child Left Behind Act, Volume III- Accountability under NCLB: Interim report*. Washington, DC: Education Publications Center.
- University of Texas at Austin, Division of Instructional Innovation and Assessment. (August, 2003). *Test Item Analysis & Decision Making*. Retrieved November 3, 2007, from www.texas.eud/academic/mec/research/pdf/itemanalysisishandout.pdf
- Washington, W. (2000). *Optional extended year program*. (Report No. AISD-OPE 98.15). Austin, TX: Austin Independent School District Office of Program Evaluation.
- Worthen, B. & Zstray, S. (1994). *What twenty years of educational studies reveal about year-round education*. Chapel Hill, NC: North Carolina Educational Policy Research Center
- Weaving Gender Equity into Math Reform. (2001). *Measuring student achievement: The impact of standardized testing on equity and excellence in mathematics* [Powerpoint]. TERC at Cambridge, MA. Retrieved September 30, 2007, from <http://wge.terc.edu/testingession.html>

Weiss, S. (2005). The growing popularity of and investment in after-school programs heightens the need for better information and closer study. *The Progress of Education Reform 2005: After-School Programs*, 6, 1-5.

Zimmerman, J. (1998). *Improving student achievement by extending school: Is it just a matter of time?* San Francisco, CA: WestEd.