1-1-2005

# The Effects of an After-School Intervention Program on the Reading and Math Proficiency Scores of Sixth Graders

Laura Kelley Gleichauf

Running head: EFFECTS OF AN AFTER-SCHOOL PROGRAM ON READING AND MATH

# The Effects of an After-School Intervention Program on the Reading and Math Proficiency Scores of Sixth Graders

Thesis submitted to
Marshall University Graduate College

In partial fulfillment of the
requirements for the degree of
Master of Education Specialist
School Psychology

by

Laura Kelley Gleichauf

Fred Jay Krieg, Ph.D, Committee Chairperson
Stephen L. O'Keefe, Ph.D.
Sandra Stroebel, Ph.D.

Marshall University Graduate College

August 2005

Abstract

The Effects of an After-School Intervention Program

on the Reading and Math Proficiency Scores of Sixth Graders

Laura Kelley Gleichauf

In January 2005, a new after-school intervention program was implemented by a public middle school with the desire to raise their students' state proficiency scores. In order to evaluate their program and report the results of their labor, the proficiency scores of these southern Ohio students were studied by comparing the arithmetic means of the March 2003 fourth grade state proficiency scores with the arithmetic means of the March 2005 sixth grade state proficiency scores in the areas of reading and math. Differences in the performances were compared using t-tests. The outcome data of the comparison school was not successful in demonstrating a statistically significant increase in state proficiency test scores. Topics including: effective instruction, program evaluation, after-school intervention, and state academic testing were explored.

Acknowledgements

I wish to thank my family and friends who have provided support and encouragement throughout this educational journey. Their thoughtfulness was especially appreciated when the road appeared perilous. An enormous thank you must go to my compassionate husband, Carlo, and our wonderful children, Logan and Alaina. A special thanks to my mentor, Estella Curry, whose unwavering confidence in me has fostered endurance. I am deeply grateful for Dr. Fred Jay Krieg and his pioneering, generous spirit. His continuing quest to learn more and tirelessly teach the most recent research based information is inspiring. I would also like to express gratitude to Dr. Stephen O'Keefe and Dr. Sandra Stroebel for their guidance and participation on the committee. Finally, I would like to thank the South Point and Ironton, Ohio school administrators for their willingness to provide data and participate in a research study. The cooperation and reports provided by the Sandra Mers, Assistant to the Principal of South Point Middle School; the approval of participation by Dean Nance, Ironton City Schools Superintendent; and the assistance and reports provided by Jennifer Golighlty, Ironton Middle School Guidance Counselor, made this project possible.

Table of Contents

# List of Tables

# The Effects of an After-School Intervention Program on the Reading and Math Proficiency Scores of Sixth Graders

## CHAPTER ONE

## Introduction

Raising high stakes test scores is a hot topic among school administrators and teachers. What tasks are schools undertaking to effectively raise test scores? The answer is a valuable commodity to school personnel because test scores translate into public perception of the quality of instruction provided by that school or district. Test scores should be only one of the factors considered in evaluating students' academic performance. However, loses or gains in federal and state funding including Title 1 funds, bonus pay and parents' decisions regarding enrollment can be based upon the reported mean performance of students in a certain district. Furthermore, proficiency tests have become very high stakes tests for individual students as well. Instead of evaluating a student's overall achievement, grade promotion and even graduation have been linked to a student's single standardized exam score.

No Child Left Behind mandated every state to have a testing program by 2005 (Great schools, 2004). Currently, students in grades three through eight must be tested annually in reading and math. In order to receive or maintain funding based on test scores, schools are searching for ways to increase their students' test results. One school district in southern Ohio was not satisfied with the scores of their March 2003 fourth graders and implemented an after-school intervention program to raise their mean state proficiency score for the same group of students as March 2005 sixth graders. This study afforded a means of evaluating the success of the program and allowed the administrators to assess the program for continuation, enhancement, alteration or deletion.

## Statement of the Problem

Can an after-school program raise proficiency test scores? The purpose of this study was to evaluate the effects of an after-school intervention program on students' sixth grade Ohio proficiency scores. Did the evidence indicate that the 2005 mean of sixth grade student proficiency scores were higher than the 2003 mean of fourth grade student proficiency scores?

The null research hypothesis was: The arithmetic means of Ohio proficiency math and reading test scores of South Point, OH, students will not differ due to after school intervention program when comparing the 2003 fourth grade scores to the 2005 sixth grade scores. The research hypothesis was: The arithmetic mean of the 2005 sixth grade Ohio proficiency math and reading test scores for South Point, OH, will be statistically significantly different than the arithmetic mean of the 2003 fourth grade Ohio proficiency math and reading test scores. The independent variable was the intervention programs. The independent values were math and reading. The dependent variable was the student performance.

Another school within the same county served as a control group. The control group did not receive any after school academic intervention. The performance of the control group students was evaluated by comparing the arithmetic means of Ohio proficiency math and reading test scores of Ironton, OH students of 2003 fourth grade scores to the 2005 sixth grade scores in the same manner as the above comparison group method described.

## Significance of the Problem

If the after-school intervention was successful in raising test scores, then South Point may consider repeating this after-school program in the future. If South Point's after school program was not successful in raising test scores, then South Point may not wish to repeat the intervention

in the same manner in the future. Sixth grade school administrators and parents of middle school children could consider using this project data as a reference for future decision making.

## Definition of Terms

The following definitions of terms were used to examine the research question of this study:

1. Academic Content Standards- "describe the *knowledge and skills* that students should attain - often called the "what" of "what students should know and be able to do." They indicate the ways of thinking, working, communicating, reasoning, and investigating, and important and enduring ideas, concepts, issues, dilemmas, and knowledge essential to the discipline" (Ohio Department of Education, 2001).

2. Adequate Yearly Progress (AYP) - The federal mandate that holds schools accountable for the performance of subgroups, as well as all students. AYP measurements focus on the performance and participation of various groups of students, based on race or ethnicity, socioeconomic status, disability and language background. AYP is measured by reviewing proficiency test scores in reading and math for all enrolled students. Schools need to focus efforts on the improvement of all students, and also in each of these subgroups in order to make AYP. The goal is to have all students at or above proficient by 2013-14 in reading and math (Ohio Department of Education, 2005).

3. After-School Program- From 3 o'clock until 5 o'clock on regularly scheduled school days, supplemental math and reading instruction offered to a pre-selected 35 students by middle school teachers. Curriculum and method of instruction were at the discretion of the subject teacher.

4. Mean- The arithmetic mean of a set of data is found by taking the sum of the data and dividing by the total number of values in the set. The mean is commonly referred to as an average (Glosser, 2005).

5. Ohio Proficiency Exam- Ohio teachers and community members developed content standards in reading, writing, mathematics, science and social studies. These standards were intended to tell teachers what facts, ideas, formulas, etc., to teach. The Ohio Proficiency Exam is a test designed as a measure to determine if students are being taught and are learning the intended content material. "Ohio law requires students in grades 4, 6, 9 and 12 to take proficiency tests that are designed to measure knowledge and skill in citizenship, mathematics, reading, science, and writing" (Ohio proficiency test resource center, n.d.).

6. Proficiency- The state or quality of being proficient. Further, proficient is defined as- to advance highly, to be highly competent or skilled (Simon & Schuster, 1979).

7. Standard Deviation- The standard deviation is a statistic that tells you how tightly all the various examples are clustered around the mean in a set of data. When the examples are pretty tightly bunched together and the bell-shaped curve is steep, the standard deviation is small. When the examples are spread apart and the bell curve is relatively flat, the standard deviation is relatively large (Niles, 2005).

## Limitations and Delimitations

This study investigated only the proficiency test performance of March 2003 fourth grade students and March 2005 sixth grade students enrolled in two Appalachian school districts in Lawrence County, Ohio. The study focused upon proficiency test results pertaining to these two time periods for data collection. The study was longitudinal; therefore, data was affected by

students in the sample who left the school and new students who came into the studied group groups.

Changes in the level of difficulty expected for a sixth grade test as opposed to the expected level of difficulty of a fourth grade proficiency test cannot be controlled. This researcher had to assume that the test manufacturers maintained the level of difficulty appropriate to assess content standard knowledge for the sixth grade proficiency exam in 2005 that was consistent with the level of difficulty appropriate to assess the academic content standard for the fourth grade proficiency exam in 2003.

# CHAPTER TWO

## Literature Review

In this literature review, research regarding effective instruction and program evaluation will first be presented. Studies that explored after-school programs and finally state academic testing will then be discussed.

### Effective Instruction

Mastropieri, Scruggs, and Graetz, (2003) further investigated work reported by Mastropieri, Scruggs, Bakken and Whedon in 1996 which identified a set of common instructional features for effective reading interventions. A list of effective intervention components in their article included: 1) use clear objectives; 2) follow a specific sequence for teaching, such as, a. state the purpose, b. provide instruction, c. model, d. guided practice, e. corrective feedback, f. independent practice, g. generalized practice; 3) inform students of the importance of the strategy; 4) monitor performance; 5) encourage students to think about strategies and text; 6) encourage appropriate attributions; and 7) teach for the generalized use of the strategy (Mastropieri, Scruggs, & Graetz, 2003).

Jean Clyde's (2003) article, "Stepping Inside the Story World," offered another strategy for effective instruction. Through the use of visual imagery, teachers can enrich their students' literacy learning and comprehension. If students imagined what the characters might be thinking, the students can better comprehend the story. As a result, the student may become a more interested reader.

### Program Evaluation

Moore (2003) recommended initiating a teaching program evaluation in the middle of the school year. A mid-year evaluation can provide valuable information as to whether the program

is moving in positive directions. It can also help answer questions and concerns related to the program. Various questions which can be addressed include: Is the program moving forward smoothly for students their families and staff? Is there positive momentum with the program? Can the goals be met? In order to answer questions such as these, a mid-year evaluation is a perfect tool (Moore, 2003).

A program evaluation in the middle of the year serves may purposes, which include:

A. Giving an accurate read of how family and staff are feeling.

B. Allowing shy, reticent members to respond by writing anonymously.

C. Giving time to make improvements.

D. Improving how families perceive your commitment.

E. Helping to reduce staff and family turnover (Moore, 2003).

Asking for feedback can be perceived as more work when an administrator feels overwhelmed. Yet the process of gathering feedback causes everyone to stop and appraise what is happening around him or her. "This essential principle of 'customer satisfaction' is to ask and be prepared to discuss, modify and make changes to improve your program," (Moore, 2003).

Initiating the process of evaluation and welcoming comments, suggestions, questions and opinions was described by Moore (2003) as the sign of a proactive leader. Families and staff often respond with views and observations that may have been overlooked. After discussion and implementation, evaluations have the potential to transform a program for the better. Possible methods of evaluation include surveys with multiple-choice checklists that cover specific topics or verbal interviews that can be guided by a list of questions or thoughts about the program (Moore, 2003).

Beard, Strachan, Davies, Patterson, Stark, Ball, Taylor, and Thomas (2005) successfully used a different evaluation technique called questionnaires. Rather than waiting until mid-year, though, Beard et al. (2005) found it effective to begin the assessment process in the very early stages. With the desired outcome being an evaluation of the education and assessment framework for the second year of *Foundation Programme*, the expectations of the trainees and their satisfaction were evaluated by the responses to questionnaires administered before and at the end of the pilot program.

Whether evaluations are performed before, mid-year, or after program completion, Moore (2003) stated that an evaluation may cause concern that drastic program changes would be the result. Nothing could be farther from the truth. The goal of evaluation according to Moore, was to gain valuable information during the year and that resulting little changes would make a big difference. Moore recommended an end of the year evaluation as well to see if the modifications made were beneficial. Surveys are a simple and effective tool for end of the year evaluations. Surveys given to teachers and parents can be used to obtain information concerning what worked, what didn't work, and improve levels of communication. The evaluation can provide concrete information regarding ways in which improvements can be made. Year end evaluation surveys are an integral part of the program evaluation process (Moore, 2003).

## After School Intervention

In Dodd and Wise's 2002 article regarding extended day school programs they stated that research has confirmed the notion that some students take three to six times longer to learn than the average student. An extension of learning time for students was recommended as a way to bridge the learning gap and provide students the additional time needed to master subject content (Dodd & Wise, 2002). Dodd and Wise (2002) found that extended day activities positively

affected the achievement of participating, low-achieving students in terms of passing grades, higher grades and/or better tests scores. However, it is not the time itself that makes the difference. Quality instruction was necessary in the extension of the school day to have an effect on student achievement. Higher success rates were found when students participated in extended day programs that were similar to their regular school day programs. Programs that link the regular school day with the extended day program by using similar, but slightly different academic strategies have better success rates than programs without linkage (Dodd & Wise, 2002).

*Preparation*

Specialized training for the staff involved in the extended day program includes a staff development program. Dodd and Wise (2002) stated that these programs should provide planning time, evaluation time and professional growth by staff members. Teachers with extensive knowledge of how to work with at-risk students will help the program be more successful. Program evaluation and continuous monitoring is vital to successful practices and outcomes. Student assessment data and evaluations for all participants including students, teachers and program staff were recommended to provide ongoing information (Dodd & Wise, 2002).

*Planning*

According to Lauer (2003), planning for an after school program is not an easy task. Providing an extra two or more hours of high quality, content rich enjoyable learning can be difficult. After six hours of a school day both student and staff energy levels are reduced and finding teachers willing to take on additional responsibilities, administrative staff to supervise the program and budgetary concerns compound the challenge. Lauer wrote that district level

administrators, parents, students and community leaders all should have a voice in the process of planning. Lauer (2003) recommended addressing four questions in planning an after school program. These included:

1.) "What should be included in start-up planning?

2.) What role does the principal play?

3.) How do you recruit, train and retain high quality staff?

4.) How can a successful program be sustained?"

Sustaining a successful program within an environment of decreased funding and dwindling resources poses an ever-changing challenge to school systems (Lauer, 2003).

## State Academic Testing

Testing has played an integral part of everyday life in the United States for well over one hundred years. Tests have determined which immigrants could enter the United States at the turn of the 20th Century, who was intellectually gifted, who needed special education, who received college scholarships, as well as who could serve in the military (Amrein, & Berliner, 2003). Annual testing of students in grades 3-8 in reading and math is now mandated due to the federal act, No Child Left Behind Act of 2001(Amrein, & Berliner, 2003).

As of 2003 eighteen states including Ohio use a system of testing by which students, teachers and schools are evaluated and compared (Amrein and Beliner, 2003). Many of these eighteen states attach a broad range of consequences to those assessments for the students, teachers and the school system. These assessments and subsequent consequences have created "high-stakes" testing (Amrein, & Berliner, 2003). For example, tests have been used to determine whether a student is to be granted a diploma or if a diploma is to be withheld (Amrein, & Berliner, 2003).

Amrein & Berliner (2003) reported that state implementation of high stakes testing often leads to an improvement of student scores on the state assessment exams. Students can be coached on specific skills or information anticipated to appear on the exam to increase scores on the state exams. By spending time on only those skills anticipated to appear on the exam, the curriculum is narrowed and students can increase their scores on the state exams (Amrein, & Berliner, 2003). However, student achievement in the eighteen high-stakes testing states has not improved on a range of measures, such as the National Assessment of Education Progress, despite higher scores on the state's own assessments. Although state scores can be raised with a narrowing of curriculum, student learning as assessed on SAT, ACT, Advanced Placement (AP) and National Assessment of Education Progress (NAED) showed no improvement in student learning (Amrein, & Berliner, 2003). Therefore, Amrien and Berliver concluded that in the eighteen states with high stakes testing policies, as indicated by four independent measures, no measurable improvement in student learning resulted (Amrein, & Berliner, 2003).

Furthermore, evidence shows that such tests actually decreased student motivation and increased the proportion of student who left school early (Amrein, & Berliner, 2003). Attaching stakes to tests apparently obstructed students' path to becoming life long self directed learners and alienated students from their own learning experience in school. It appeared that high-stakes tests did not motivate students; in fact such tests decrease student motivation and lead to higher student retention and drop out rates (Amrein, & Berliner, 2003).

Hoff (2005) reported further evidence highlighting a downside to high-stakes testing in his *Education Week* report, "High-stakes testing produces incentives for cheating, harms teacher morale and results in important decisions about student's lives being made based on occasionally faulty data." In a review of newspaper articles, two researchers found examples of educators

cheating to improve test scores, of narrowing teaching to the test, and unfair testing practices such as requiring students to take tests shortly after a siblings murder (Hoff, 2005).

## Evaluating Scores

Nataraj., McCaffrey, Lockwood, Sloan, Naftel, & Barney (2002) discussed a real threat to the evaluation of state scores of student achievement involves the reliability and validity of school level test scores. Nataraj et al. defines reliability in terms of test scores in that the "reliability of a score is the extent to which it yields the same results on repeated trials" (Nataraj et al., 2002). Aggregate school-level data are susceptible to random errors. "Every error in student scores contributes to errors in aggregated scores" (Nataraj et al., 2002). Student level errors tend to balance out in aggregate scores and group aggregate scores will often be of greater test reliability than individual scores (Nataraj et al., 2002). "Student achievement varies within schools and this contributes to the variability in school level measures" (Nataraj et al., 2002).

Nataraj et al. (2002) found that relative to variability between schools, the variability among students within a school was large. They found that school-to-school variance accounts for only about 11%-13% of the total variance in fourth grade reading or math scores in North Carolina and they estimated that "the confidence interval for the average fourth-grade reading or math score in a school with sixty-eight students per grade level would extend roughly from the $25^{th}$ to the $75^{th}$ percentile among schools of that size" (Nataraj et al., 2002). Programs are designed to affect school performance, not reduce variability errors. Therefore, small samples of program schools will not provide reliable estimates of program effects (Nataraj et al., 2002). Nataraj et al.(2002) discussed Kane and Staiger's report which estimated that for the average-size schools, "when looking at gains between two grade levels (e.g. fourth to fifth grade) 49% of the variance between schools was due to sampling variation and other non persistent factors.

This estimate increased to 73% when looking at change scores (e.g. comparing fourth-grade scores over time)."

Validity of research data can be defined as "an essential quality in research data having to do with whether the data are, in fact, what they are believed or purported to be" (Mertler & Charles, 2005). This means valid data come from measuring what you wanted to measure based upon the primary purpose or focus of the research (Mertler & Charles, 2005).

Nataraj et al. (2002) warned that school scores may not reflect student learning. Nataraj et al. (2002) stated that when high-stakes were attached to performance, inflation of test scores were a high threat to validity. When high stakes were involved, scores may have increased more than actual learning as a result of teaching narrowly to the test, cheating, or coaching (Nataraj et al., 2002). "Inflation of gains can arise even in the absence of inappropriate practices or high stakes. For example, test scores typically fall when a new test is introduced, and then rise rapidly over time as rapidly as students and teachers become increasingly familiar with the test" (Nataraj et al., 2002).

Nataraj et al. (2002) also reported a second issue affecting learning as contamination by other factors, such as average initial achievement of students entering first grade. The change in composition of cohorts also makes it difficult to measure the contribution of the schools. A third contaminating factor is high student mobility that may reflect the contributions of several schools rather than the particular school where the student took the test (Nataraj et al., 2002).

Further limitations of state school accountability data included the metrics by which data was reported. Each reporting metric whether it is mean scale scores, median or mean national percentile ranks, or percentage of students reaching one of several pre-determined performance

standards has limitations as an outcome measure in evaluation of school programs (Nataraj et al., 2002).

School-level averages reported in raw scores are sensitive to the performance of all students. Percentiles or percent passing do not reflect the performance of all students tested. As a result, small changes will be reflected in the schools' mean score, however this sensitivity comes at the price of increased variability (Nataraj et al., 2002).

## Conclusion

In conclusion, the above literature related to the proposed research hypotheses and questions of this study by reiterating the prevalence of high stakes testing and the importance of well–planned, effective after school programs as they can have immense impact on student test scores. As school systems are judged, and even funded, based upon proficiencies of their students, the importance of an intervention program and the assessment of its effectiveness made the research hypothesis and the answers it provided of importance to the students, teachers, and administration as they evaluate, modify, and strive to improve their program to attain an admirable goal of not only improving student test scores, but also increasing their learning and achievement.

# CHAPTER THREE

## Design of the Study

This chapter will describe the research procedures used to evaluate the effectiveness of an after-school program implemented to raise proficiency test scores of sixth grade students.

## Research Design

The primary purpose of this quantitative study was to evaluate the effects of an after-school intervention program on students' sixth grade Ohio proficiency scores. The researcher evaluated if the evidence indicated that the 2005 means of sixth grade student proficiency scores were significantly statistically different when compared with the 2003 means of fourth grade student proficiency scores in the areas of reading and math. The comparison group study was longitudinal comparing the proficiency exam scores of the South Point, Ohio, students over a two year period. The control group study was also longitudinal comparing the proficiency exam scores of the Ironton, Ohio, students over a two year period.

## Population

The subjects for this study consisted of 2003 fourth grade students and 2005 sixth grade students enrolled in two Appalachian school districts in Ohio. The study examined test results obtained from two proficiency exams occurring over a two year period, one fourth grade test in March 2003 and one sixth grade test in March 2005.

The control sample consisted of all Ironton March 2003 fourth graders tested and all March 2005 sixth graders tested. According to the Ohio Department of Education, all students enrolled in the named school for 120 consecutive days are included in the yearly testing report card data for that particular school (ODE, n.d, p 4). As a result, the researcher compared the Ohio proficiency exam test results of the corresponding Ironton students. In the 2003 control group

there were 68 fourth grade students tested in the area of Reading and 108 fourth grade students tested in the area of Mathematics. The gender breakdown of the participants was not available at the time of the study. The most recent population information available was from the Ironton Middle School 2002-2003 School Year Report Card Building Report (ODE, n.d.) which contains the comparison group. The ethnicity breakdown of the middle school (grades 4 through 6) consists of 89.7% Caucasian, 7.4% African-American, and 2.9% Multi-racial. 56.4 % of the students are economically disadvantaged (ODE, n.d.). Of this fourth through sixth grade population, 19.3 % of the students have disabilities (ODE, n.d.). The 2005 control group of Ironton sixth graders numbered 132.

The comparison sample consisted of all March 2003 fourth graders and all March 2005 sixth graders tested in the South Point district. As a result, the researcher compared the Ohio proficiency exam test results of largely the same South Point students. South Point has two elementary schools serving children in grades Kindergarten through Fifth grade. Therefore, the students comprising the comparison group attended either Burlington Elementary or South Point Elementary. The gender breakdown of the participants was not available.

The most up-to-date population information available was from the Burlington Elementary School and South Point Elementary School 2002-2003 Year Report Cards which contains the comparison sample (ODE, n.d.) as they would have been in fourth grade. The ethnicity breakdown of the Burlington Elementary School consisted of 89.2% Caucasian and 9.8% African-American. Fifty-six percent of the students were economically disadvantaged. Of the 299 students enrolled, 19.5 % of Burlington Elementary School students had disabilities (ODE, n.d.).

The ethnicity breakdown of the South Point Elementary School consisted of 90.9% Caucasian, 5.3% African-American, 3.0% Multiracial, and 0.8 % was not reported (ODE, n.d.). 45.9% of the students were economically disadvantaged. Of the 554 students enrolled, 19 % of South Point Elementary School students had disabilities (ODE, n.d.).

In the 2003 comparison group there were 126 fourth grade students tested in both areas, Reading and Mathematics. The 2005 comparison group of sixth graders numbered 128.

## Sample Selection

Within the comparison sample was a subgroup of 35 subjects that participated in the after-school program to be evaluated. In December 2004, every sixth grade student in the comparison population was given a pre-test comprised of forty-six questions directly copied from previous Ohio sixth-grade proficiency exams. The resulting scores were rank ordered and divided into seven groups, with groups one and two comprising the lowest scores and groups six and seven comprising the highest scores. The lowest two groups were not considered because it was decided by school administrators that two months of additional instruction would not raise those students' scores sufficiently to achieve proficiency. The high achieving and passing groups were not considered to receive additional services due to lack of need for intervention. The students whose scores fell into groups three and four were selected to receive after-school additional instruction in reading and math. Their scores fell below proficient, but were close enough to warrant consideration for after-school academic intervention in the areas of reading and math to become proficient.

## Data Collection

The data for this study was collected by examining the records for the March 2003 fourth grade Ohio proficiency exam scores and March 2005 sixth grade Ohio proficiency exam

scores**.** Parent permission to view individual student records was not necessary because the data collected was the group arithmetic mean and standard deviation. Student confidentiality was maintained throughout the study as individual student data was not investigated, recorded or reported.

The South Point middle school assistant to the principal provided the researcher with copies of the frequency distributions pertaining to the South Point March 2003 fourth grade Ohio proficiency scores and the 2005 sixth grade Ohio proficiency scores. The Ironton middle school guidance counselor provided the researcher with copies of the frequency distributions pertaining to the Ironton March 2003 fourth grade Ohio proficiency scores and the 2005 sixth grade Ohio proficiency scores. The researcher used this information to test for significance.

Frequency distributions contain, by subject area, the mean scaled score and the standard deviation. The standard deviation was squared to obtain the variance value needed. The reported information was used to perform t-tests and evaluate any changes in proficiency scores.

## Method of Data Analysis

The researcher performed two sets of two t-tests, one set was to evaluate the significance between independent arithmetic means of the control group performance in the areas of Reading and Mathematics and another set of t-tests was to evaluate the significance between the independent arithmetic means of the comparison group in the areas of Reading and Mathematics. A t-test of independent means was chosen because the difference between the fourth grade performance and the sixth grade performance was being explored. Both groups were tested once as fourth graders and once as sixth graders.

Regarding each set of t-tests: One t-test was used to calculate any significant difference of the mean of the 2003 Reading scores and 2005 Reading scores. The second t-test was used to

calculate any difference of the mean of 2003 Mathematics scores and 2005 Mathematics scores. Mertler & Charles (2005) state that a t-test "analyzes the difference between the means of two groups, to determine whether the difference is significant- that is, whether the difference of two points can, or cannot, be attributed to chance errors made in selecting participants." A t-test is computed by using the mean for group one, the mean for group two, the number of participants in group one, the number of participants in group two, the variance for group one and the variance for group two. The variance is calculated by squaring the standard deviation. The outcome of the computation will result in the ability to determine if the null hypothesis should be rejected or accepted (Mertler & Charles, 2005).

## Results

The first set of t-tests pertained to the control group data. In the area of Reading for the control group, the mean scaled score for Group 1 (2003 Ironton fourth grade) was equal to 210. The mean scaled score for Group 2 (2005 Ironton sixth grade) was equal to 240. The number of students (N) in Group 1 was 68. The number of students (N) in Group2 was 132. The variance for Group 1 was 324. The variance for Group 2 was 1,369. The obtained value is t= -6.307125529. The value was negative since a larger value was subtracted from a smaller number (Group 1 minus Group 2). At the .05 significance level, with 198 degrees of freedom, the value needed for rejection of the null hypothesis on a two –tailed test was 1.960. The obtained value was more extreme than the critical value; therefore the null hypothesis regarding the control group in the area of Reading should not be accepted. When comparing Ironton's 2003 fourth grade proficiency scores to Ironton's sixth grade proficiency scores in the area of Reading, a statistically significant difference existed. The proficiency scores improved.

In the area of Mathematics for the control group, the mean scaled score for Group 1 (2003 Ironton fourth grade) was equal to 238. The mean scaled score for Group 2 (2005 Ironton sixth grade) was equal to 212. The number of students (N) in Group 1 was 108. The number of students (N) in Group2 was 132. The variance for Group 1 was 1089. The variance for Group 2 was 1,225. The obtained value was t= 5.873767205. The value was positive since a smaller value was subtracted from a larger number (Group 1 minus Group 2). At the .05 significance level, with 238 degrees of freedom, the value needed for rejection of the null hypothesis on a two – tailed test was 1.960. The obtained value was more extreme than the critical value; therefore the null hypothesis regarding the control group in the area of Mathematics could not be accepted. When comparing Ironton's 2003 fourth grade proficiency scores to Ironton's 2005 sixth grade proficiency scores in the area of Mathematics, a statistically significant difference did exist. For the control group in the area of Mathematics, proficiency scores were lower in the 2005 than they were in 2003.

The second set of t-tests pertained to the comparison group data. In the area of Reading for the comparison group, the mean scaled score for Group 1 (2003 South Point fourth grade) was equal to 226. The mean scaled score for Group 2 (2005 South Point sixth grade) was equal to 232. The number of students (N) in Group 1 was 126. The number of students (N) in Group2 was 128. The variance for Group 1 was 361. The variance for Group 2 was 1,269. The obtained value was t= -1.6573266. The value was negative since a larger value was subtracted from a smaller number (Group 1 minus Group 2). At the .05 significance level, with 252 degrees of freedom, the value needed for rejection of the null hypothesis on a two –tailed test is 1.960. The obtained value did not exceed the critical value; therefore the null hypothesis regarding the comparison group in the area of Reading cannot be rejected and may be accepted. The obtained

value was not extreme enough to determine that in the area of Reading, the difference between the performance of the 2003 South Point fourth grade students and the performance of the 2005 South Point sixth grade students occurred by anything other than chance.

In the area of Mathematics for the comparison group, the mean scaled score for Group 1 (2003 South Point fourth grade) was equal to 233. The mean scaled score for Group 2 (2005 South Point sixth grade) was equal to 210. The number of students (N) in Group 1 was 126. The number of students (N) in Group2 was 128. The variance for Group 1 was 784. The variance for Group 2 was 1,444. The obtained value was t= 5.484647471. The value was positive because a smaller value was subtracted from a larger number (Group 1 minus Group 2). At the .05 significance level, with 252 degrees of freedom, the value needed for rejection of the null hypothesis on a two –tailed test was 1.960. The obtained value is more extreme than the critical value; therefore the null hypothesis regarding the comparison group in the area of Mathematics could not be accepted. When comparing South Point's 2003 fourth grade proficiency scores to South Point's 2005 sixth grade proficiency scores in the area of Mathematics, a statistically significant difference did exist. For the comparison group in the area of Mathematics, proficiency scores were lower in the 2005 than they were in 2003.

The following tables facilitate quick comparisons between performances by year/grade tested and comparisons between the control group and the comparison group.

Control Group Statistics*

| | Mean Scaled Score | Standard Deviation | N | Variance |
|---|---|---|---|---|
| 2003 4th grade Reading | 210 | 18 | 68 | 324 |
| 2005 6th grade Reading | 240 | 37 | 132 | 1369 |
| 2003 4th grade Math | 238 | 33 | 108 | 1089 |
| 2005 6th grade Math | 212 | 35 | 132 | 1225 |

**2003 Ohio Fourth Grade Proficiency Tests Frequency Distributions & the 2005 Ohio Fifth Grade Proficiency Tests Frequency Distributions reported by district.*

Comparison Group Statistics*

| | Mean Scaled Score | Standard Deviation | N | Variance |
|---|---|---|---|---|
| 2003 4th grade Reading | 226 | 19 | 126 | 361 |
| 2005 6th grade Reading | 232 | 36 | 128 | 1296 |
| 2003 4th grade Math | 233 | 28 | 126 | 784 |
| 2005 6th grade Math | 210 | 38 | 128 | 1444 |

**2003 Ohio Fourth Grade Proficiency Tests Frequency Distributions & the 2005 Ohio Fifth Grade Proficiency Tests Frequency Distributions reported by district.*

# CHAPTER FOUR

## Findings and Discussion

It was shown in the control group data that for the students tested in the fourth grade and two years later in sixth grade, a statistically significant difference in Reading existed. The result of comparing mean scaled scores indicated an overall improvement in group performance on the Ohio Proficiency Test in Reading. The results for the control group also indicated that for the students tested in the fourth grade and two years later in sixth grade, a statistically significant difference in Mathematics existed. However, this result indicated an overall decline in group performance on the Ohio Proficiency Test. The percentage of students at or above the proficient level decreased in both comparisons.

Control Group Results

| | Mean Scaled Score | Median Scaled Score | State Max Scaled Score | % At or Above Proficient | N |
|---|---|---|---|---|---|
| 2003 4th grade Reading | 210 | 213 | 268 | 78% | 68 |
| 2005 6th grade Reading | 240 | 243 | 349 | 75% | 132 |
| Result | higher | | | decrease 3 % | |
| 2003 4th grade Math | 238 | 248 | 320 | 77% | 108 |
| 2005 6th grade Math | 212 | 215 | 368 | 66 % | 132 |
| Result | lower | | | decrease 11% | |

*Information compiled from ODE Public 4th reading 02 vs 03; March 2005 preliminary results, 2003 Ohio Fourth Grade Proficiency Tests Frequency Distributions and the 2005 Ohio Fifth Grade Proficiency Tests Frequency Distributions reported by district.*

It was shown in the comparison group data that for the students tested in the fourth grade and two years later in sixth grade, a statistically significant difference in Reading did not exist. This result indicated that the effort put forth by the sixth grade students, teachers and

administrators to improve the test scores of the sixth grade students did not produce statistically significant overall change in group performance on the Ohio Proficiency Test. However, the results did reveal in the comparison group data that for the students tested in the fourth grade and two years later in sixth grade, the outcome was a statistically significant difference in Mathematics. Unfortunately, however, this result indicated an overall decrease in group performance on the Ohio Proficiency Test. The percentage of students at or above the proficient level decreased in both comparisons.

Comparison Group Results

| | Mean Scaled Score | Median Scaled Score | State Max. Scaled Score | % At or Above Proficient | N |
|---|---|---|---|---|---|
| 2003 4th grade Reading | 226 | 224 | 268 | 95% | 126 |
| 2005 6th grade Reading | 232 | 239 | 349 | 72% | 128 |
| Result | n/a | | | decrease 23% | |
| 2003 4th grade Math | 233 | 233 | 320 | 78% | 126 |
| 2005 6th grade Math | 210 | 211 | 368 | 63% | 128 |
| Result | lower | | | decrease 15% | |

*Information compiled from ODE, Public 4th reading 02 vs 03; March 2005 preliminary results, 2003 Ohio Fourth Grade Proficiency Tests Frequency Distributions and the 2005 Ohio Fifth Grade Proficiency Tests Frequency Distributions reported by district.*

## Description of Findings

The control group figures revealed greater gains than the comparison group figures in student performance on the Ohio Proficiency Test. While in one area (Mathematics) there was a decrease in overall group performance, in the other area (Reading), the control group findings did signal improvement. The comparison group figures revealed a statistically insignificant

difference in one area (Reading) and a decrease in overall group performance in the other area (Mathematics).

The results for both the control and the comparison groups generated implication questions. Considering that the control group and the comparison group Mathematic mean scaled scores both decreased, was the portion of the sixth grade Ohio proficiency test that evaluated Mathematics more difficult in 2005 than in previous years? Did it evaluate different skills than anticipated? What methods of Reading instruction resulted in gains for the control group? What methods of Reading instruction were not effective for the comparison group? Each answer would provide useful, tangible information. Furthermore, after reviewing the literature and the project results, the comparison school program administrators may consider asking themselves how well the after-school program was intertwined with the regular school day curriculum. Were different strategies used after school to teach the information covered earlier in the day? To promote positive ideas, this researcher would suggest discussion with the teachers regarding how the after-school program could have been more closely connected to the regular school day curriculum. Also, based on the students' performance, what specific topics within the areas of Reading and Mathematics need further explanation? What does the research say is an effective means of transferring that particular knowledge to the students to increase their understanding?

## Summary

The teachers and administrators of the comparison group demonstrated their desire to help their students perform better on the Ohio Proficiency Test by conducting two month after-school intervention project. It was the intent of this researcher to provide a compilation of the data in order to allow for evaluation of the after-school intervention program compared to the control group data and report the results of the comparison group's extra effort. The proficiency

scores of southern Ohio students were studied by comparing the means of the March 2003 fourth grade state proficiency scores with the means of the March 2005 sixth grade state proficiency scores in the areas of reading and math. Differences in the performances were compared using t-tests.

Had the comparison group data produced a statistically significant increase in performance in the areas of Reading and Math, then the comparison school might have considered using this information to replicate their design in years to come and distributed their design to other schools. However, the outcome data by the comparison school was not successful in demonstrating a statistically significant increase in test scores. In the area of Reading, the change in scores was insignificant. In the area of Mathematics, the change was a decrease in the proficiency scores. Therefore, the comparison school administrators and teachers might consider using this information as a seed from which to grow and a reference for future decision making. This southern Ohio district opened up its state test result records for evaluation to determine how well their after-school intervention program worked. Just as the literature reviewed earlier stated, evaluations provide valuable information. By utilizing the resulting data, the district may now consider exploring the methods of their after-school program design before implementing another after-school intervention program in an effort to raise Ohio proficiency test scores.

References

Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning [Electronic version]. *Educational Leadership, 60*, 32-38.

Beard, J., Strachan, A., Davies, H., Patterson, F., Stark, P., Ball, S., Taylor, P., & Thomas, S. (2005). Developing an education and assessment framework for the foundation programme. *Medical Education*, 39(8), 841- 852.

Clyde, J.A. (2003). Stepping inside the story world: the subtext strategy- a tool for connecting and comprehending. *The Reading Teacher*, 57(1), 150-160.

Dodd, C., & Wise, D. (2002). Extended-day programs: time to learn [Electronic version]. *Leadership, 32* (1), 24-26.

Glosser, G. (2005). *Mrs. Glosser's Math Goodies*. Retrieved on July 17, 2005 from http://www.mathgoodies.com/lessons/vol8/advanced_mean.html.

Great School, Inc. (2004). *About test scores in Ohio*. Retrieved July 16, 2005 from http://www.greatschools.net/issues/oh/ach.html.

Hoff, D. J. (2005). High-stakes testing [Electronic version]. *Education Week*, 24 (29), 1.

Lauer, R. H. (2003). After-school programs: everybody's doing it! *Principal, 82*, 28-32.

Lusky, M. B., & Hayes, R. L. (2001). Collaborative consultation and program evaluation [Electronic version]. *Journal of Counseling & Development, 79* (1), 26-39.

Mastropieri, M. A., Scruggs, T. E.; Graetz, J. E. (2003). Reading comprehension instruction for secondary students: challenges for struggling students and teachers [Electronic version]. *Learning Disability Quarterly, 26* (2), 103-117.

McKechnie, J.L.(Ed.) (1979). *Webster's New Twentieth Century Dictionary* (2nd ed.). New York: Simon & Schuster.

Mertler, C. A. & Charles, C. M.(2005). *Introduction to educational research* (5th ed.). Boston: Allyn & Bacon.

Moore, K. B. (2003) Evaluating your program at mid-year [Electronic version]. *Early Childhood Today, 17* (4), 12-14.

Nataraj Kirby, S., McCaffrey, D. F., Lockwood, J. R., Sloan McCombs, J., Naftel, S., & Barney, H. (2002). Using state school accountability data to evaluate federal programs: a long uphill road [Electronic version]. *Peabody Journal of Education, 77* (4) 122- 146.

Niles, R. (2005)*Standard deviation*. Retrieved on July 15, 2005 from http://nilesonline.com/stats/stdev.shtml.

Ohio Department of Education (2005).*List of acronyms.* Retrieved on July 15, 2005 from http://www.ode.state.oh.us/employee/acronyms.asp.

Ohio Department of Education (2001). *Academic Content Standards Frequently Asked Questions.* Retrieved on July 15, 2005 from http://www.ode.state.oh.us/academic_content_standards/acsfaq.asp

Ohio Department of Education (n.d.) *Burlington Elementary School 2002-2003 School Year Report Card*. Retrieved on August 8, 2005 from http://www.ode.state.oh.us/reportcardfiles/2002-2003/BUILD/004317.pdf.

Ohio Department of Education (n.d.) *Ironton middle school 2002-2003 school year report card*. Retrieved on August 8, 2005 from http://www.ode.state.oh.us/reportcardfiles/2002-2003/BUILD/033290.pdf.

Ohio Department of Education (n.d.) *Lawrence county ohio sixth grade proficiency tests grade 6 preliminary results-public schools march 2005 test administration.* Retrieved on June 20, 2005 from

https://webapp1.ode.state.oh.us/proficiency_reports/data/csvtoasp.asp?filename=g6pub6805.csv&county=lawrence.

Ohio Department of Education (n.d.), *Public 4th reading 02 vs 03*. Retrieved on June 20, 2005 from https://www.ode.state.oh.us/news/onlinemediaroom/march2003proficiency/P4THREADING0203.xls.

Ohio Department of Education (n.d.) *South point elementary school 2002-2003 school year report card*. Retrieved on April 10, 2005 from http://www.ode.state.oh.us/reportcardfiles/2002-2003/BUILD/033274.pdf.