University of Nevada, Reno

An Alternative Approach to School Accountability: Measuring School Achievement with Student Growth Percentile (SGP)

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in
Curriculum, Teaching and Learning

by

Selcuk Ozdemir

Dr. Pamela Cantrell/Dissertation Advisor

2012

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

SELCUK OZDEMIR

entitled

**An Alternative Approach to School Accountability:
Measuring School Achievement with Student Growth Percentile (SGP)**

be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

____
Pamela Cantrell, Ph.D., Advisor

____
Bob Ives, Ph.D., Committee Member

____
Bill Thornton, Ph.D., Committee Member

____
Steve Rock, Ph.D., Committee Member

____
Christine Cheney, Ph.D., Graduate School Representative

____
Marsha H. Read, Ph. D., Associate Dean, Graduate School

May 2012

ABSTRACT

A serious challenge for many schools and districts across the nation is that significant numbers of students enter the next grade level with performance levels well below proficiency. Since the implementation of NCLB, states have been primarily using Status Models to report AYP results. Status Models rely on a single year's assessment as an indicator of how many students are achieving proficiency on academic content standards. Essentially, Status Models are a count of the number of students meeting a specified target. The basic question under this model is: On average how are students performing this year? Comparisons over time are made by comparing the results for the students at each grade level one year with the results for students who were at the same grade level in previous years.

One of today's significant topics in educational curriculum and administration is how to measure student improvement. There are many important implications for this measurement system. A reliable and accurate measurement system, for example, would enable evaluators to observe what teachers are making a difference as well as what type of curriculum and learning environment best promote student learning. In an era of educational accountability, the topic of how to accurately measure student progress is an important hinge upon which the door of improving education turns.

Over the past decade, educational researchers and policy makers have endeavored to identify valuable growth models, methods that measure the growth of students' skills, knowledge, and abilities acquired in public school systems over time (United States Department of Education, 2005). In contrast with status models ─ the bases for adequate yearly progress (AYP) under No Child Left Behind (NCLB, 2001) ─ growth models describe progress in student performance rather than a fixed measure of current-state achievement. Because many current

accountability systems have been limited to the use of status designs, growth models appeal to educators and policy makers alike.

While Growth Models have been gaining more widespread use for measuring student achievement in primary and secondary education, school personnel may not have the background to understand these models. This paper reviews the continuing research and applications of statistical Growth Models. We describe the current convention of using descriptive Status Models and review the advantages and limitations of these models. Growth Models are then described as a statistical analytic technique to address the limitations of Status Models. Four of the most common Growth Models currently being implemented include the Value Table, Growth to Proficiency, Value-Added and Student Growth Percentile Models. Each of these Growth Models is described and the advantages and disadvantages of these models are examined.

Estimating and understanding how students grow in learning is perhaps the most important goal for any educational system. The nature of Status Models (using single year's assessment as an indicator) makes it impossible to determine whether or not the score represents normal academic progress. Growth Models have been proposed as an alternative to Status Models that address these limitations.

Growth Models offer practical benefits for states seeking to address the limitations of the Status Models that are currently used to assess student academic achievement and determine school accountability. Growth Models offer flexibility and improved measurement reliability when they are comparing data for the same students over time. The purpose of this dissertation is to investigate the Student Growth Percentile (SGP) as an alternative growth model approach to the measurement of student and school achievement. The study aims to build understanding of how Student Growth Percentile plays out with state and district data and how various definitions of growth impact schools. Longitudinal data from the statewide assessment program (Criterion Reference Test) from five years (N=300,000) will be used. Results will establish criterion-referenced interpretations of growth for the district schools. Findings will make important contributions to ongoing studies on the use of growth models in state accountability and will provide insight into processes and results of a criterion-referenced growth model.

Dedicated to my parents:

My mother Esin and my father Abdurrahman

And My Family

My wife Esmeray, and my children Gokcen Yagmur and Burkay Efe

ACKNOWLEDGEMENTS

I would never been able to finish my dissertation without guidance of my committee members, support from my family and help from my friends.

My graduate school experience and this dissertation were enabled by many people. I would like to express my sincere gratitude to all those who provided assistance. I would like to express my sincere gratitude to my advisor Dr. Pamela Cantrell for the continuous support of my Ph.D study and research, for her patience, friendship, enthusiasm, and vast knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Bob Ives, Dr. Christine Cheney, Dr. Bill Thornton and Dr. Steven Rock for their encouragement, insightful comments, and friendship.

Three other people deserve special recognition as mentors who supplied inspiration and coaching along the way. Several academics gave their time and insight to help me develop and advance my ideas. I must express my gratitude to Dr. Leping Liu, Dr. Cahit Evrensel and Dr. Gokhan Pekcan for being a wise, kind, and great mentor. I am also grateful for having an exceptional friend Huseyin Sahin, who supported me to succeed in this journey. His enthusiasm and a fresh perspective always keep me focused.

My deepest gratitude and appreciation of all goes to my family: my mother,
and my father who committed her life to my education. None of this would be possible without my family. Dad and mom, your support and sacrifice made me stronger and better. Without my family’s love and support I would not have completed this degree. My parents were always thinking of me and supporting me.

None of this would be possible without my wife Esmeray, who had live with a husband who had limited time to spend with her. Esmeray, you are the smartest person I know, and I can always put my full faith in you, knowing that I will get the best support and advice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER ONE

INTRODUCTION

The Elementary and Secondary Education Act (ESEA) of 1965 was passed as a part of President Lyndon B. Johnson's War on Poverty. ESEA appropriated federal money for states to improve educational opportunities for disadvantaged children. The No Child Left Behind Act (NCLB) was signed by President George W. Bush on January 8, 2002 and reauthorized the ESEA. The landscape of educational assessment in the United States changed in dramatic ways following the passage and subsequent implementation of NCLB.

NCLB increases federal requirements for states, public schools, and districts. The most significant change requires all public schools to bring the academic proficiency of every public school student up to the state standards for reading and mathematics by the 2013-2014 school year (Cohen, 2002). At the time of this writing the Obama administration has begun accepting applications to grant waivers of some NCLB requirements to states, on condition that those states adopt the Common Core State Standards, develop plans to overhaul the lowest performing schools, and base teacher and principal evaluations on student progress over time. Additionally, NCLB currently requires schools to close academic gaps between economically advantaged students and students who are from different economic, racial, and ethnic backgrounds as well as students with disabilities (U.S. Department of Education, 2002).

NCLB requires that state officials administer tests to all public school students to measure progress in academic achievement. It is the responsibility of individual states' officials to determine their own proficiency standards, known as Adequate Yearly Progress (AYP). AYP is the minimum level of academic achievement that students in states, school districts, and schools must meet or exceed each year, according to Federal No Child Left behind (NCLB) legislation. Over time the percentage of students in a district that must meet the pre-established state proficiency standards progressively increases until the target of 100%

is reached in 2013-2014. If students in a school district do not meet these proficiency levels, the law mandates that sanctions and corrective actions be applied (Yell & Drasgow, 2009). NCLB regulations include guidelines to assist states in defining AYP and goals for schools that are designated as low performing as indicated by AYP measures.

Since the implementation of NCLB, states have been primarily using Status Models to report AYP results. Status Models rely on a single year's assessment as an indicator of how many students are achieving proficiency on academic content standards. Essentially, Status Models are a count of the number of students meeting a specified target. The basic question under this model is: On average how are students performing this year? Comparisons over time are made by comparing the results for the students at each grade level one year with the results for students who were at the same grade level in previous years.

A serious challenge for many schools and districts across the nation is that significant numbers of students enter the next grade level with performance levels well below proficiency. One of today's salient topics in educational curriculum and administration is how to measure student improvement. There are many important implications for this measurement system. A reliable and accurate measurement system, for example, would enable evaluators to observe what teachers are making a difference as well as what type of curriculum and learning environment best promote student learning. In an era of educational accountability, the topic of how to accurately measure student progress is an important hinge upon which the door of improving education turns.

Over the past decade, educational researchers and policy makers have endeavored to identify valuable growth models, methods that measure the growth of students' skills, knowledge, and abilities acquired in public school systems over time (United States Department of Education, 2005). In contrast with status models ─ the bases for adequate yearly progress (AYP) under No Child Left Behind (NCLB, 2001) ─ growth models describe progress in student performance rather than a fixed measure of current-state achievement. Because many current accountability systems have been limited to the use of status designs, growth models appeal to educators and policy makers alike.

The purpose of this dissertation is to review the continuing research and related literature on Growth Models, and evaluate the recommendations on the use of Growth Models, either as a stand-alone accountability system or as a supplementary method integrated into the accountability measures that are already in place. This sectiion review will include:

- The main tenets of the Growth Models.
- A description and critique of four types of Growth Models (Value Table Models, Growth to Proficiency Models, Value-Added Models and Student Growth Percentile Models), and
- Recommendations for the application of Growth Models based on these analyses.

CHATPER TWO

LITERATURE REVIEW

Policymakers invested in education may decide to implement Growth Models for a variety of purposes, but the overarching goal is to improve their ability to assess school efficacy in promoting growth (Betebenner & Shang, 2007). Decisions derived from the assessments of school efficacy are based on the assumption that higher achievement will lead to improved progress toward state NCLB achievement and proficiency standards. However, decisions about implementing a Growth Model or integrating a Growth Model into existing accountability standards should be based on careful reasoning. Proponents of Growth Models argue that these models offer a more accurate depiction of school performance because any given school's ability to foster academic progress is more accurately revealed. Growth Models offer two additional important advantages over Status Models in measuring school accountability. First, Growth Models control for mobility of students between schools from year to year. Second, Growth Models account for students' prior achievement and the effects of their background characteristics (sex, ethnicity, SES, etc.) (Goldschmidt, 2005). However, compared to Status Models, Growth Models tend to be more expensive to develop and maintain and are more difficult for stakeholders to understand.

STATUS MODELS

Description

NCLB requires each state to evaluate school performance and student achievement by determining the degree of achievement attained on state academic standards. The majority of states currently use Status Models to determine AYP. Status Models involve evaluating an entire group at a specific point in time to determine progress toward state determined achievement and proficiency goals. The achievement levels of each grade level of students one year are compared to target achievement levels and the achievement levels of students who were in the same grade level the previous year (Goldschmidt & Choi, 2007).

Advantages

Status Models take a cross section of a subgroup's or a school's level of student proficiency at one point in time and compare that proficiency level with an established target. In AYP, that target is the annual measurable objective (AMO), which is the level of proficiency the state established as an annual goal for schools and students. For example, the AMO might require that 45% or more of the students in each of the state's elementary schools score at the proficient or higher level of achievement in order for that school to make AYP in a particular year. Status Models emphasize the importance of all pupil groups and schools meeting a certain minimum level of achievement.

Disadvantages

One limitation of the Status Models is that they fail to recognize real improvements in student achievement unless they result in higher percentages of students meeting or exceeding proficiency standards in a given year. Schools whose students are demonstrating learning gains, but whose gains overall fall short of the AMO standard will not be judged to be making adequate yearly progress. Conversely, schools whose students meet the AMO standard will be judged to be making AYP even if little is done to advance student achievement beyond that minimum proficiency target.

Further, Status Models compare student performance in groups to a pre-determined target, which does not account for the influence of academic differences between individual students or for student background characteristics as influential moderating factors (Auty, 2008; Ballou et al, 2004; Goldschmidt, 2005). Status Models also do not manage imprecise data, missing data, and small sample sizes for subgroups at the school level well. Alternative models that take such factors into account could allow for a better overall picture of academic progress across grades for all students.

Because Status Models emphasize whether or not schools make AYP, instructional resources may be focused on students who are closest to the proficiency threshold. Students less likely to attain proficiency from a given amount of instructional effort—e.g., those farther below the proficiency threshold—may receive less attention (U.S Department of Education, 2006). Status Models also tend to focus the school's attention and resources on those students who are not achieving proficiency rather than those who are above proficiency.

The nature of Status Models makes it impossible to determine whether or not the score represents typical academic progress using these models. Determining whether or not the score represents typical academic progress requires a model that measures that progress by tracking the achievement scores of the same students from one year to the next.

These limiting characteristics of Status Models have sparked mounting interest in alternative methods of assessment, and states have urged the U.S Department of Education to consider alternative ways to measure and report student progress (Goldschmidt & Choi, 2007). Growth Models have been proposed as an alternative to Status Models to address these limitations.

GROWTH MODELS

Currently states or school districts define and measure academic growth in several ways (Auty,Bielawskie, Deeter, Hirata, Hovanets-Lassila, Rheim, et al., 2008; O'Malley, 2008; Goldstein, 2006). Growth Models generally refer to models of education accountability that measure progress by tracking the achievement scores of each student from one year to the next with the intent of determining whether or not, on average, the student made progress (Goldschmidt et al. 2005). This means that student records from at least two different test administrations have to be combined or matched.

Gong, Perie, and Dunn (2006) defined growth as the "change in the performance (learning) between two (at least) specified points in time" (p. 4). Growth Model assessment requires the ability to track individual student records across schools, and from multiple administrations of tests. Growth Models also include the analysis of the achievement of same-student cohort groups over two or more years; the requirement that test scores allow meaningful comparisons across at least adjacent grades; the implementation of statistical procedures for calculating growth metrics at the individual student and/or school level; and, typically, an explicit link with traditional status or improvement indices (e.g., Are students on track to achieve proficiency within a specified number of years?). Additionally, Growth Model analytic methods require that student data can be tracked longitudinally. This requirement can be met with the use of student ID systems where students are assigned a unique identification number that is recorded with each test a student takes for as long as that student is in the system statewide (Auty et al, 2008). Data must be accessible through a central database that assigns and maintains the ID numbers and also accumulates the required data associated with the ID numbers for analysis and accountability reports.

Key differences between Growth Models and Status Models include whether performance levels or scaled scores are used to calculate the growth metric; the extent to which the models require test scores derived from vertically-scaled tests; the statistical sophistication (lack of transparency) of the methodology; and whether the progress of students is judged against growth targets set by policy and/or in relation to observed growth. While student background characteristics and other variables related to the school context may be, and often are, factors in research studies about achievement growth, such variables cannot be considered in reporting results under the NCLB accountability system where all students are held to the same proficiency targets (Government Accountability Office, 2008). Some Growth Models require student scores to be reported on a common scale in which differences in scores across grades are consistent and meaningful (CCSSO, 2009). However it is possible to create a common scale for existing tests that were designed separately across grades.

Growth Models can help schools and districts see how much academic progress their students are making by measuring that achievement across two points in time. While Growth Models share many common features, they differ greatly with respect to other features. While there are many ways to measure growth and design accountability systems, there are a limited number of general methods that underlie those possibilities (CCSSO, 2010). Below we address four common Growth Models, including (1) Value Table Models, (2) Growth to Proficiency Models, (3) Value-Added Models and (4) Student Growth Percentile Models. We identify important characteristics of each of the models, vertical scaling issues, reporting formats for growth results, and grade levels for which growth is calculated. We describe how each type of Growth Model is being applied in a sample state, and summarize some advantages and disadvantages of each type of Growth Model.

Value Table Models

Description. Value Table Models, for capturing student progress, are based on the principle that accountability can best motivate behavior on the part of school personnel if the expectations are very transparent to educators. Value Table Models are based on standards determined by states for calculating student growth. Value Table models evaluate student transitions across performance levels with the focus on change over the course of a specific number of years, typically two years (CCSSO, 2009). Students are expected to progress across the performance levels so that proficiency is achieved within a predetermined time frame.

Hill, Gong, Marion, DePascale, Dunn, and Simpson (2006) described value Tables as a system that examines students' achievement level from one year to the next by assigning a value to the change in the achievement level. To develop a Value Table Model, a state would need to determine several categories of achievement level for each grade and each discipline. For example, a state might decide to use five levels of achievement with the highest level equivalent to the proficient level on the state's high stakes testing. Each of the five levels would be defined by test scores. Next the state would have to decide on point values to be assigned to a student who moves up from one level to another from one year to the next. For example, if level five is proficient, a student would be assigned more points for moving from level two to level five, and fewer points for moving from level two to level three.

The specific values awarded to students are typically set by an advisory panel using a process of ranking transitions, discussing ranks and averaging ranks over multiple rounds, much like a standard-setting activity. Students are expected to make a set number of transitions each year so that they reach proficiency by a defined number of years, most often three or four.

These models do not require vertical scaling. Vertical scaling allows states to compare student scores from different grades. Nevada and some other states do not use tests that have vertical scaling. In these states, each grade has had its own score scale, making it difficult to compare student performance across grades.

The Value Table Model represents how states or districts want to see individual students within schools progress across performance levels from year to year, based the values placed on different achievement outcomes. The points assigned within a value Table are then averaged across the students in a school. A school's growth scores, based on the averages of the individual student scores, will be higher when more of the students within the school demonstrate the achievement outcomes that are most highly valued by the state or district (Hill, Marion, DePascale, Dunn, Simpson, 2006). Methods for reporting results from Value Table Models differ from state to state, but growth reports usually include progress across value Table levels and the percentage of students meeting the growth targets. States typically report the growth results in summary Table format by subject for all schools and districts.

Value Table Models only compare progress across two years of testing. The current year's results are compared to results for the previous year to determine how many students have gained how many levels of achievement, and also to determine how many students have reached proficiency levels. Growth in Value Table Model can be reported for 4th grade through 10th grade. States implementing Value Table Models include Delaware, Iowa, Michigan, and Minnesota.

Value Table Model Case Study: **Delaware**. Delaware has developed an approach for measuring student growth that relies on awarding points to the school based on the change in students' performance across vertically-articulated achievement standards. A value Table is constructed whereby a school (or district) is awarded varying amounts of points depending on how much progress across achievement standards, each student makes from one year to the next. In Delaware, the stakeholder group worked in teams to review every position in the value Table and rank order the positions from highest valued growth to lowest valued growth. Each cell in the value Table was assigned a number, the stakeholder group ranked the positions, and then the average was determined for each position. Next, the average in each position was multiplied by a constant to create whole numbers and smoothed to more validly reflect the group's value statements. This value Table was then used to refine and inform the discussion about what value should be in each of the positions in the value Table. The final value Table on student growth for the State of Delaware can be seen in Table 39.

Table 39
Delaware's Final Value Table

	Year 2 Levels				
Year 1 Levels	Level 1A	Level 1B	Level 2A	Level 2B	Proficient
Level 1A	0	150	225	250	300
Level 1B	0	0	175	225	300
Level 2A	0	0	0	200	300
Level 2B	0	0	0	0	300
Proficient	0	0	0	0	300

The resulting value Table reflects the values, or weights, the stakeholder group placed on student growth among the achievement levels. In general, Delaware valued movement into Proficiency the most, as reflected by the greatest number of points. The Stakeholder group felt it was unlikely that many students far below Proficient (in Levels 1A and 1B) could validly jump to Level 3, but could make growth to the next higher levels, and so the next largest point values were assigned for students in the lowest achievement levels moving up a plausible amount. Every cell in the value Table was completed through this type of deliberation and discussion. The growth Table reflects values that are consistent with proficiency levels required by NCLB. This Table does not value growth above proficient. Rather, growth to proficiency is valued less than being proficient. There is no compensation for growth above proficiency and if all of the students in a subgroup were proficient in year 1 and stayed proficient, the average score for the subgroup would be 300. Likewise, if all of the students in the subgroup moved from anywhere below proficiency in year 1 to proficiency in year 2, the average score for the subgroup would be 300 points. The maximum average score, therefore, is 300, which is equivalent to 100% proficient, and all students are meeting the standards.

Advantages. Value Table Models are relatively easy for educators to understand. They generate growth scores through simple calculations, and use performance levels that have meaning for educators that are clearly related to state goals for growth over time. Value Table Models also do not require the technical complexity of vertically scaled scores.

Disadvantages. Disadvantages to Value Table Models include the difficult task of creating performance level subcategories and values that appropriately reflect stakeholder values for growth outcomes. Value Table Models also rely on policy decisions about expected growth that may not be realistic because stakeholders create the value Tables and determine the expected growth. This model also does not condition or alter expectations based on student demographics or student characteristics.

Growth to Proficiency Model

Description. Growth to Proficiency models, also known as Growth to Standards or On Track are designed to show whether students are on track to meet standards for proficient and above (Center for Public Education, 2010). In these approaches, students below proficiency in any year may be classified as proficient if their performance indicates that they are on a growth trajectory that is likely to result in proficiency within the specified time span.

Growth to Proficiency Models take initial student performance and provide a yearly growth target for students by which students could reach proficiency in a set number years (e.g., three or four years). Each year, states using Growth to Proficiency Models compare actual student performances to target performances in order to determine if students have academically progressed or regressed during the school year. These models require finding the difference between each student's current achievement scores and the minimum proficiency scores for all students who have not yet reached proficiency. These differences are then divided by the number of years within which the state has targeting them to reach proficiency. The result is the annual gain each student must make to reach proficiency in the targeted year.

Growth to Proficiency Models require an assessment with a vertical scale. The vertical scaling allows for direct comparison of achievement scores across grade levels. Without vertical scaling it would be impossible to determine annual targeted gains.

Annual report for Growth to Proficiency Models include whether students are on track to meet proficiency based on targeted gains. Typically states provide a report that includes summary Tables with the percentage of students meeting growth target for each school. For Growth to Proficiency Models, growth targets are typically established for three years, or no later than the end of a particular grade, such as eighth or tenth grade.

Growth to Proficiency Models are unable to calculate growth targets for a student without the prior year's score. For this reason, Growth in Growth to Proficiency Models are calculated for all non-proficient students in all subgroups for fourth grade through tenth grades, but not third grade. States implementing Growth to Proficiency Models include Alaska, Arizona, Arkansas, Missouri, and North Carolina.

Growth to Proficiency Model Case Study**:** **Arizona**. Arizona's model looks at progress individual students make toward proficiency from one year to the next. The goal is proficiency within three years for grades three and four, or by the eighth grade for grades five through seven. Annual growth targets are set based on that goal. Students are making sufficient progress when they meet the annual growth target. Scores for individual students are aggregated by the relevant subgroups. In Arizona, a separate annual growth target is set for each student in reading and in mathematics. The growth target is set by subtracting the student's previous year scale score from the scale score for proficiency in the target grade and dividing by the number of intervening years. Because Arizona's vertically scaled series of tests ends at eighth grade, the scale score for eighth grade will be used for grades five through seven. The targets are rounded to the nearest whole number. For example, suppose that a student scores 403 on the third grade math test in 2005. The passing score on the sixth grade math test is 496. The student's math score must improve 31 points each year to make up the 93 point difference in three years. Demographic factors are not used to set the target.

The Arizona model calculates the actual annual progress made by each student toward proficiency in state standards and compares these gains to the annual targets. For example, if the student described in the previous a student scores 442 on the 4th grade math test in 2006, the student's actual growth is 39 points in one year. This actual growth exceeds the targeted growth of 31 points, and this student would be making adequate progress according to Arizona's Growth to Proficiency Model.

Advantages. Advantages to Growth to Proficiency Models include giving schools credit for producing gains in student performance even if their students score below proficient, and encouraging schools to focus on all students below the proficiency level, not just those students who immediately under the proficiency level (Center for Public Education, 2010). Growth to Proficiency Models permit schools to be rewarded for students who are making progress toward reaching proficiency as well as for students who are proficient.

Disadvantages. Growth to Proficiency Models may have limited application because many statewide assessment systems do not have vertically scaled tests. In addition, there is no empirical basis for determining whether growth targets set by policy are reasonable. As with the Value Table Models, Growth to Proficiency Models do not take student characteristics and demographics into consideration.

Value-Added Model

Description. Value-Added Models are designed to predict student performance up to three years in the future. The primary assumption of Value-Added Models is that past test achievement is a good predictor of current and future achievement. Value-Added Models use sophisticated regression formulas to estimate how much value a school (or in some cases a teacher or a specific program) has added to a student's learning by projecting how much each of these variables contributes to each student's progress. Projections are based on past and current student performance. These models use student scores in all content areas and information from previous years to predict future test performance (Sanders, 2003). Value-Added models predict student performance based on a student's past performance and the performance of prior cohorts in target grades for which there is a predetermined proficiency standard (CCSSO, 2009). With these types of Growth Models, credit for student growth is given to schools for students who have not met proficiency but are projected to reach proficiency within a specified number of years.

Value-Added Models require assessments with a vertical scale (Ballou et al., 2004; Doran & Cohen, 2005). Scores must be on a vertical scale across grades in order to project growth across time.

The results of Value-Added Models are typically reported with summary pages for all students by grade, and these models can be used to report results at all levels, from individual students through teachers, schools, districts, and even states. Individual student growth data from assessment results can be accessed by educators through school districts. At least two years of growth data must be sued in Value-Added Models.

Growth in Value-Added Models can be calculated for 4th grade through 10^{th} grade. States implementing Value-Added Models include Tennessee, Ohio, Texas, and Pennsylvania.

Value-Added Model Case Study: **Ohio.** Ohio's model uses up to five years of available test scores for individual students, merged longitudinally, to provide the best estimates of future student achievement trajectories. This model uses all available test scores for each child to create an individual growth trajectory in the appropriate subject. Student-level projections can be created for any subject and any test metric. Test scores for all content areas, including science and social studies, are analyzed to improve the precision of the estimates. In this way, data on other content areas are included to try to improve the achievement projections in the core areas of reading and mathematics. By using all of an individual student's previous achievement data to make these projections, the need to adjust for a student's socioeconomic condition, ethnic background, prior knowledge, etc. is not necessary. Any predictive power based on these student characteristics would be incorporated into the each individual's previous data. This approach ensures that students with the same prior academic achievement will have the same analysis regardless of the neighborhood in which the attending school is located or the student's demographic characteristics.

Advantages. Value-Added Models allow for more precise growth information to be generated, if the assumptions of the statistical methodology are met (Center for Public Education, 2010). They also allow for the inclusion of student demographics and characteristics in achievement projections.

Disadvantages. There are many challenges in implementing Value-Added Model. The models require test data for each student over at least two years. Complex statistical procedures may be proprietary and statistical expertise is required to implement the models. Expensive computer software and hardware may be required. Teachers and administrators may have difficulty understanding understand the statistics and interpreting the results (Hibpshman, 2004).

Student Growth Percentile (SGP)

Description. There are several ways to measure student growth from year-to-year. Student Growth Percentile Models have been getting some attention lately because they control for each student's starting proficiency. This approach focuses on estimating the observed growth of a student in relation to students with the same academic achievement history in order to establish a normative baseline for growth. Student Growth Percentile Models measure how much a student has improved, or grown, academically from one year to the next compared to the student's academic peers. An academic peer is a student who has had the same or similar test scores over a period of years. This approach is in contrast to most other Growth Models where growth targets based on policy, and not on empirical information about typical growth.

Pediatric growth charts are an easily understood analogy used to explain Student Growth Percentile Models (Betebenner, 2008a, 2008c). When describing the physical growth of a child, growth is measured at two different times using a quantity, such as inches, that is easily understood by all interested parties. Growth becomes meaningful, however, only when normatively referenced to the growth obtained by similar children during the same timeframe of measurement. Defining whether or not the amount of growth obtained is typical is determined by comparing it relative to the growth exhibited by other children. Thus, a criterion is obtained by which to reference the amount of growth expected at any given age or stage. Similarly, Student Growth Percentile Models offer information on student academic growth by normatively comparing the amount of academic growth a student makes. They provide a criterion of expected growth by comparison to the amount of academic growth obtained by similar students during the same timeframe.

For example, if a student's SGP score is determined to be 70%, that means only 30% of the students who had the same prior achievement had the same or higher achievement. That student's growth was substantially above the median in relation to the student's academic peers. If, on the other hand, a student's SGP score is determined to be 20%, that means 80% of the students who had the same prior achievement had higher achievement—that student's growth was substantially below the median in relation to the student's academic peers. The growth percentile threshold serves to establish a baseline for schools by which to assess the number of students achieving or exceeding expected growth. Growth can then be evaluated at the classroom level, grade level, school level and district level.

Student Growth Percentile Models does not require vertical scaling. At each grade level, scores are compared to other students at the same grade level to determine each student's relative standing compared to other students at the same grade level.

This model reports a school's median percentile annual growth across all students, based on an analysis of up to five-years of individual student data. Student Growth Percentile data reports generally present the results in interactive graphical pages and Tables. States also report the summary results by subject for all schools and districts. Individual student growth data is also accessible through this model.

Generally two to five years of data are used for Student Growth Percentile Models. Growth in Student Growth Percentile Models is calculated for fourth grade through tenth grade. All the analyses to calculate student growth percentiles can be done in R (R Development Core Team, 2010) using the R package called *SGP: An R Package for the Calculation and Visualization of Student Growth Percentiles*, developed by Betebenner and Van Iwaarden (2011). States implementing Student Growth Percentile Model include Colorado, Nevada, Indiana, Massachusetts, Virginia, and Rhode Island.

Student Growth Percentile Model Case Study: **Colorado.** Colorado has developed a Student Growth Percentile Model to answer three essential questions about student, school and district performance (Colorado Growth Model, 2009; Betebenner, 2008):

- What is the growth rate of a student, a school and a district?
- What should be the growth rate for a student to reach a desired level of achievement within a period of time?
- What are the highest sustained growth rates that exist today and under what conditions could they improve?

To answer these questions, the Colorado Growth Model uses a common measure to describe how much growth each student makes and how much growth is needed to reach state standards. In doing so, the Colorado growth Model provides a complete history of all students' individual-level test scores from the Colorado Student Assessment Program. The model depicts academic growth in an interactive display that relates normative information about student progress toward the criteria of reaching different state proficiency levels. The Colorado Growth Model uses quantile regression methodology combined with all available prior test score data for each student to determine students' growth targets. Quantile regression is more suiTable for developing results based on percentiles and medians than least squares regression. The Colorado Growth Model sets growth targets for all students, whether they are proficient or not. Further, the Colorado Growth Model holds all disaggregated groups accounTable for reaching the same growth targets. These growth targets are based on prior growth for each student relative to academic peers.

In Colorado's model, a student growth percentile corresponding to a student's current score requires all available consecutive prior scores. For the purposes of state-wide calculations of student growth percentiles, growth percentile analyses are performed for each subject and each grade separately. Each grade and subject cohort's data are analyzed so that each student receives a growth percentile in each subject in which the student was tested. As with any analysis of large-scale observational data, missing or incomplete data are an issue. Because the purpose is to describe growth in the most recent year, at a minimum, it is necessary to have the student's prior year's scores.

According to Colorado's model, a typical school or district in the state would have a median student growth percentile of 50. No matter how student growth percentiles are aggregated, whether at the subgroup, school, or district level, that median, and its interpretation, remain the same.

Advantages. Student Growth Percentile Models ensure better informed decisions about adequate growth in that those decisions are based on prior growth. This is in contrast to most other Growth Models where growth targets are set by policy. Student Growth Percentile Models take into account each student's academic starting points and the rate of growth the student has maintained over several years. Student Growth Percentile Models allow districts and states to determine, and focus on, what is necessary to help students attain adequate academic achievement. Student Growth Percentile Models describe a student's growth by locating the student's current score within the distribution of students who had identical prior achievement. These models offer empirically based growth targets for all students, including those who have already exceeding state proficiency levels of achievement.

Disadvantages. Establishing criteria to determine adequate growth will be challenging for districts and states (Betebenner & Shang, 2008). Some of the technical challenges include being able to interpret the scale used for measuring academic growth relative to that obtained by other students, and possessing the technical expertise to perform the statistical procedures that will establish the growth percentile baseline.

Comparison of Growth Models

The characteristics of the four types of Growth Models we described are summarized in Table 40. These characteristics may inform the choices of state officials deciding which models to adopt. For example, states that do not have vertically scaled assessments would have some difficulty adopting one of the two models that requires vertically scaled assessments. States that want to ensure that their growth models can be understood by staff, parents and other stakeholders may choose not to adopt a model that requires complex statistical analyses.

Table 40

Features of Growth Models

Features	Value Table	Growth to Proficiency	Value-Added	Student Growth Percentile
Vertical Scaling	No	Yes	Yes	No
Years of Data	2	1	2+	2+
Complex Statistics	No	No	Yes	Yes
Student Characteristics	No	No	Yes	No

Conclusions

Growth Models are the new frontier of academic evaluation, proficiency determination, and teacher/program/school assessment. Researchers are just beginning to understand the accuracy of growth model measures and how to best interpret the results, although there is general agreement that Growth Models are an improvement over current Status Models.

Growth Models conceptually align well with one of the fundamental goals of education –student learning. Our educational systems expect continuous learning from all students and measures of student growth should help evaluate those expectations. Growth Models can be valuable in analyzing students' scores over time to improve understanding of student learning. Growth Models focus on the educational development of individual students.

Growth Models can also be important tools for evaluating schools, teachers, and education programs, but like any measurement model, Growth Models have their limitations. While some tests are clearly better than others for particular applications, there is no perfect measure of achievement (Ballou 2002, McCaffrey, et al. 2003). Many of the tests administered by states are not designed for measuring growth. For example, if the scores are not vertically scaled, they do not directly reflect achievement gains. Measurement error is one of the critical components states should also consider (Ballou 2002, Doran 2004). When scores from the tests are subtracted from each other, as in Value Table Models, the measurement error of the difference is greater than the measurement error of the individual scores (Ballou 2002), and yet educators rely on these differences to measure growth.

Even the best data collection system cannot assure that all data will be produced and reported for every student. In any given year, some students are absent during testing. Some students transfer into school during the school year from other states. Others transfer out. These factors, and others, lead to missing data for some students. Missing data can have a large impact on growth results, depending on the characteristics of students who typically fall into this group. Students who are highly mobile, for example, tend to be lower achievers. A high incidence of mobility in a school would produce gaps in the data and could distort the results reported for the school and its teachers (McCaffrey, et al. 2003).

In addition, growth models – when compared with status measures – can be very difficult for non-statisticians to understand. States implementing Growth Models have invested time and energy in developing and disseminating information about those models.

Growth data can be helpful in many ways, and it is tempting to create one Growth Model and use it for multiple purposes. Policymakers and educators should resist this temptation. Although developing a single model could save a lot of time and money over developing multiple models, researchers strongly discourage using just one model, because pulling distinct pieces of information from one single model would likely lead to false conclusions (Ballou 2002). For example, a Growth Model developed for high-stakes school accountability, such as NCLB, should not be used for program and teacher evaluation when these evaluation controls for other variables such as socioeconomic status (Gong, Perie, and Dunn 2006).

Growth Models need to be adequately reliable, valid, and useful. However, developing an assessment system that is adequately reliable, valid, places demands on the developers. These include, but are not limited to, being clear about how the information will be used, and using tests that produce results with adequate reliability and validity for the purpose. States must have the technological capacity to track students' progress and other characteristics through the school system and match them with each school the students attend. The calculations for Growth Models take many forms and cannot be easily illustrated due to their complex statistical methodology. Because of this complexity, hiring or assigning trained experts to oversee the design and implementation of the model is critical. On the other hand, transparency and good communication also play critical roles in successful implementation of

Growth Models. Not surprisingly, funding is also a key factor. There will likely be some costs involved in software and hardware, as well as personnel. Much of the cost involved in implementing a Growth Model involves infrastructure. Constructing or revising current infrastructure to supply the necessary data for the specific growth model to be implemented is a significant contributor to the overall cost of implementation.

A number of states have already implemented Growth Models into their assessment processes, each using a variation of the Growth Models detailed in the body of this paper. Each type of Growth Model has strengths and weaknesses that states must consider when deciding which type of models will best serve the needs of the students within their districts, while also addressing the concerns of stakeholders, and meeting the goals established by NCLB (U.S. Department of Education, 2005).

As additional states consider adopting similar models, they will need to make important decisions in determining which type of Growth Models best suits their needs. States must first identify what they hope to achieve by implementing a Growth Model. Each of the Growth Models implemented by individual states has its own unique set of defined goals that fulfill the purpose of incorporating Growth Models into the accountability measures required by NCLB. However, in all instances, the overarching goal of implementing a Growth Model is to address the deficiencies of the current Status Model measurements of accountability and proficiency (Ho, 2008; O'Malley, Auty, Bielawski, et al, 2009). For example, while Value-Added Models are commonly used for school effectiveness and teacher evaluation, Student Growth Percentile Models are designed to describe students' academic gains relative to their academic peers. Models that are designed to consider more factors have the benefit of greater precision, yet require more complex statistical analyses and may be difficult to explain to stakeholders. Further, some states may decide that the growth model will be the only method of assessment while others will choose to integrate Growth Models with Status Models.

Growth Models offer practical benefits for states seeking to address the limitations of the Status Models that are currently used to assess student academic achievement and determine school accountability. Growth Models offer flexibility and improved measurement reliability when they are comparing data for the same students over time. Progress can then be defined as the degree to which students' estimated improvements compare to a statewide or local target. Growth Models offer a mechanism by which more precise measures of student growth toward proficiency and student academic achievement may be obtained. However, Growth Models require effort on the part of states to put model requirements in place before the model is implemented.

CHAPTER THREE

METHODS

The purpose of this dissertation is to compare growth models with respect to their ability to a) accurately project students' scores at a future point in time on the basis of past performance and (b) investigate the Student Growth Percentile (SGP) as an alternative growth model approach to the measurement of student and school achievement and (c) to build understanding of how Student Growth Percentile plays out with state and district data and (d) how correctly classify students' proficiency categories at a future point on the basis of previous performance.

Research Questions

Using results from the Nevada Criterion Referenced Test (CRT) statewide assessment program from 2005-2010 for Elementary Schools, and Nevada AYP targets, The following research questions serve as the primary focus of the study for each disaggregated group (total of 9 different groups) using a Student Growth Percentile (SGP) Model:

Research Question 1: How well do growth trends for the 2005-2010 data predicts which students will meet AYP targets for 2009-2010? Research Question 2: Does the SGP target variable accurately predict CRT proficiency? Research Question 3: How do the growth trends for Washoe County School District (WCSD) compare to the growth trends for Nevada based on CRT data from 2005-2009? Research Question 4:How are the growth trends similar and different across disaggregated groups? Research Question 5: Does SGP correctly classify students' proficiency categories at a future point on the basis of previous performance? Research Question 6: Does the SGP model accurate in classifying students into the different segment of the population based on grade and ethnicity?

Data

Assessment data for the normative analysis will be gathered from the 2005-2006, 2006-2007, 2007-2008, 2008-2009 and 2009-2010 school years. Data from all students in elementary school who took the Nevada statewide criterion-reference test (CRT) in reading or mathematics will be included (N~=300,000). The Student Growth Percentile method uses the student's historical CRT results and currently requires results from at least two grades. Therefore, no results will be available for grade 3 (the first grade of CRT testing) or for science (because science is tested only in grades 5, 8 and high school). Demographic data will be collected through the state's student information system, and assessment data will be collected through the Nevada Department of Education.

As discussed in chapter 2 to test the frequency of SGP target variable accurately predicting CRT proficiency and we looked over effectiveness of SGP to predict CRT performance of students in grades 3-8. Chi-Square of SGP target achievement (dichotomous) and CRT proficiency Frequencies of SGP Target Variable Accurately Predicting CRT Proficiency calculated.

In this section of the report we perform frequencies on the median regression, the covariates in this model are previous test scores, and the dependent variables are the current test scores. To address the first research I will calculate the individual SGPs for the entire population. Two master data files will be created, one for Reading and one for Mathematics. The following variables will be used for the initial SGP calculation; unique student ID, students grade level from 2005-2009 and available scale scores from 2005 to 2009. The unique student IDs will contain no personally identifying information, and will not be related to any district or state identifier. Variables in the Reading data file will be as follows; StudentID, GRADE2006, GRADE2007, GRADE2008, GRADE2009, ReadSS2006, ReadSS2007, ReadSS2008, ReadSS2009. Variables in the Math data file will be as follows:

StudentID, GRADE2006, GRADE2007, GRADE2008, GRADE2009, MathSS2006, MathSS2007, MathSS2008, MathSS2009).

SGP Calculation

Calculation of SGP will be performed using R (R Development Core Team, 2009), a software language and environment for statistical computing, with the SGP package (Betebenner, 2009). (There are also other softwares with quantile regression capability such as SAS and Stata, but I preferred to use R because the original SGP package was written in the R environment by Dr. Betebenner). SGP is an R Package that contains functions to calculate and visualize Student Growth Percentiles (SGP) and percentile growth projections/trajectories. I will construct a master file that contains the necessary R syntax to calculate the SGPs.

Calculation of a student's growth percentile is based upon the estimation of the conditional density associated with a student's score at time *t* using the student's prior scores at times 1, 2,…, *t* - 1 as the conditioning variables. (If the distribution is conditional on some information, then the density is called conditional density). Estimation of the conditional density will be performed using quantile regression (Koenker, 2005). Regression analyses include any techniques for analyzing several variables, and focus on the relationship between a dependent variable and one or more independent variables. Quantile regression is a type of regression analysis used in statistics that as introduced in Koenker and Bassett (1978) may be viewed as a natural extension of classical least squares estimation of conditional mean models to the estimation of an ensemble of models for conditional quantile functions. In this study, I will use quantile regression analysis to estimate the conditional density for current achievement using each student's achievement history. Each student will be compared to other students statewide with similar CRT test score histories for this analysis. This makes for a fair comparison because it allows me to describe the likely range of scores observed among all students with a similar CRT test score history, and therefore to see how quickly the student improved given his or her past test scores.

After calculating student level SGPs, the next phase will be to calculate the growth projections and analyze how well the SGP data predict which student will meet AYP targets for 2009-2010. This analysis will allow me to investigate what level of growth is necessary for the student to reach desirable levels of achievement in the future. These growth projections will allow me to quantify and discuss what it will take for a student to reach desired levels of achievement. Percentile growth projections/trajectories will be calculated using the coefficient matrices derived from the student growth percentile analyses. Calculation of growth projections will be performed using R with the SGP package (Betebenner, 2009). The growth projections will determine what the student growth percentile should be to the AYP target. The analysis will provide three different targets:

-*Target 1*-Did not meet AYP (Emergent/Developing or Approaching Standards),

-*Target 2*- Met with AYP (Meets Standard),

-*Target 3*- Met with AYP (Exceeds Standards).

For example, Johnny's current SGP is 10, Target 2 is 65 and Target 3 is 99. This means that if Johnny wants to meet the AYP target, his growth percentile should be at least 65 for the following year. If he wants to fall in the Exceeds Standard category in AYP, his growth percentile should be 99. I will calculate the individual growth projections and analyze how well these projections predict the 2009-2010 AYP results.

SGP – AYP

To address the second research question (How do the growth trends for Washoe County School District (WCSD) compare to the growth trends for Nevada based on CRT data from 2005-2009?) I will calculate the state level student growth percentiles and compare those with WCSD results. The comparison will include each disaggregated group's (total of 9 different groups) SGP results.

To address the third research question (How are the growth trends similar and different across disaggregated groups?), I will use the student level SGP results and will rely on the original projections calculated for the first question. These will be applied to each disaggregated group. An advantage of quantifying growth at the student level is that it is generally an easy task to combine the individual level growth results to retrieve a school level aggregate. For example, after growth percentiles are calculated for each of the 500 students at a school, the distribution of growth percentiles for those 500 students represents how much the students at that school grew in the previous year. Summarizing this distribution's "average" would supply a single number describing the growth of student's at a given school. Because it is not appropriate to calculate a typical average using percentiles, the median is used as the single number which best describes where the middle of the distribution of student growth percentiles lies.

Individual level SGP results will be aggregated to the subpopulations specified by NCLB, and the growth trend will be compared across these groups. The median student growth percentile will be used for this analysis.

Table 1: *Estimated Rate of Accurately Predicting CRT Proficiency in Math and Reading*

		Math*		Reading	
		N	%	N	%
Overall	Inaccurate Prediction	6031	32.8%	2289	12.4%
	Accurate Prediction	12360	67.2%	16106	87.6%
4th	Inaccurate Prediction	551	11.9%	466	10.0%
	Accurate Prediction	4096	88.1%	4179	90.0%
5th	Inaccurate Prediction	1324	28.6%	319	6.9%
	Accurate Prediction	3308	71.4%	4314	93.1%
6th	Inaccurate Prediction	1277	28.0%	782	17.1%
	Accurate Prediction	3285	72.0%	3779	82.9%
7th	Inaccurate Prediction	2879	63.3%	722	15.8%
	Accurate Prediction	1671	36.7%	3834	84.2%
African American	Inaccurate Prediction	133	29.4%	62	13.7%
	Accurate Prediction	319	70.6%	390	86.3%
Asian	Inaccurate Prediction	269	30.5%	96	10.9%
	Accurate Prediction	612	69.5%	785	89.1%
Hispanic	Inaccurate Prediction	2317	33.4%	1163	16.8%
	Accurate Prediction	4623	66.6%	5779	83.2%
Multi-Race	Inaccurate Prediction	240	31.0%	80	10.3%
	Accurate Prediction	534	69.0%	694	89.7%
Native American	Inaccurate Prediction	103	34.6%	43	14.4%
	Accurate Prediction	195	65.4%	255	85.6%
Pacific Islander	Inaccurate Prediction	51	34.5%	23	15.5%
	Accurate Prediction	97	65.5%	125	84.5%
White	Inaccurate Prediction	2918	32.8%	822	9.2%
	Accurate Prediction	5980	67.2%	8078	90.8%
LEP	Inaccurate Prediction	775	22.9%	583	17.2%
	Accurate Prediction	2616	77.1%	2810	82.8%
IEP	Inaccurate Prediction	434	20.0%	282	13.0%
	Accurate Prediction	1733	80.0%	1893	87.0%
FRL	Inaccurate Prediction	2960	32.6%	1462	16.1%
	Accurate Prediction	6127	67.4%	7629	83.9%

** Transition cut scores of 255 for 6th, 267 for 7th, 267 for 8th*

Table 2: *Estimated Rate of Accurately Predicting CRT Proficiency in Math, Separated by Transition Cut Scores and Cut Scores of 300*

		Math		Math	
		Transition Cut Scores*		Cut Scores of 300**	
		N	%	N	%
Overall	Inaccurate Prediction	6031	32.8%	4242	23.1%
	Accurate Prediction	12360	67.2%	14149	76.9%
4th	Inaccurate Prediction	551	11.9%	551	11.9%
	Accurate Prediction	4096	88.1%	4096	88.1%
5th	Inaccurate Prediction	1324	28.6%	1324	28.6%
	Accurate Prediction	3308	71.4%	3308	71.4%
6th	Inaccurate Prediction	1277	28.0%	410	9.0%
	Accurate Prediction	3285	72.0%	4152	91.0%
7th	Inaccurate Prediction	2879	63.3%	1957	43.0%
	Accurate Prediction	1671	36.7%	2593	57.0%
African American	Inaccurate Prediction	133	29.4%	72	15.9%
	Accurate Prediction	319	70.6%	380	84.1%
Asian	Inaccurate Prediction	269	30.5%	210	23.8%
	Accurate Prediction	612	69.5%	671	76.2%
Hispanic	Inaccurate Prediction	2317	33.4%	1471	21.2%
	Accurate Prediction	4623	66.6%	5469	78.8%
Multi-Race	Inaccurate Prediction	240	31.0%	174	22.5%
	Accurate Prediction	534	69.0%	600	77.5%
Native American	Inaccurate Prediction	103	34.6%	70	23.5%
	Accurate Prediction	195	65.4%	228	76.5%
Pacific Islander	Inaccurate Prediction	51	34.5%	36	24.3%
	Accurate Prediction	97	65.5%	112	75.7%
White	Inaccurate Prediction	2918	32.8%	2209	24.8%
	Accurate Prediction	5980	67.2%	6689	75.2%
LEP	Inaccurate Prediction	775	22.9%	499	14.7%
	Accurate Prediction	2616	77.1%	2892	85.3%
IEP	Inaccurate Prediction	434	20.0%	263	12.1%
	Accurate Prediction	1733	80.0%	1904	87.9%
FRL	Inaccurate Prediction	2960	32.6%	1925	21.2%
	Accurate Prediction	6127	67.4%	7162	78.8%

** Transition cut scores of 255 for 6th, 267 for 7th, 267 for 8th*

*** Cut scores of 300 for 6th, 7th, and 8th*

Chi Square Tests for Independence on SGP Target Achievement and CRT Proficiency

This section of the dissertation conducts Chi-square tests for independence on 2 (SGP target achievement) X 2 (CRT proficiency) designs. The SGP target achievement variable was created by categorizing with the value 1, students for whom actual 2010-2011 SGP values were equal to or greater than the predicted SGP values. The value of 0 was given to students for whom actual 2010-2011 SGP values were less than the predicted SGP values. So, for the SGP target achievement variable each student had a value of 1 or 0. Chi-square tests are performed in math and reading for overall students in grades 4-7, for individual grades, for each ethnicity, and for special subgroups. Vertically scaled scores were created by test centers. The data collection design used to establish the vertical scale and all state students took operational test forms at their grade level. Samples of students at each grade level took from the operational form either above or below their grade level.

Overall Students

Math. A total of 18,391 students comprised the overall student sample for math in grades 4-7. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 4149.15, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* = -.475, $p < .001$ (approaching a large effect) according to Cohen's (1988) criteria.[1]

As displayed in Table 3 and visually in Figure 1, among students who achieved their SGP target, 98.2% were CRT proficient and 1.8% were not CRT proficient. Among students

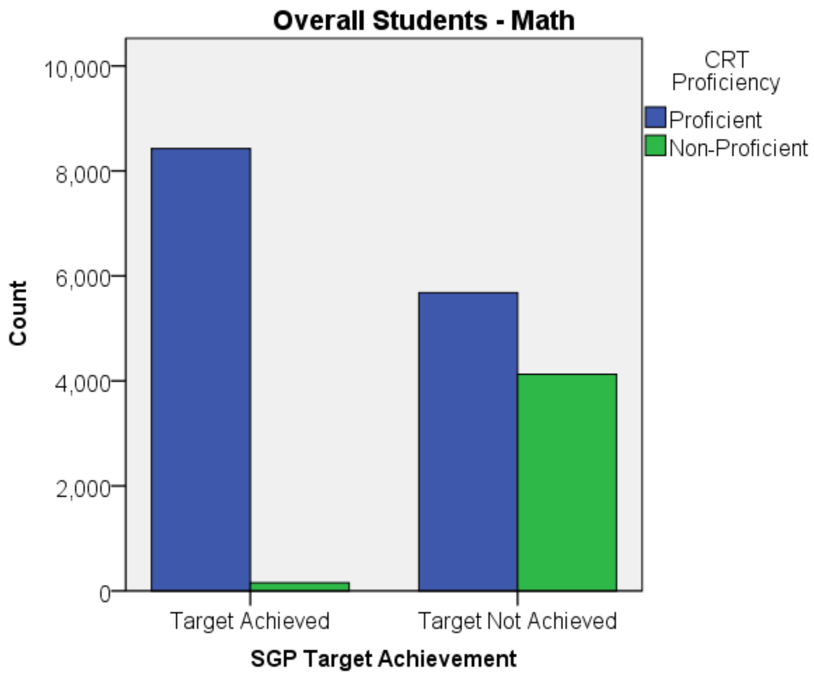
[1] Cohen's (1988) criteria for effect sizes was .10 for a small effect, .30 for a medium effect, and .50 for a large effect.

who did not achieve their SGP target, 57.9% were CRT proficient and 42.1% were not CRT proficient.

Table 3: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Students Overall in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**8426**	157	8583
	%	**98.2%**	1.8%	100.0%
SGP Target *Not* Achieved	***N***	5680	**4128**	9808
	%	57.9%	**42.1%**	100.0%

Figure 1: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Students Overall in Math*



Reading. A total of 18,395 students comprised the overall student sample for reading in grades 4-7. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 9753.13, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between

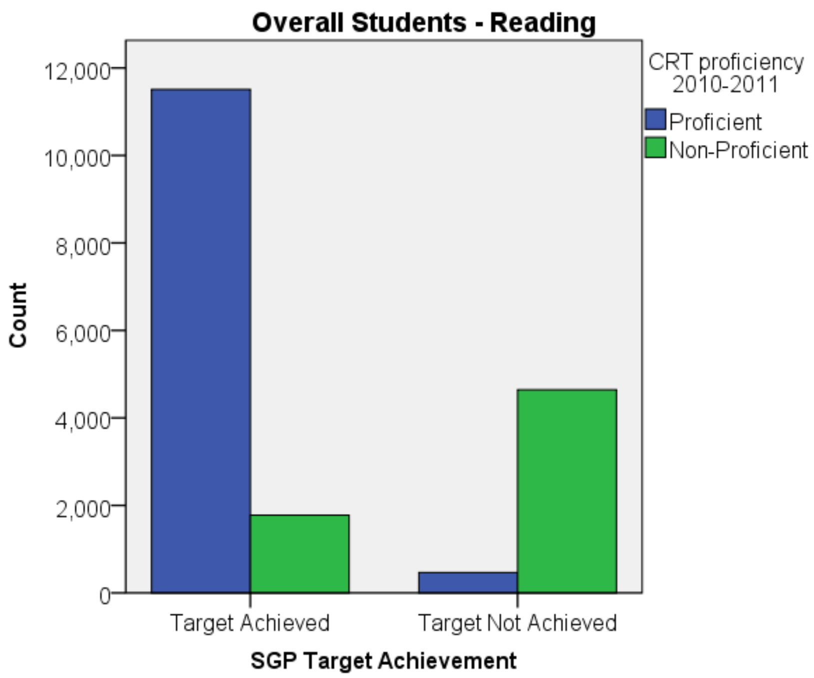
SGP target achievement and CRT proficiency demonstrated a large effect, *phi* = .728, $p <$ .001 according to Cohen's (1988) criteria.

As displayed in Table 4 and visually in Figure 2, among students who achieved their SGP target, 86.6% were CRT proficient and 13.4% were not CRT proficient. Among students who did not achieve their SGP target, 9.1% were CRT proficient and 90.9% were not CRT proficient.

Table 4: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Students Overall in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**11508**	1777	13285
	%	**86.6%**	13.4%	100.0%
SGP Target *Not* Achieved	***N***	466	**4644**	5110
	%	9.1%	**90.9%**	100.0%

Figure 2: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Students Overall in Reading*



CHAPTER FOUR

RESULTS

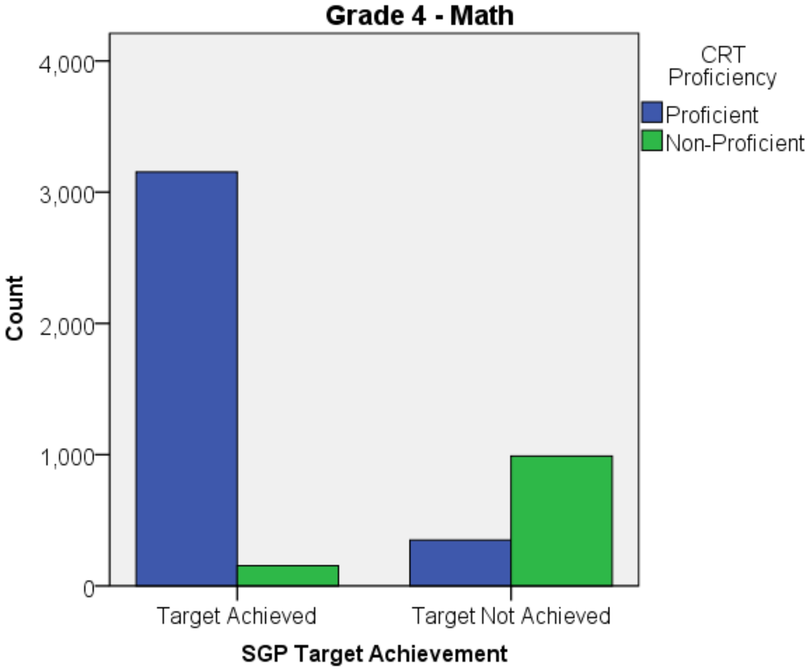
This section analyzes Chi-square tests for independence on four individual grades: 4^{th}, 5^{th}, 6^{th}, and 7^{th}. A total of 4,647 students were in our sample from grade 4 for math. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 2457.73, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.728, p < .001 according to Cohen's (1988) criteria.

Math. As displayed in Table 5 and visually in Figure 3, among students who achieved their SGP target, 95.3% were CRT proficient and 4.7% were not CRT proficient. Among students who did not achieve their SGP target, 26.1% were CRT proficient and 73.9% were not CRT proficient.

Table 5**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 4 Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**3154**	154	3308
	%	**95.3%**	4.7%	100.0%
SGP Target *Not* Achieved	***N***	350	**989**	1339
	%	26.1%	**73.9%**	100.0%

Figure 3: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 4 Students in Math*



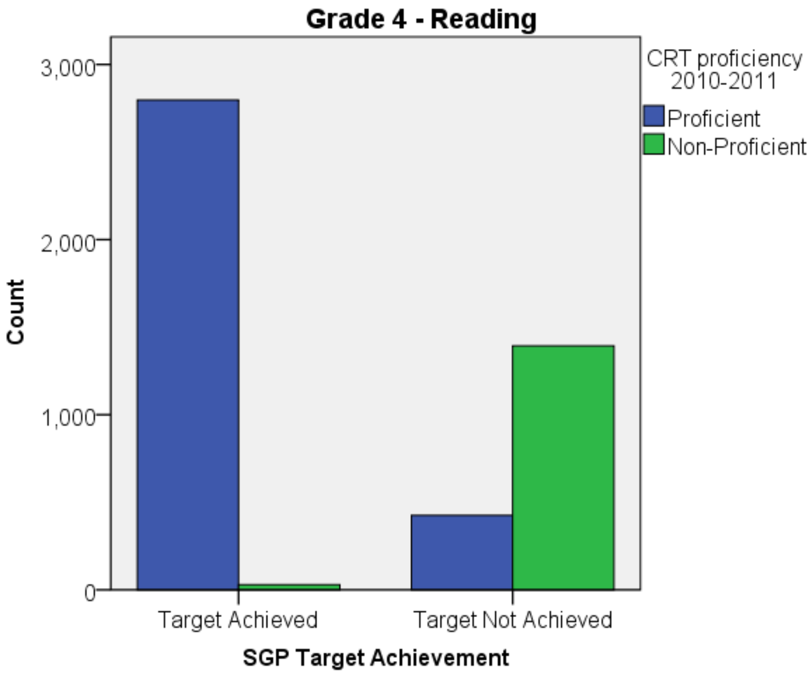
Reading. A total of 4,645 students were in our sample from grade 4 for reading. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 2969.31, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.800, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 6 and visually in Figure 4, among students who achieved their SGP target, 98.9% were CRT proficient and 1.1% were not CRT proficient. Among students who did not achieve their SGP target, 23.4% were CRT proficient and 76.6% were not CRT proficient.

Table 6: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 4 Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**2797**	30	2827
	%	**98.9%**	1.1%	100.0%
SGP Target *Not* Achieved	***N***	425	**1393**	1818
	%	23.4%	**76.6%**	100.0%

Figure 4: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 4 Students in Reading*



Grade 5 Students

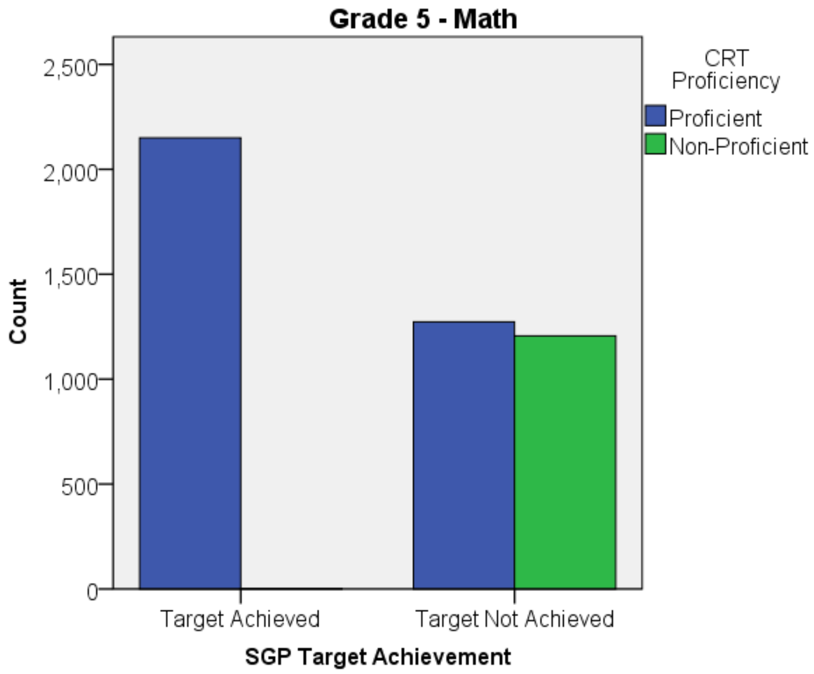
Math. A total of 4,632 students were in our sample from grade 5 for math. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1403.23, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.551, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 7 and visually in Figure 5, among students who achieved their SGP target, 99.9% were CRT proficient and 0.1% were not CRT proficient. Among students who did not achieve their SGP target, 51.4% were CRT proficient and 48.6% were not CRT proficient.

Table 7: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 5 Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**2150**	3	2153
	%	**99.9%**	0.1%	100.0%
SGP Target *Not* Achieved	***N***	1273	**1206**	2479
	%	51.4%	**48.6%**	100.0%

Figure 5: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 5 Students in Math*



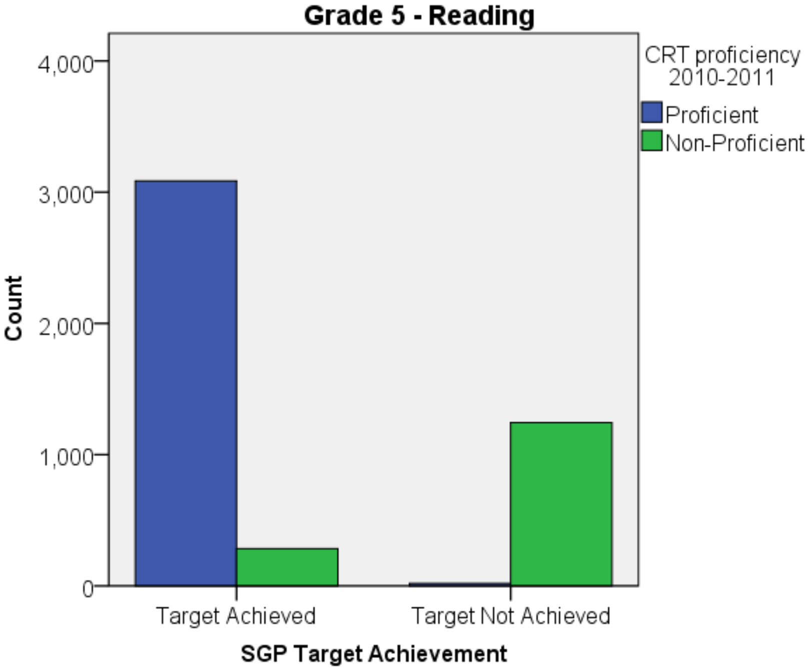
Reading. A total of 4,633 students were in our sample from grade 5 for reading. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 3367.67, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.853, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 8 and visually in Figure 6, among students who achieved their SGP target, 91.6% were CRT proficient and 8.4% were not CRT proficient. Among students who did not achieve their SGP target, 1.5% were CRT proficient and 98.5% were not CRT proficient.

Table 8**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 5 Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**3086**	284	3370
	%	**91.6%**	8.4%	100.0%
SGP Target *Not* Achieved	***N***	19	**1244**	1263
	%	1.5%	**98.5%**	100.0%

Figure 6: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 5 Students in Reading*



Grade 6 Students

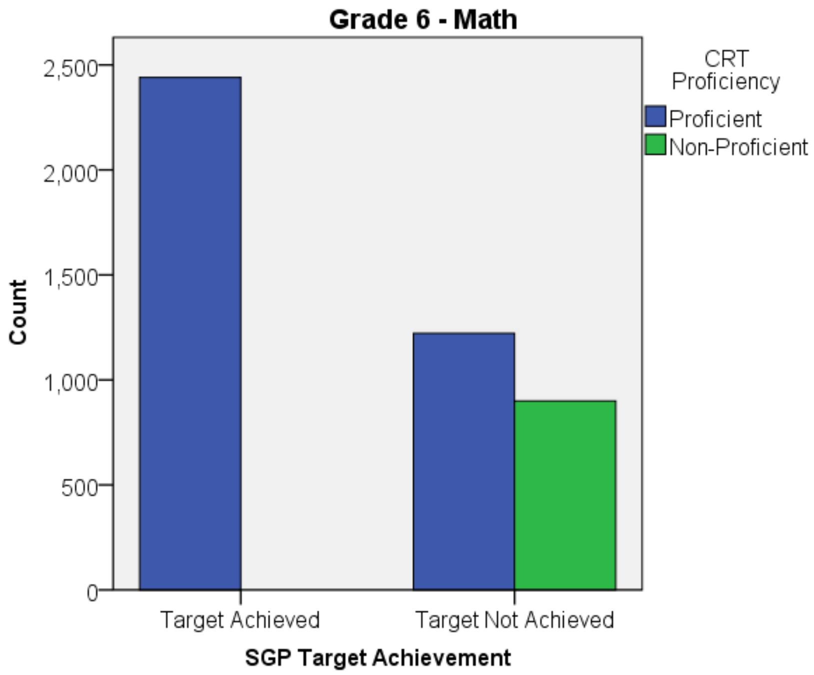
Math. A total of 4,562 students were in our sample from grade 6 for math. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1286.53, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.532, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 9 and visually in Figure 7, among students who achieved their SGP target, 100% were CRT proficient and 0% were not CRT proficient. Among students who did not achieve their SGP target, 57.6% were CRT proficient and 42.4% were not CRT proficient.

Table 9: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 6 Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**2440**	0	2440
	%	**100.0%**	0.0%	100.0%
SGP Target *Not* Achieved	***N***	1222	**900**	2122
	%	57.6%	**42.4%**	100.0%

Figure 7: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 6 Students in Math*



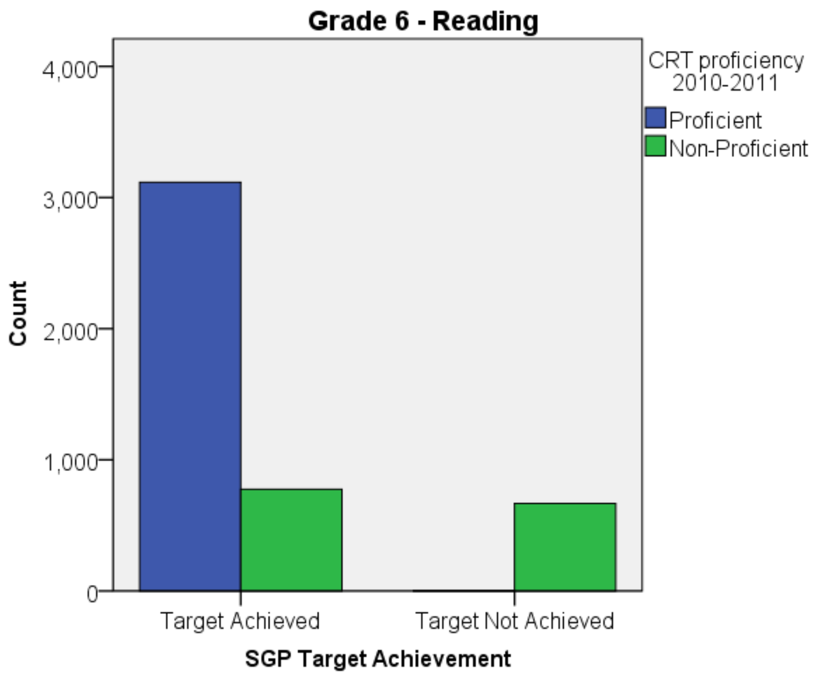
Reading. A total of 4,561 students were in our sample from grade 6 for reading. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1672.87, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.606, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 10 and visually in Figure 8, among students who achieved their SGP target, 80.1% were CRT proficient and 19.9% were not CRT proficient. Among students who did not achieve their SGP target, 0.4% were CRT proficient and 99.6% were not CRT proficient.

Table 10**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 6 Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**3116**	775	3891
	%	**80.1%**	19.9%	100.0%
SGP Target *Not* Achieved	***N***	3	**667**	670
	%	0.4%	**99.6%**	100.0%

Figure 8**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 6 Students in Reading*



Grade 7 Students

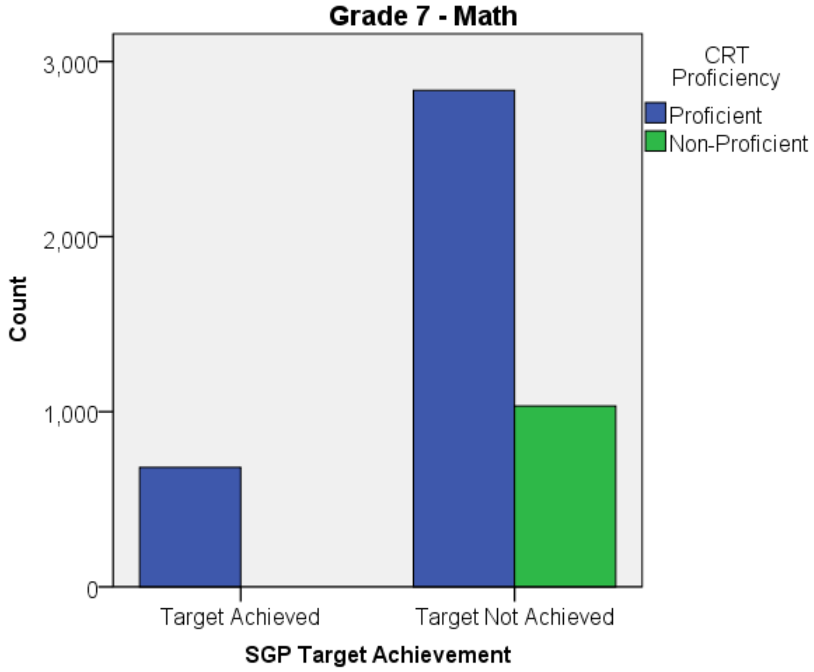
Math. A total of 4,550 students were in our sample from grade 7 for math. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 234.11, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a small effect, *phi* =.228, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 11 and visually in Figure 9, among students who achieved their SGP target, 100% were CRT proficient and 0% were not CRT proficient. Among students who did not achieve their SGP target, 73.3% were CRT proficient and 26.7% were not CRT proficient.

Table 11: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 7 Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**682**	0	682
	%	**100.0%**	0.0%	100.0%
SGP Target *Not* Achieved	***N***	2835	**1033**	3868
	%	73.3%	**26.7%**	100.0%

Figure 9: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 7 Students in Math*



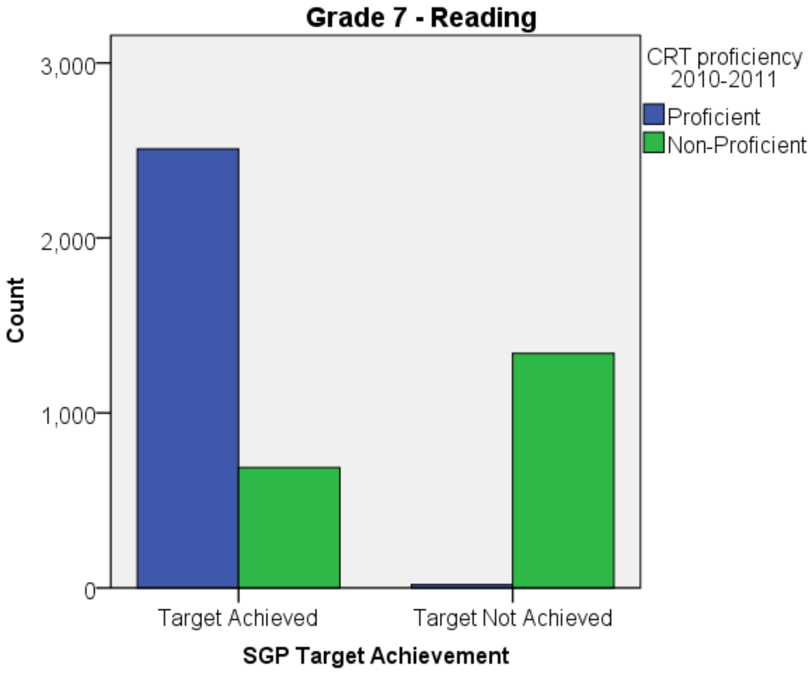
Reading. A total of 4,556 students were in our sample from grade 7 for reading. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 2290.94, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.710, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 12 and visually in Figure 10, among students who achieved their SGP target, 78.5% were CRT proficient and 21.5% were not CRT proficient. Among students who did not achieve their SGP target, 1.4% were CRT proficient and 98.6% were not CRT proficient.

Table 12**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Grade 7 Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**2509**	688	3197
	%	**78.5%**	21.5%	100.0%
SGP Target *Not* Achieved	***N***	19	**1340**	1359
	%	1.4%	**98.6%**	100.0%

Figure 10: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Grade 7 Students in Reading*



Ethnicity Subpopulation

This section shows Chi-square tests for independence on seven ethnicities: African American, Asian, Hispanic, Multi-Racial, Native American, Pacific Islander, and White.

African American Students

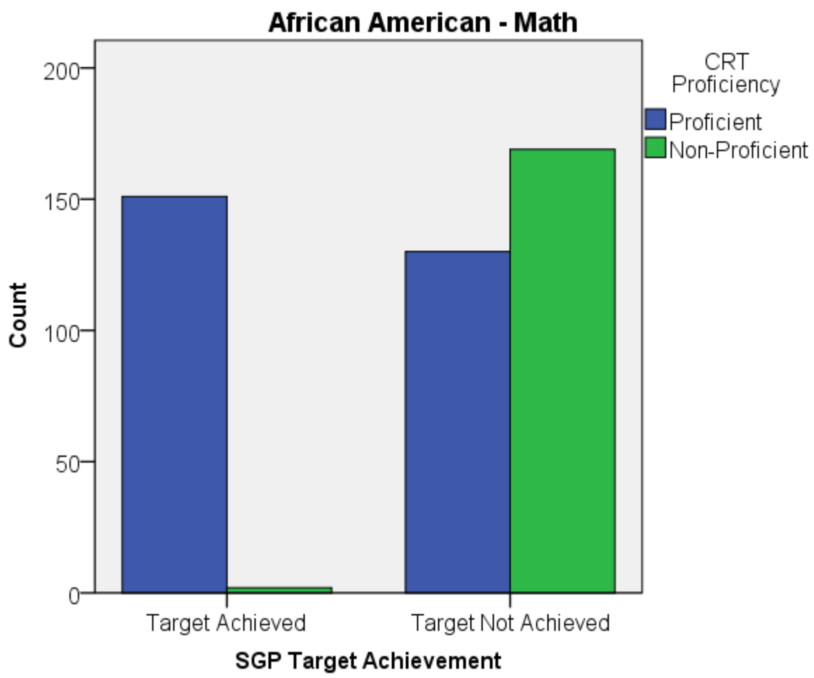
Math. In our math sample, 452 students self-described as African American. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 128.85, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.539, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 13 and visually in Figure 11, among students who achieved their SGP target, 98.7% were CRT proficient and 1.3% were not CRT proficient. Among students who did not achieve their SGP target, 43.5% were CRT proficient and 56.5% were not CRT proficient.

Table 13: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for African American Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	*N*	**151**	2	153
	%	**98.7%**	1.3%	100.0%
SGP Target *Not* Achieved	*N*	130	**169**	299
	%	43.5%	**56.5%**	100.0%

Figure 11: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for African American Students in Math*



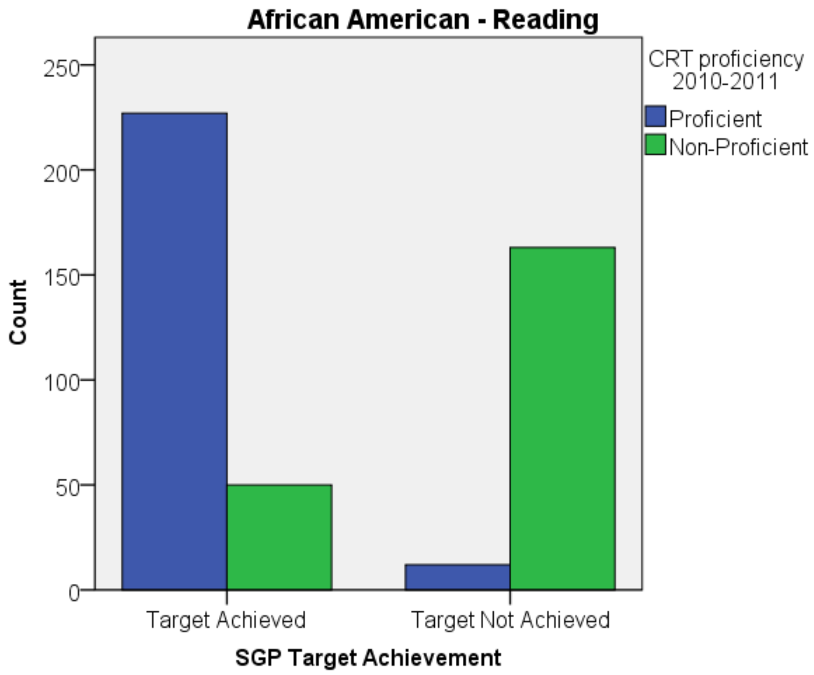
Reading. In our reading sample, 163 students self-described as African American. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 239.70, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.733, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 14 and visually in Figure 12, among students who achieved their SGP target, 81.9% were CRT proficient and 18.1% were not CRT proficient. Among students who did not achieve their SGP target, 6.9% were CRT proficient and 93.1% were not CRT proficient.

Table 14**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for African American Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**227**	50	277
	%	**81.9%**	18.1%	100.0%
SGP Target *Not* Achieved	***N***	12	**163**	175
	%	6.9%	**93.1%**	100.0%

Figure 12**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for African American Students in Reading*



Asian Students

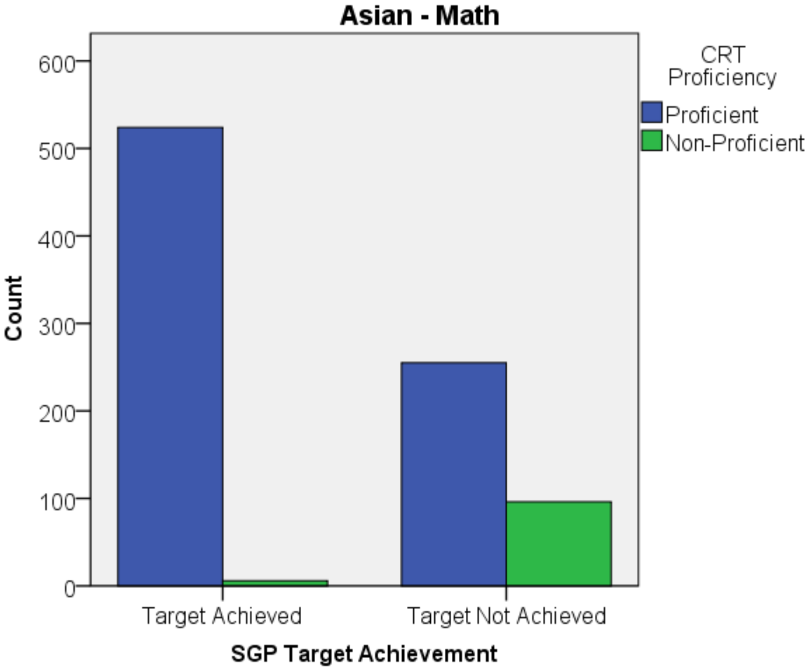
Math. In our math sample, 881 students self-described as Asian. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 139.24, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.401, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 15 and visually in Figure 13, among students who achieved their SGP target, 98.9% were CRT proficient and 1.1% were not CRT proficient. Among students who did not achieve their SGP target, 72.6% were CRT proficient and 27.4% were not CRT proficient.

Table 15**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Asian Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**524**	6	530
	%	**98.9%**	1.1%	100.0%
SGP Target *Not* Achieved	***N***	255	**96**	351
	%	72.6%	**27.4%**	100.0%

Figure 13**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Asian Students in Math*



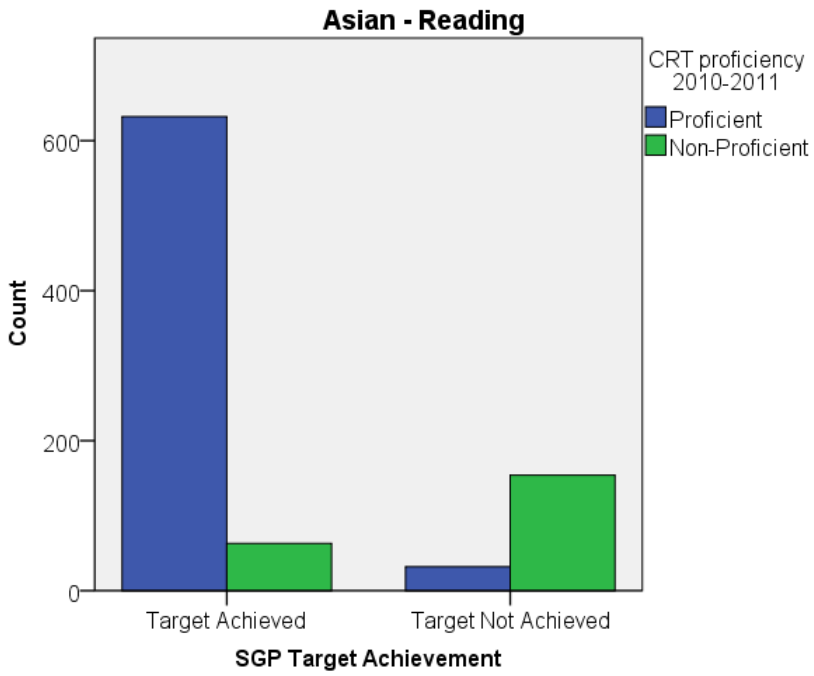
Reading. In our reading sample, 881 students self-described as Asian. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 425.72, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.698, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 16 and visually in Figure 14, among students who achieved their SGP target, 90.9% were CRT proficient and 9.1% were not CRT proficient. Among students who did not achieve their SGP target, 17.2% were CRT proficient and 82.8% were not CRT proficient.

Table 16: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Asian Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**632**	63	695
	%	**90.9%**	9.1%	100.0%
SGP Target *Not* Achieved	***N***	32	**154**	186
	%	17.2%	**82.8%**	100.0%

Figure 14: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Asian Students in Reading*



Hispanic Students

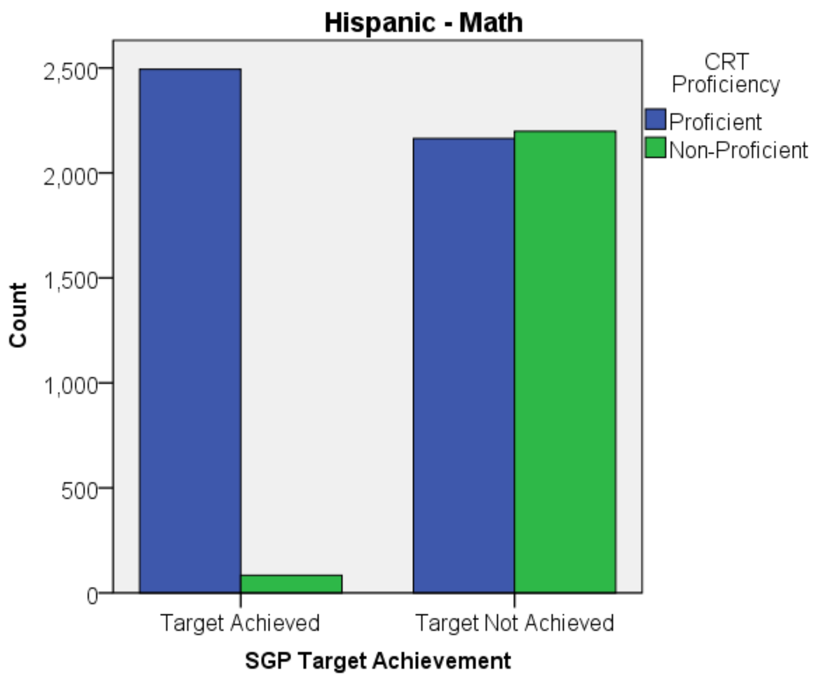
Math. In our math sample, 6,940 students self-described as Hispanic. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1628.79, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.485, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 17 and visually in Figure 15, among students who achieved their SGP target, 96.7% were CRT proficient and 3.3% were not CRT proficient. Among students who did not achieve their SGP target, 49.6% were CRT proficient and 50.4% were not CRT proficient.

Table 17**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Hispanic Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**2494**	84	2578
	%	**96.7%**	3.3%	100.0%
SGP Target *Not* Achieved	***N***	2164	**2198**	4362
	%	49.6%	**50.4%**	100.0%

Figure 15: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Hispanic Students in Math*



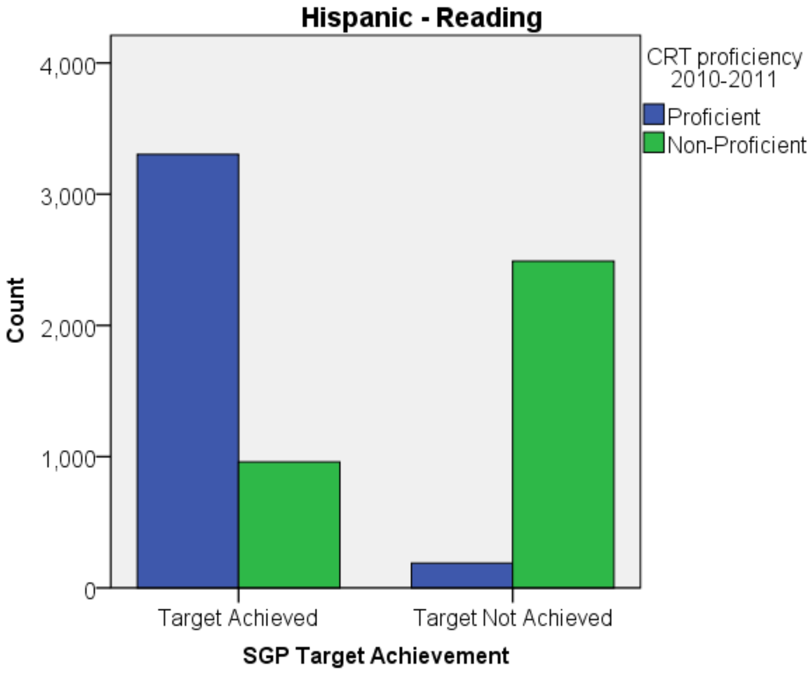
Reading. In our reading sample, 6,942 students self-described as Hispanic. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 3261.13, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.686, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 18 and visually in Figure 16, among students who achieved their SGP target, 77.5% were CRT proficient and 22.5% were not CRT proficient. Among students who did not achieve their SGP target, 7.1% were CRT proficient and 92.9% were not CRT proficient.

Table 18**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Hispanic Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**3303**	960	4263
	%	**77.5%**	22.5%	100.0%
SGP Target *Not* Achieved	***N***	189	**2490**	2679
	%	7.1%	**92.9%**	100.0%

Figure 16: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Hispanic Students in Reading*



Multi-Race Students

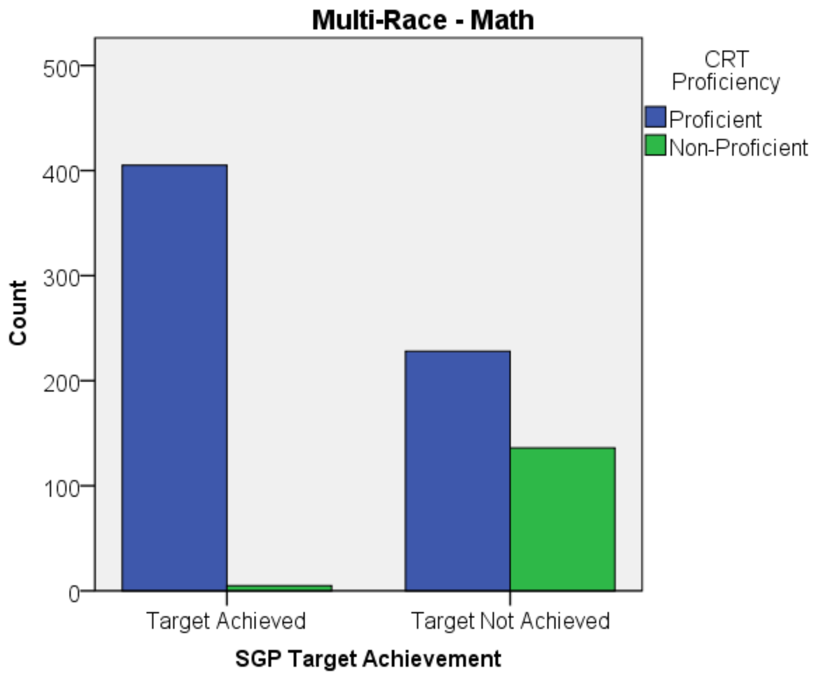
Math. In our math sample, 774 students self-described as Multi-Race. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 166.65, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.467, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 19 and visually in Figure 17, among students who achieved their SGP target, 98.8% were CRT proficient and 1.2% were not CRT proficient. Among students who did not achieve their SGP target, 62.6% were CRT proficient and 37.4% were not CRT proficient.

Table 19: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Multi-Race Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**405**	5	410
	%	**98.8%**	1.2%	100.0%
SGP Target *Not* Achieved	***N***	228	**136**	364
	%	62.6%	**37.4%**	100.0%

Figure 17: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Multi-Race Students in Math*



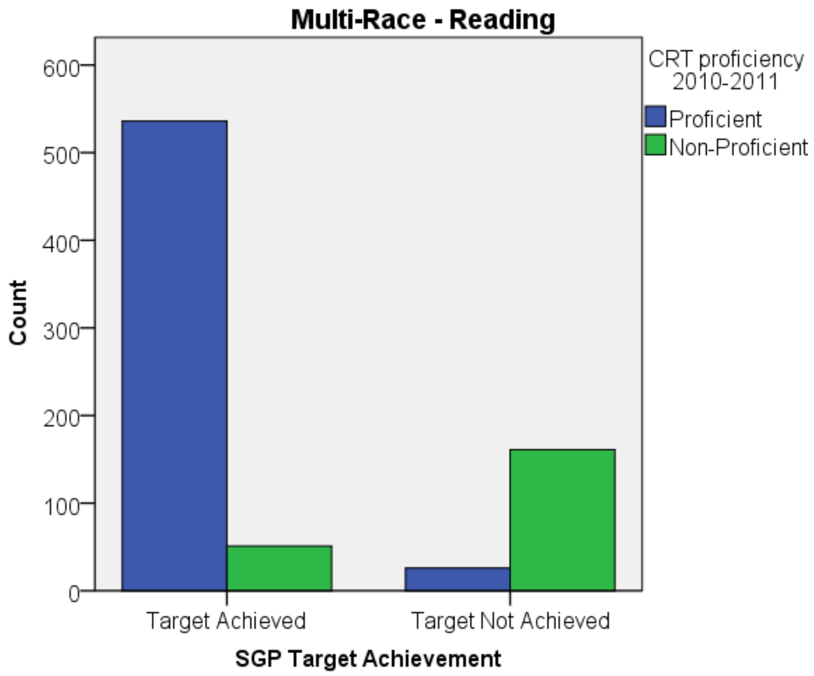
Reading. In our reading sample, 774 students self-described as Multi-Race. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 423.40, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.743, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 20 and visually in Figure 18, among students who achieved their SGP target, 91.3% were CRT proficient and 8.7% were not CRT proficient. Among students who did not achieve their SGP target, 13.9% were CRT proficient and 86.1% were not CRT proficient.

Table 20: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Multi-Race Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	*N*	**536**	51	587
	%	**91.3%**	8.7%	100.0%
SGP Target *Not* Achieved	*N*	26	**161**	187
	%	13.9%	**86.1%**	100.0%

Figure 18: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Multi-Race Students in Reading*



Native American Students

Math. In our math sample, 93 students self-described as Native American. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 64.96, p <

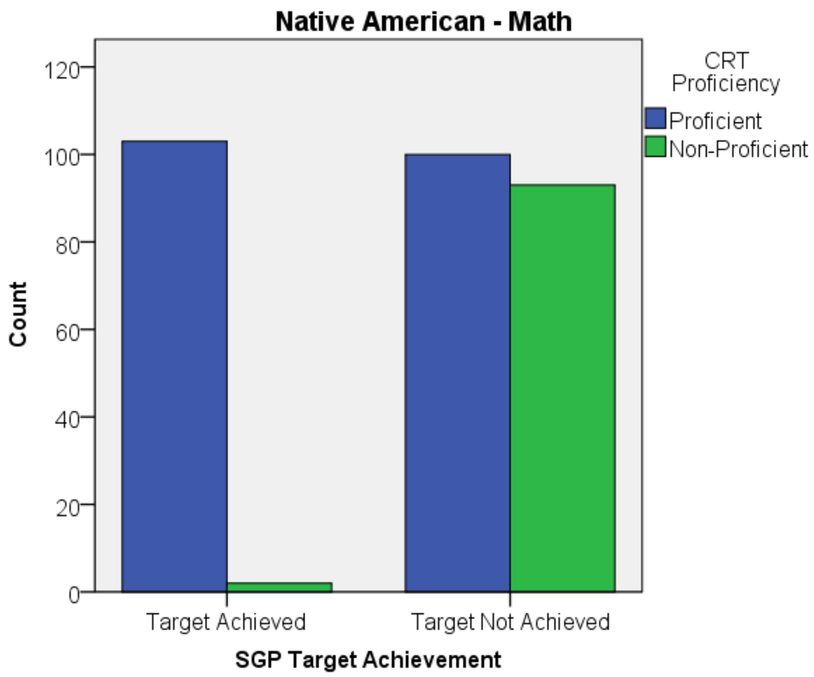
0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.474, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 21 and visually in Figure 19, among students who achieved their SGP target, 98.1% were CRT proficient and 1.9% were not CRT proficient. Among students who did not achieve their SGP target, 51.8% were CRT proficient and 48.2% were not CRT proficient.

Table 21**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Native American Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**103**	2	105
	%	**98.1%**	1.9%	100.0%
SGP Target *Not* Achieved	***N***	100	**93**	193
	%	51.8%	**48.2%**	100.0%

Figure 19**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Native American Students in Math*



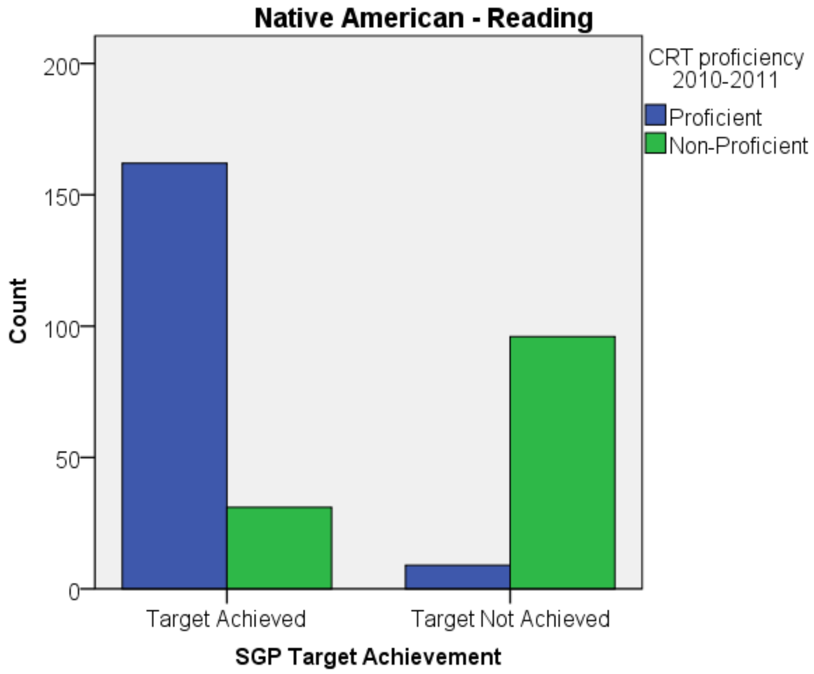
Reading. In our reading sample, 93 students self-described as Native American. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 154.88, *p* < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.728, *p* < .001 according to Cohen's (1988) criteria.

As displayed in Table 22 and visually in Figure 19, among students who achieved their SGP target, 83.9% were CRT proficient and 16.1% were not CRT proficient. Among students who did not achieve their SGP target, 8.6% were CRT proficient and 91.4% were not CRT proficient.

Table 22**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Native American Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**162**	31	193
	%	**83.9%**	16.1%	100.0%
SGP Target *Not* Achieved	***N***	9	**96**	105
	%	8.6%	**91.4%**	100.0%

Figure 20**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Native American Students in Reading*



Pacific Islander Students

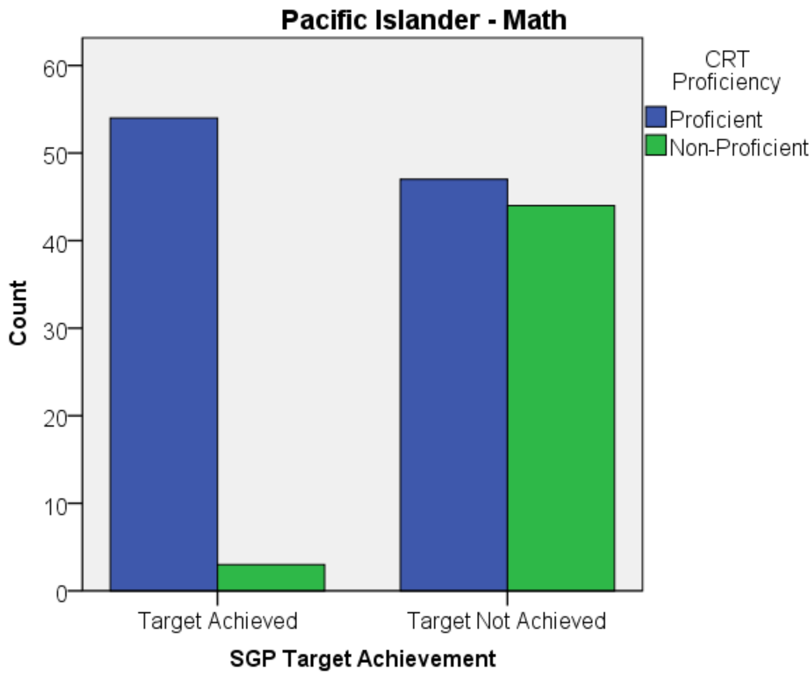
Math. In our math sample, 148 students self-described as Pacific Islander. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 28.07, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.450, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 23 and visually in Figure 21, among students who achieved their SGP target, 94.7% were CRT proficient and 5.3% were not CRT proficient. Among students who did not achieve their SGP target, 51.6% were CRT proficient and 48.4% were not CRT proficient.

Table 23: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Pacific Islander Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**54**	3	57
	%	**94.7%**	5.3%	100.0%
SGP Target *Not* Achieved	***N***	47	**44**	91
	%	51.6%	**48.4%**	100.0%

Figure 21: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Pacific Islander Students in Math*



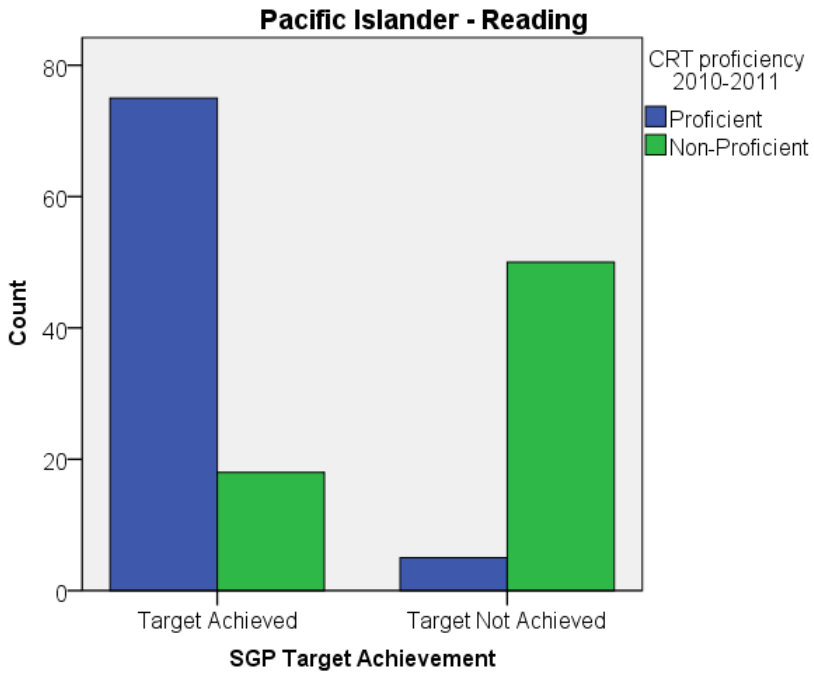
Reading. In our reading sample, 148 students self-described as Pacific Islander. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 68.40, $p <$ 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.694, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 24 and visually in Figure 22, among students who achieved their SGP target, 80.6% were CRT proficient and 19.4% were not CRT proficient. Among students who did not achieve their SGP target, 9.1% were CRT proficient and 90.9% were not CRT proficient.

Table 24: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Pacific Islander Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**75**	18	93
	%	**80.6%**	19.4%	100.0%
SGP Target *Not* Achieved	***N***	5	**50**	55
	%	9.1%	**90.9%**	100.0%

Figure 22: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Pacific Islander Students in Reading*



White Students

Math. In our math sample, 8,898 students self-described as White. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1704.66, $p < 0.001$.

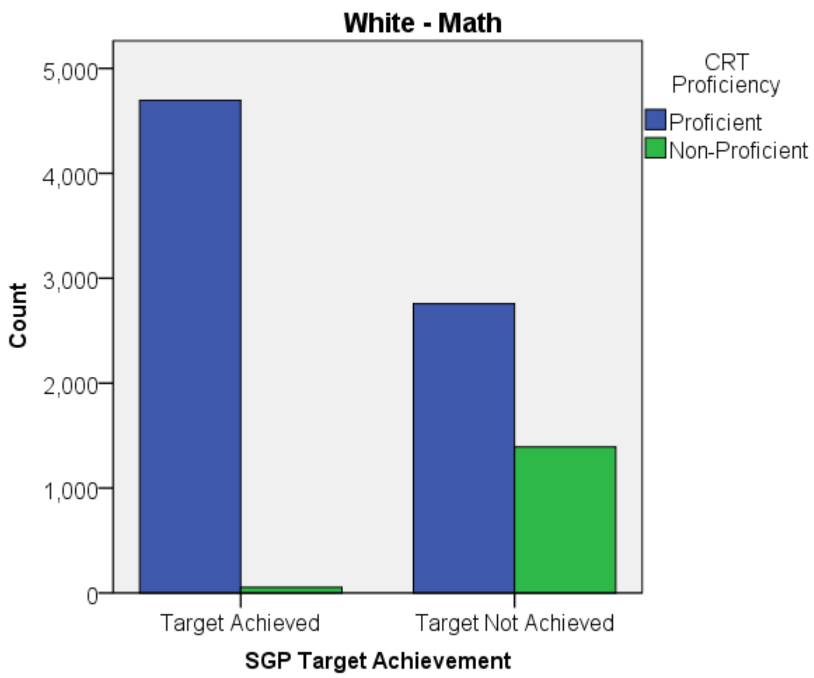
That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.438, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 25 and visually in Figure 23, among students who achieved their SGP target, 98.8% were CRT proficient and 1.2% were not CRT proficient. Among students who did not achieve their SGP target, 66.4% were CRT proficient and 33.6% were not CRT proficient.

Table 25**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for White Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**4695**	55	4750
	%	**98.8%**	1.2%	100.0%
SGP Target *Not* Achieved	***N***	2756	**1392**	4148
	%	66.4%	**33.6%**	100.0%

Figure 23**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for White Students in Math*



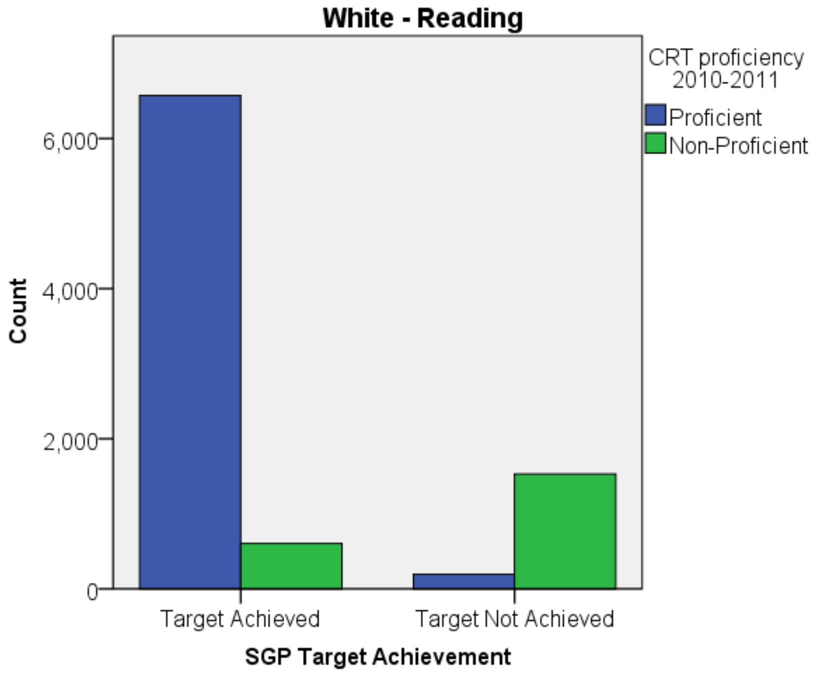
Reading. In our reading sample, 8,900 students self-described as White. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 4920.73, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.744, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 26 and visually in Figure 24, among students who achieved their SGP target, 91.6% were CRT proficient and 8.4% were not CRT proficient. Among students who did not achieve their SGP target, 11.2% were CRT proficient and 88.8% were not CRT proficient.

Table 26**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for White Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**6573**	604	7177
	%	**91.6%**	8.4%	100.0%
SGP Target *Not* Achieved	***N***	193	**1530**	1723
	%	11.2%	**88.8%**	100.0%

Figure 24: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for White Students in Reading*



Special Subpopulations

This section shows Chi-square tests for independence on three subpopulations: students who have Limited English Proficiency (LEP), students with an Individualized Education Plans (IEP), and students who receive Free or Reduced Lunch (FRL).

LEP Students

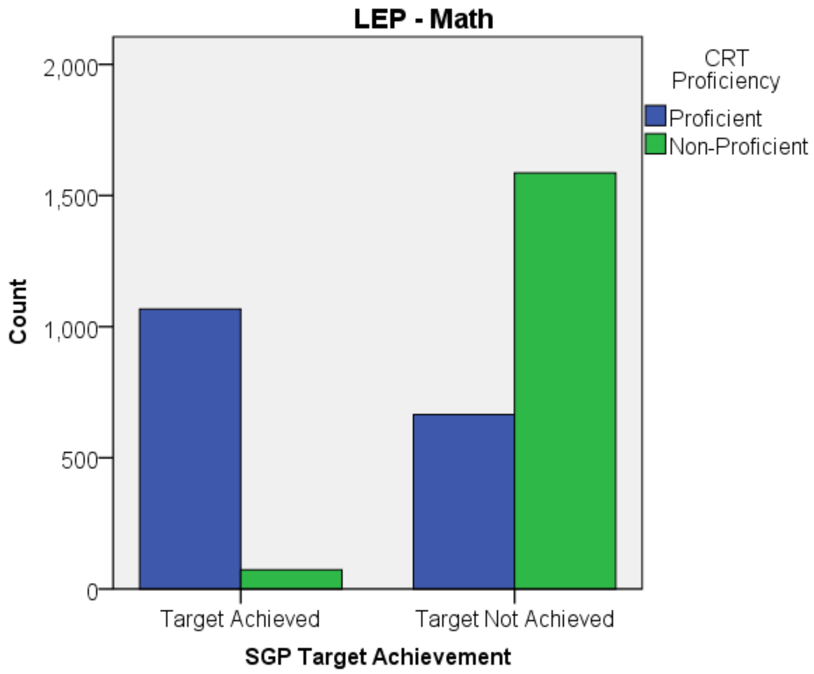
Math. In our math sample, 3,391 students were in the LEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1239.97, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.605, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 27 and visually in Figure 25, among students who achieved their SGP target, 93.6% were CRT proficient and 6.4% were not CRT proficient. Among students who did not achieve their SGP target, 29.5% were CRT proficient and 70.5% were not CRT proficient.

Table 27: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for LEP Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	*N*	**1067**	73	1140
	%	**93.6%**	6.4%	100.0%
SGP Target *Not* Achieved	*N*	665	**1586**	2251
	%	29.5%	**70.5%**	100.0%

Figure 25: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for LEP Students in Math*



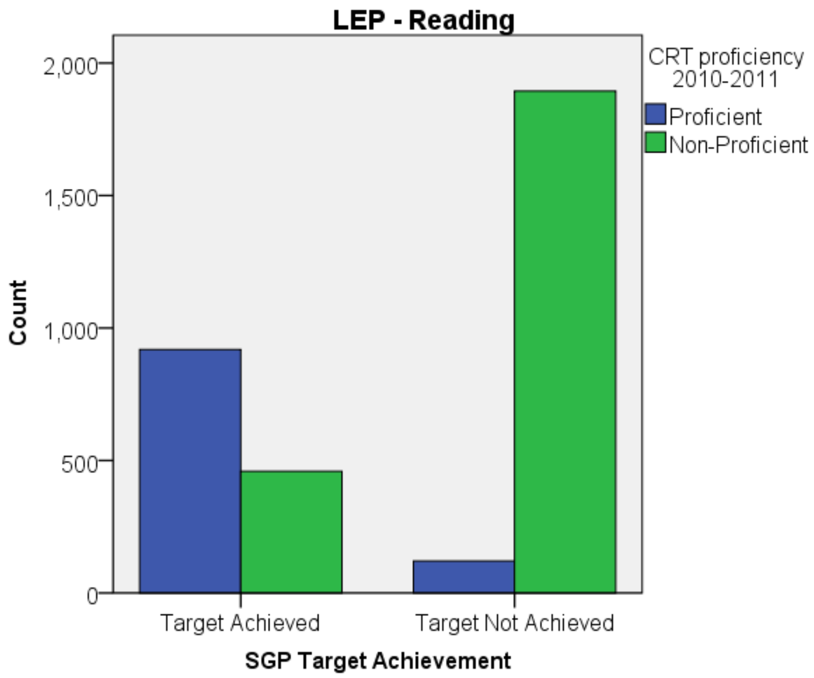
Reading. In our reading sample, 3,393 students were in the LEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1414.99, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.646, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 28 and visually in Figure 26, among students who achieved their SGP target, 66.7% were CRT proficient and 33.3% were not CRT proficient. Among students who did not achieve their SGP target, 6.0% were CRT proficient and 94.0% were not CRT proficient.

Table 28: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for LEP Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**919**	459	1378
	%	**66.7%**	33.3%	100.0%
SGP Target *Not* Achieved	***N***	121	**1894**	2015
	%	6.0%	**94.0%**	100.0%

Figure 26: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for LEP Students in Reading*



Non-LEP Students

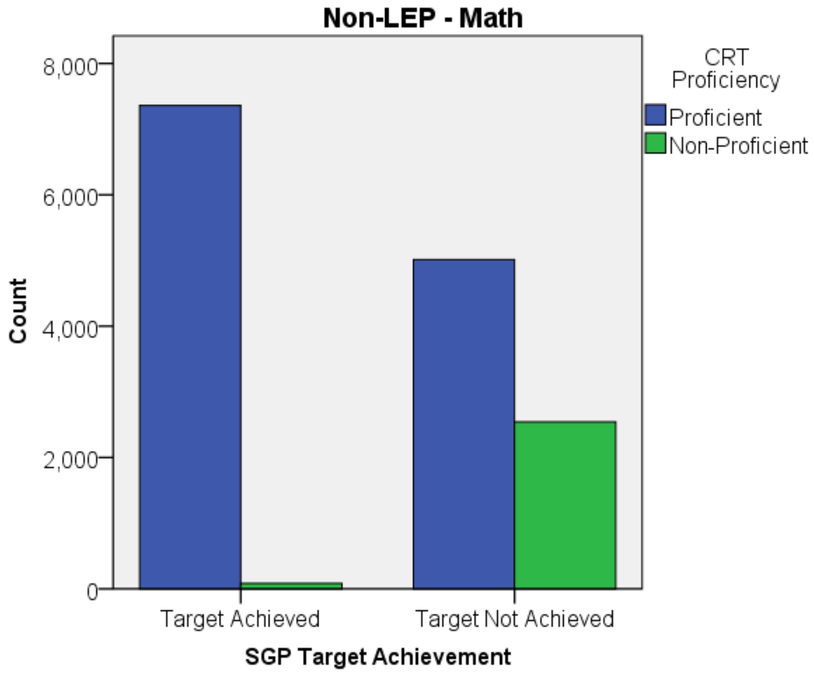
Math. In our math sample, 7,557 students were in the Non-LEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 2741.81, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.428, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 29 and visually in Figure 27, among students who achieved their SGP target, 98.9% were CRT proficient and 1.1% were not CRT proficient. Among students who did not achieve their SGP target, 66.4% were CRT proficient and 33.6% were not CRT proficient.

Table 29**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-LEP Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**7359**	84	7443
	%	**98.9%**	1.1%	100.0%
SGP Target *Not* Achieved	***N***	5015	**2542**	7557
	%	66.4%	**33.6%**	100.0%

Figure 27**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-LEP Students in Math*



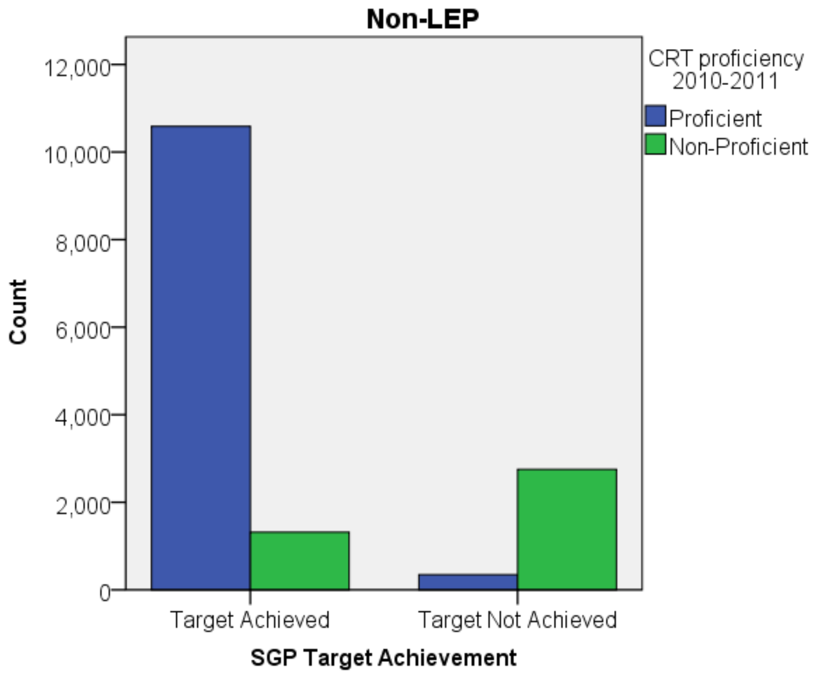
Reading. In our reading sample, 15,002 students were in the Non-LEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 7516.30, *p* < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.708, *p* < .001 according to Cohen's (1988) criteria.

As displayed in Table 30 and visually in Figure 28, among students who achieved their SGP target, 88.9% were CRT proficient and 11.1% were not CRT proficient. Among students who did not achieve their SGP target, 11.1% were CRT proficient and 88.9% were not CRT proficient.

Table 30: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-LEP Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**10589**	1318	11907
	%	**88.9%**	11.1%	100.0%
SGP Target *Not* Achieved	***N***	345	**2750**	3095
	%	11.1%	**88.9%**	100.0%

Figure 28: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-LEP Students in Reading*



IEP Students

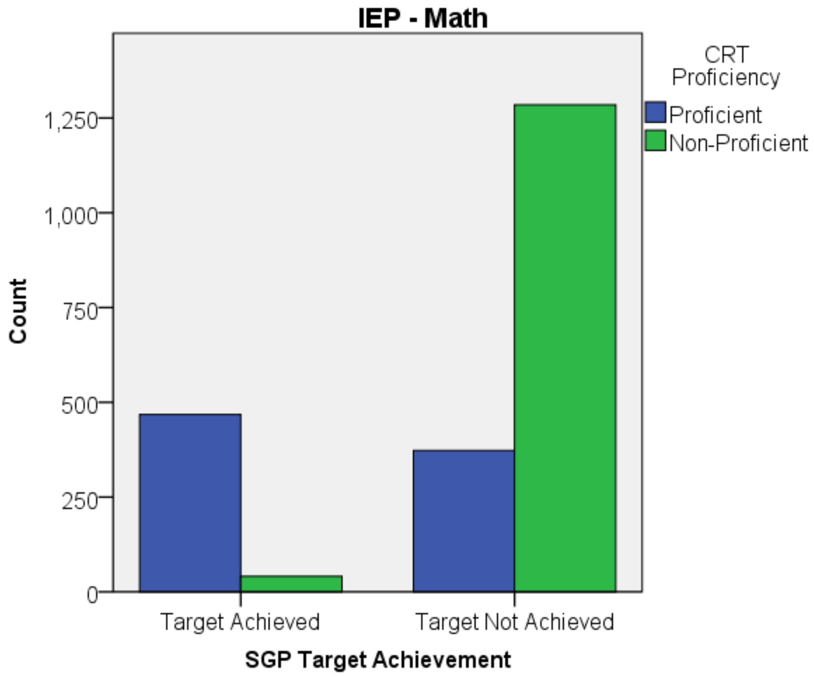
Math. In our math sample, 2,167 students were in the IEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 788.01, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.604, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 31 and visually in Figure 29, among students who achieved their SGP target, 91.9% were CRT proficient and 8.1% were not CRT proficient. Among students who did not achieve their SGP target, 22.5% were CRT proficient and 77.5% were not CRT proficient.

Table 31: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for IEP Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	*N*	**468**	41	509
	%	**91.9%**	8.1%	100.0%
SGP Target *Not* Achieved	*N*	373	**1285**	1658
	%	22.5%	**77.5%**	100.0%

Figure 29: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for IEP Students in Math*



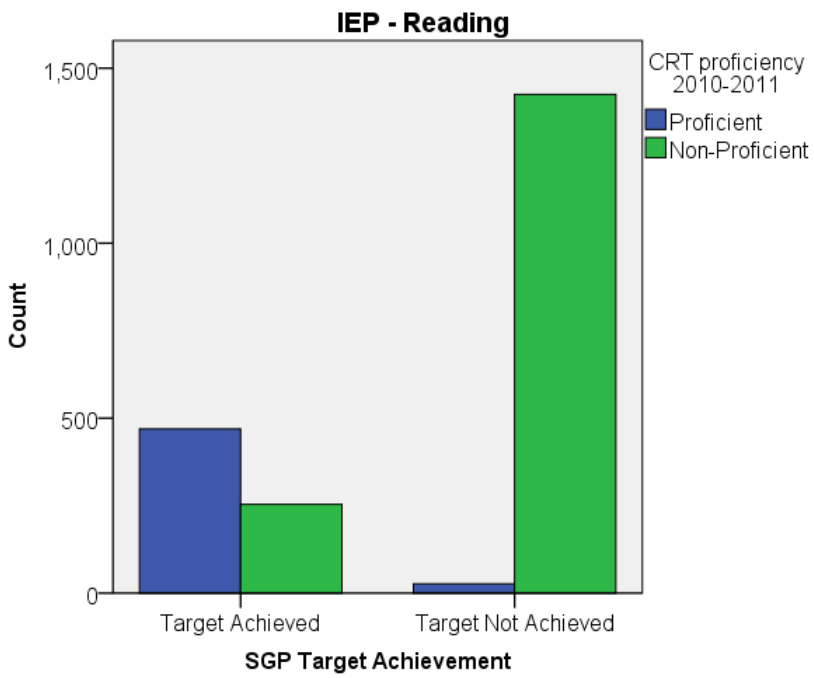
Reading. In our reading sample, 2,175 students were in the IEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1084.95, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.707, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 32 and visually in Figure 30, among students who achieved their SGP target, 64.9% were CRT proficient and 35.1% were not CRT proficient. Among students who did not achieve their SGP target, 1.9% were CRT proficient and 98.1% were not CRT proficient.

Table 32**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for IEP Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**469**	254	723
	%	**64.9%**	35.1%	100.0%
SGP Target *Not* Achieved	***N***	27	**1425**	1452
	%	1.9%	**98.1%**	100.0%

Figure 30**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for IEP Students in Reading*



Non-IEP Students

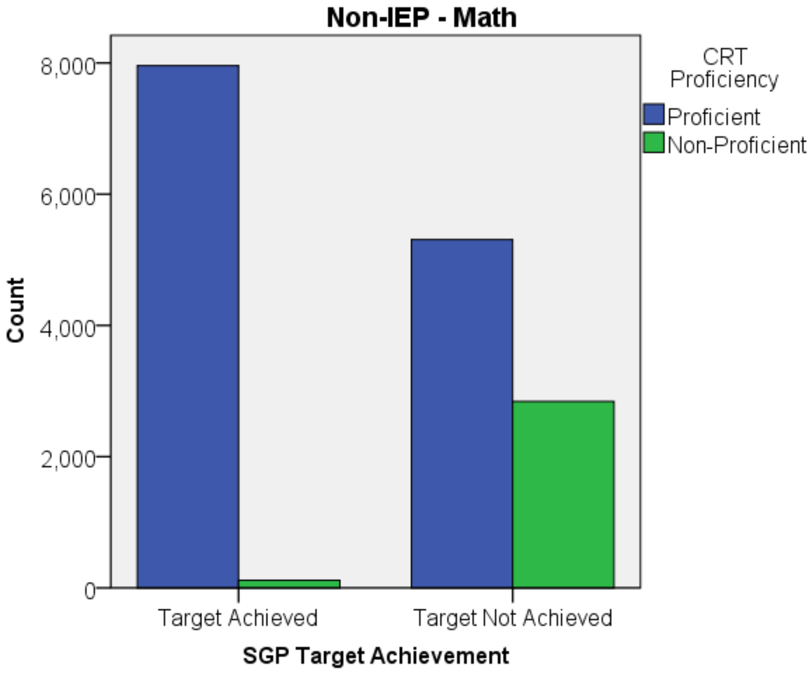
Math. In our math sample, 16,224 students were in the Non-IEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 3040.46, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.433, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 33 and visually in Figure 31, among students who achieved their SGP target, 98.6% were CRT proficient and 1.4% were not CRT proficient. Among students who did not achieve their SGP target, 65.1% were CRT proficient and 34.9% were not CRT proficient.

Table 33**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-IEP Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**7958**	116	8074
	%	**98.6%**	1.4%	100.0%
SGP Target *Not* Achieved	***N***	5307	**2843**	8150
	%	65.1%	**34.9%**	100.0%

Figure 31**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-IEP Students in Math*



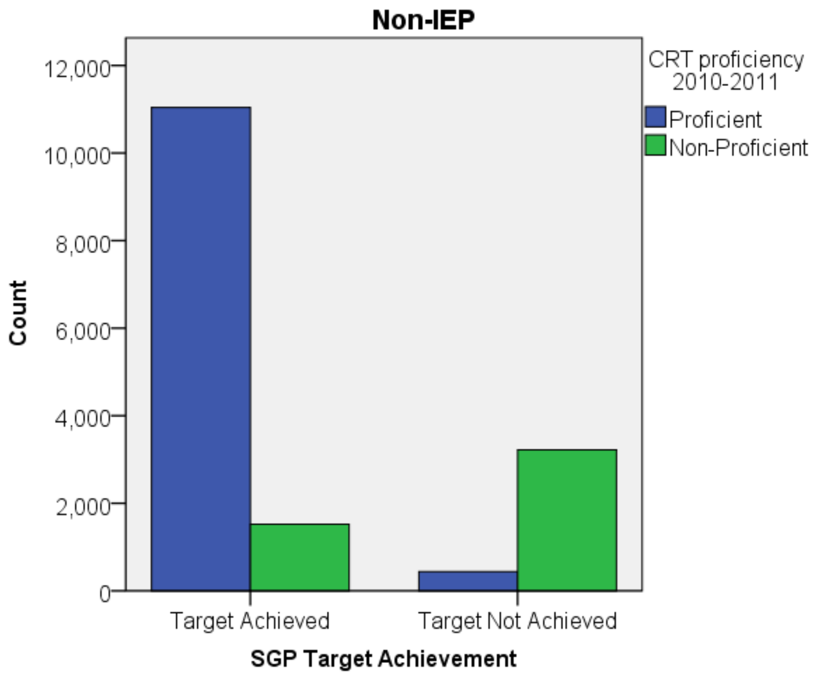
Reading. In our reading sample, 16,220 students were in the Non-IEP subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 7879.58, *p* < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.697, *p* < .001 according to Cohen's (1988) criteria.

As displayed in Table 34 and visually in Figure 32, among students who achieved their SGP target, 87.9% were CRT proficient and 12.1% were not CRT proficient. Among students who did not achieve their SGP target, 12.0% were CRT proficient and 88.0% were not CRT proficient.

Table 34: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-IEP Students in Reading*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**11039**	1523	12562
	%	**87.9%**	12.1%	100.0%
SGP Target *Not* Achieved	***N***	439	**3219**	3658
	%	12.0%	**88.0%**	100.0%

Figure 32: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-IEP Students in Reading*



FRL Students

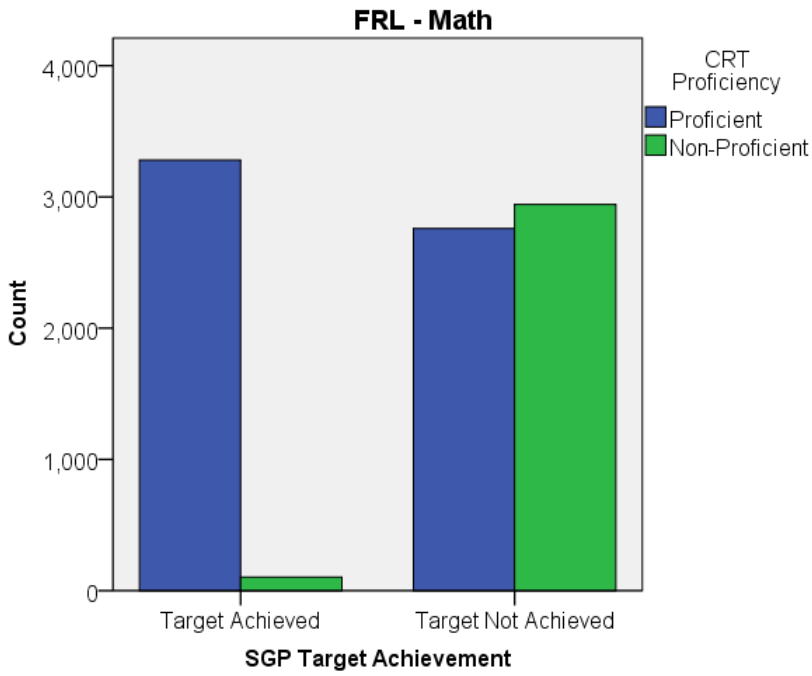
Math. In our math sample, 9,087 students were in the FRL subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 2242.08, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.497, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 35 and visually in Figure 33, among students who achieved their SGP target, 96.9% were CRT proficient and 3.1% were not CRT proficient. Among students who did not achieve their SGP target, 48.4% were CRT proficient and 51.6% were not CRT proficient.

Table 35**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for FRL Students in Math*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**3281**	104	3385
	%	**96.9%**	3.1%	100.0%
SGP Target *Not* Achieved	***N***	2760	**2942**	5702
	%	48.4%	**51.6%**	100.0%

Figure 33**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for FRL Students in Math*



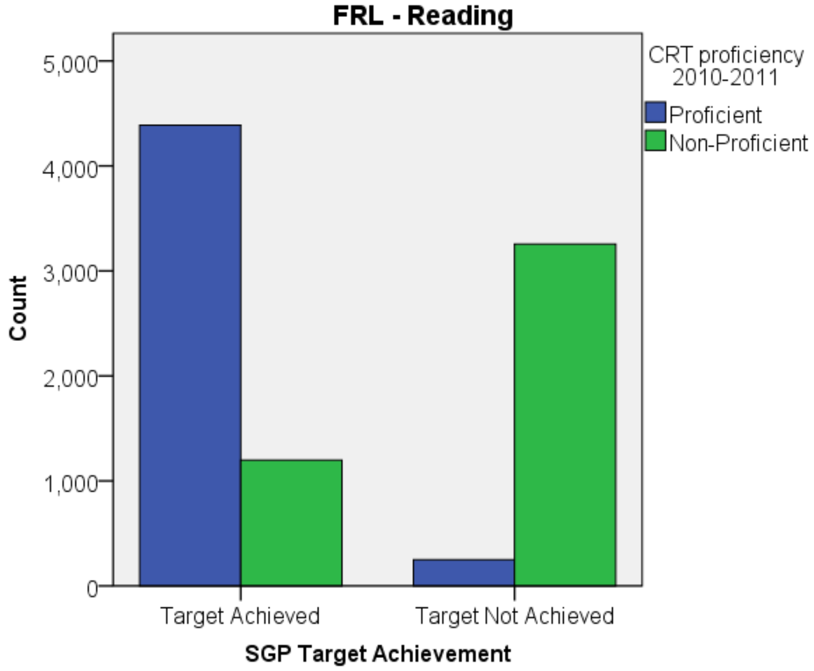
Reading. In our reading sample, 9,091 students were in the FRL subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 4396.71, p < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.696, p < .001 according to Cohen's (1988) criteria.

As displayed in Table 36 and visually in Figure 34, among students who achieved their SGP target, 78.6% were CRT proficient and 21.4% were not CRT proficient. Among students who did not achieve their SGP target, 7.1% were CRT proficient and 92.9% were not CRT proficient.

Table 36**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for FRL Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**4388**	1198	5586
	%	**78.6%**	21.4%	100.0%
SGP Target *Not* Achieved	***N***	249	**3256**	3505
	%	7.1%	**92.9%**	100.0%

Figure 34: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for FRL Students in Reading*



Non-FRL Students

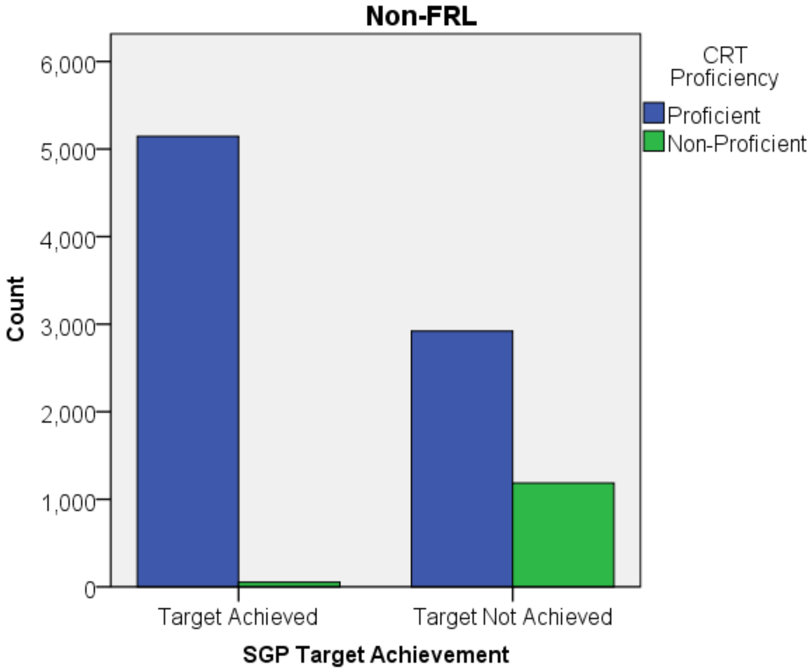
Math. In our math sample, 9,304 students were in the Non-FRL subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 1540.59, *p* < 0.001. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a medium effect, *phi* =.407, *p* < .001 according to Cohen's (1988) criteria.

As displayed in Table 37 and visually in Figure 35, among students who achieved their SGP target, 99.0% were CRT proficient and 1.0% were not CRT proficient. Among students who did not achieve their SGP target, 71.1% were CRT proficient and 28.9% were not CRT proficient.

Table 37: *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-FRL Students in Math*

		CRT Proficient	CRT Non-Proficient	Total
SGP Target Achieved	***N***	**5145**	53	5198
	%	**99.0%**	1.0%	100.0%
SGP Target *Not* Achieved	***N***	2920	**1186**	4106
	%	71.1%	**28.9%**	100.0%

Figure 35: *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-FRL Students in Math*



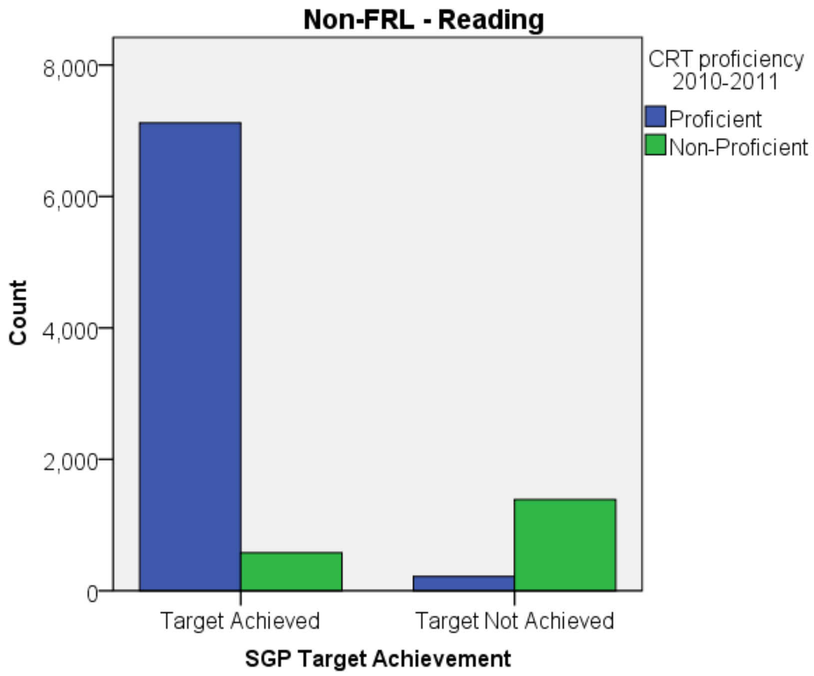
Reading. In our reading sample, 9,304 students were in the Non-FRL subgroup. A Chi-square test for independence (with Yates Continuity Correction) indicated a significant association between SGP target achievement and CRT proficiency status, χ^2 (1) = 4961.91, $p < 0.001$. That is, there is a significant difference between the CRT proficiency of students who score at or above their SGP target and the CRT proficiency of students who score lower than their SGP target. The effect size of the association between SGP target achievement and CRT proficiency demonstrated a large effect, *phi* =.731, $p < .001$ according to Cohen's (1988) criteria.

As displayed in Table 38 and visually in Figure 36, among students who achieved their SGP target, 92.5% were CRT proficient and 7.5% were not CRT proficient. Among students who did not achieve their SGP target, 13.5% were CRT proficient and 86.5% were not CRT proficient.

Table 38**:** *Crosstabulation Table of SGP Target Achievement by CRT Proficiency for Non-FRL Students in Reading*

		CRT Proficient	**CRT Non-Proficient**	**Total**
SGP Target Achieved	***N***	**7120**	579	7699
	%	**92.5%**	7.5%	100.0%
SGP Target *Not* Achieved	***N***	217	**1388**	1605
	%	13.5%	**86.5%**	100.0%

Figure 36**:** *Bar Chart Displaying SGP Target Achievement by CRT Proficiency for Non-FRL Students in Reading*



CHAPTER FIVE

DISCUSSION

This dissertation compares some growth models that are commonly used by states and districts to measure growth in achievement scores. There are several areas in the findings of the study that are noteworthy.

Significance and Conclusion

The results of this dissertation will make important contributions to ongoing studies on the use of growth models in state accountability as well as research on teacher and principal effectiveness. The expectation is to have achieved a new and more complete description of growth models, including a detailed explanation of the Student Growth Percentile method, as well as improving public education in the United States by placing highly skilled and effective teachers in all classrooms via foundational questions on educator effectiveness.

Many policy makers, stakeholders, and even some education researchers believe that the results from growth models should be used for accountability purposes, to measure school or classroom effectiveness. The focus in this dissertation was on growth projections for diagnostic purposes. Such projections are particularly important and useful at a diagnostic level since students who are at risk for not meeting NCLB goals can be identified early and appropriate remedial action can be taken.

While variants of growth models are currently in use, there has been no systematic study comparing the relative merits of these models for assessing students' future projected scores. Such projections play a role in many policy decisions. Many policy makers, stakeholders, and even some education researchers believe that the results from growth models should be used for accountability purposes, to measure school or classroom effectiveness. The focus in this dissertation was on growth projections for diagnostic purposes. Such projections are particularly important and useful at a diagnostic level since students who are at risk for not meeting NCLB goals can be identified early and appropriate remedial action can be taken. Proponents of the use of growth models for accountability purposes are either unaware of or ignoring the potential causality problems. When these models are used for accountability purposes, concerns about the validity of the inferences have been voiced by many scholars.

This dissertation will help not just Nevada, but all states planning to develop and implement a statewide student achievement growth model. Here in the Nevada, the Nevada Department of Education has adopted a growth model to enhance the understanding and transparency of student achievement in the state. This research will serve as a guide to growth model implementation and use measuring school and teacher effectiveness which is a highly complex and hugely relevant area of research across the nation. While variants of growth models are currently in use, there has been no systematic study comparing the relative merits of these models for assessing students' future projected scores.

The contribution to the research literature and policies will also greatly enhance the district and state education communities' efforts to improve student learning, school improvement, and teacher effectiveness. Estimating the effects of teachers by modeling longitudinal data on student achievement raises a number of important statistical and psychometric questions and issues and requires careful consideration about how these issues should be addressed. This study will make important contributions to the practical and technical use of SGP on using student learning.

Limitations and Future Areas of Research

This dissertation provides useful information on the accuracy of the projected scores obtained from different models. However, using empirical data, observed scores were treated as "true values" and one-year projected scores as "estimates." However, it can be argued that true values (observed scores in this study) are also estimates that contain a degree of measurement error. Although the analyses and results from this study have provided a new level of understanding, future studies should investigate which modeling approach yields the most accurate projected scores by designing a simulation study where the true parameters are known with no measurement error.

REFERENCES

Cohen, J. W. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Auty, W., Bielawski, P., Deeter, T., Hirata, G., Hovanetz-Lassila, C., Rheim, J., et al. (2008). *Implementer's guide to Growth Models*. Washington D.C.: The Council of Chief State School Officers. Retrieved from http://www.ccsso.org/projects/scass/projects/accountability_systems_and_reporting_consortium/8705.cfm

Ballou, D. (2002). *Sizing up test scores.* Education Next. Retrieved from http://www.educationnext.org/2002/16.html

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37 - 65.

Betebenner, D., Shang, Y. (2007). *Reference growth charts for educational outcomes.* Lynch School of Education, Dept of Educational Research, Measurement, and Evaluation. Boston. Retrieved on from http://www2.bc.edu/~shangyi/referencegrowthcharts.pdf

Betebenner, D. (2008a). *A primer on student growth percentiles*. New Hampshire: National Center for the Improvement of Educational Assessment. Retrieved from http://www.cde.state.co.us/cdedocs/Research/PDF/Aprimeronstudentgrowthpercentiles.pdf

Betebenner, D. (2008b). *Colorado's growth model: A harmonious marriage of normative and standards-based growth*. Dover: National Center for the Improvement of Educational Assessment. Retrieved from http://www.ccsso.org/content/PDFs/73_Betebenner.pdf

Betebenner, D. (2008c). *Norm- and criterion-referenced student growth*. Dover: National Center for the Improvement of Educational Assessment. Retrieved on from http://www.ccsso.org/content/PDFs/194_Bbetebenner.pdf

Betebenner, D. W., & Van Iwaarden, A. (2011). SGP: An R Package for the Calculation and Visualization of Student Growth Percentiles. R package version 0.4-5.0. URL http://cran.r-project.org/web/packages/SGP/index.html.

Betebenner, D., Shang, Y. (2008). *On defining a year's growth*. Paper presented at the 2008 AERA Annual Meeting, New York, NY (pp. 1-21) Retrieved from http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date

Cohen, M. (2002). Assessment and accountability: Lessons from the past, challenges for the future. Paper presented at the No Child Left Behind: What will it take? Conference hosted by the Thomas B. Fordham Foundation. Retrieved from http://www.edexcellence.net/foundation/topic/topic.cfm?topic_id=5

Colorado Department of Education. (2009). Education reform: Colorado growth model. Retrieved from http://cde.state.co.us/cdeassess/growthmodel.html

Council of Chief State School Officers. (2009). *Guide to United States Department of Education Growth Model Pilot Program 2005 – 2008*. Publications, CCSSO; Washington, DC. Retrieved from http://www.ccsso.org/content/pdfs/CCSSO%20ASR%2009%20Guide%20to%20Growth%20Pilot%20projects.pdf

Council of Chief State School Officers. (2010). *State Growth Models for School Accountability: Progress on Development and Reporting Measures of Student Growth.* Publications, CCSSO; Washington, DC. Retrieved from http://www.ccsso.org/Documents/2010/State_Growth_Models_2010.pdf

Center for Public Education. (2010). Measuring student growth: A guide for informed decision making. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Policies/Measuring-student-growth-At-a-glance/Measuring-student-growth-A-guide-to-informed-decision-making.html

Doran, H. C., & Cohen, J. (2005). The confounding effect of linking bias on gains estimated from value-added models. In R. Lissitz (Ed.), *Value-added models in education: Theory and applications.* Maple Grove, MN: JAM Press.

Goldschmidt, P., & Choi, K. (2007). *The practical benefits of Growth Models for accountability and the limitations under NCLB*. Northridge: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from http://www.ccsso.org/content/pdfs/P%20goldschmidt%20Policy%20Brief%20v3.pdf

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., & Williams, A. (2005). *Policymakers' guide to Growth Models for school accountability: How do accountability models differ?* The Council of Chief State School Officers. Washington, DC. 1 – 36. Retrieved from http://www.ccsso.org/content/pdfs/Growth%20Models%20Policymaker%20Guide%202005.pdf

Goldstein, J. (2006). Measuring growth in student achievement: Can different statistical models lead to different consequences for schools? Ph.D. dissertation, University of Connecticut. Retrieved on September 8, 2010, from Dissertations & Theses @ University of Connecticut (Publication No. AAT 3217033).

Gong, B., Perie, M., & Dunn, J. (2006). *Using student longitudinal growth measures for school accountability under* No Child Left Behind: *An update to inform design decisions.* Retrieved on March 13, 2010, from http://www.nciea.org/publications/GrowthModelUpdate_BGMAPJD07.pdf

Government Accountability Office. (2008). No child left behind act: States face challenges measuring academic growth that education's initiatives may help address. Retrieved October 3, 2009, from http://www.gao.gov/new.items/d06661.pdf.

Hill, R., Gong, B, Marion, S, DePascale, C., Dunn, J. and Simpson, M. A. (2006). Using value Tables to explicitly value growth. In R. Lissitz (Ed.), *Longitudinal and value added models of student performance* (pp. 255-290). Maple Grove, MN: JAM Press.

Ho, A.D. (2008). *The problem with "proficiency": Limitations of statistics and policy under no child left behind.* Educational Researcher. 37, 351 – 360.

Hibpshman, Terry. (September 2004). *A Review of Value-Added Models*. Kentucky Education Professional Standards Board.

Jennings, J., & Corcoran, S. (2009). Beware of geeks bearing formulas. *Phi Delta Kappan, 90*(9), 635 - 639.

Keeves, J., Hungi, M., & Afrassa, T. (2005). Measuring value added effects across schools: Should schools be compared in performance? *Studies in Educational Evaluation, 31,* 247 - 266.

Lissitz, R, W. & Huynh, H. (2003). Vertical equating for state assessments: issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation*, 8(10). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=10.

McCaffrey, D.F., Lockwood, J.R., Koretz, D.M. & Hamilton, L.S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: The Rand Corporation.

No Child Left Behind Act of 2001. Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

O'Malley, K. (2008). *Review of student growth models used by states*. Retrieved on October, 8, 2008 from http://www.pearsonassessments.com/NR/rdonlyres/BBBD784B-B0BC-4B0BACB2-7069F1239BB3/0/StudentGrowthModels_0608.pdf

O'Malley, K., Auty, W., Bielawski, P., Bernstein, R., Boatman, T., Deeter, T., Goldschmidt, P., & et al. (2009). *Guide to United States Department of Education growth model pilot program 2005 – 2008.* Council of Chief State School Officers. Publications, Washington DC. 1 – 47. Retrieved from http://www.ccsso.org/Projects/Accountability_Systems/Resources/

R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for statistical computing. Retrieved on March 8, 2010, from http://www.R-project.org

Sanders, W. (2003). *Beyond No Child Left Behind*. *Value-added approaches to school accountability: Results and lessons from an evaluation.* Presented at the Annual Meeting of the American Educational Research Association in Chicago, IL. Retrieved from http://www.sas.com/govedu/edu/no-child.pdf

Tennessee Department of Education. (2006) Proposal to the U.S. Department of Education. NCLB Growth Model Pilot Program; Washington, DC, 1 – 34.

U.S. Department of Education. (2002). *No Child Left Behind: A desktop reference.* Education Publications Center. Washington, DC.

U.S. Department of Education. (2005). *Growth Models: Flexibility and Accountability*. Retrieved from http://www.ed.gov/admins/lead/account/growthmodel/factsheet.html.

U.S. Department of Education (2006) *No Child Left Behind Growth Models: Ensuring grade-level proficiency for all students by 2014.* Retrieved from www.ed.gov/adminis/lead/account/growthmodels/proficiency.pdf.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, (2010). *Interim Report on the Evaluation of the Growth Model Pilot Project*, Washington, D.C.,Retrieved from http://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp.pdf

Yell, M., & Drasgow, E. (2009). *No child left behind: A guide for professionals.* (2^{nd} ed.). Columbus, Ohio: Pearson Education Inc.