Describing State Level Mathematical Growth Using the Student Growth Percentile (SGP) Methodology

Darin C Kelberlau, University of Nebraska-Lincoln

Available at: https://works.bepress.com/darin_kelberlau/1/
DESCRIPTING STATE LEVEL MATHEMATICAL GROWTH USING THE
STUDENT GROWTH PERCENTILE (SGP) METHODOLOGY

by

Darin C. Kelberlau

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Psychological Studies in Education
(Quantitative, Qualitative, & Psychometric Methods)

Under the Supervision of Professor Delwyn L. Harnisch

Lincoln, Nebraska

August, 2015
The purpose of this study was to describe growth in student mathematics performance at the student and group levels as measured by the statewide mathematics test in Nebraska. Student groupings analyzed for differences in growth patterns were cohorts (elementary and middle school aged students), growth categories, grade levels, gender, ethnicity, students receiving special education (SPED) services, students receiving support services due to English not being their native language (ELL). Building differences were defined by average number of students per grade level (size), levels of student performance, levels of students receiving SPED services, and levels of students receiving ELL services.

The Student Growth Percentile (SGP) methodology was the growth model employed to quantify learning growth in mathematics as measured by the Nebraska State Accountability (NeSA) mathematics assessment. This methodology meets federal requirements for accountability, provided student level growth scores, and compared students with similar testing history. The NeSA-M results (2010-2014) were analyzed using the SGP approach for grades 3-8. Analyses were run utilizing the open source SGP Package within the R Studio software application.

The SGP assumptions were verified prior to testing for growth differences between student groups. Median SGPs matched expectations, and a variety of
perspectives ensured that performance and growth scores were uncorrelated. These approaches confirmed that no relationship existed between performance and growth. Scatterplots were generated to evaluate possible ceiling and floor effects. The plots demonstrated the possibility of experiencing all levels of growth regardless of performance.

Comparisons between groups of learners revealed significance due to the power of the study. Effect sizes were all small, which suggests little practical significance. The availability of more longitudinal data offers an opportunity for further investigation.
Acknowledgements

I want to thank many people who played key roles as I completed not only my dissertation, but also my coursework and related tasks over the years. First, I’d like to thank my adviser, committee chair, and mentor, Dr. Delwyn Harnisch. He has provided much guidance in structuring my program and many meaningful experiences in assessment and research design. Without his support and confidence in me, I would not be where I am today.

I also thank Dr. Charles Ansorge for his willingness to serve on my committee. I have great respect for his dedication to teaching, statistics and technology integration. His expertise and approach to designing online courses has greatly influenced me as a teacher.

This study would not have been possible without the work of Dr. Damian Betebenner with The National Center for the Improvement of Educational Assessment, Inc. He has made such an impact on utilizing the SGP methodology with states. His openness in sharing his expertise through replying to emails and taking my calls was of much value to me. I can’t thank him enough for assisting me with adding in Nebraska structure to the open source code. This allowed me to move forward with my study and make my analysis as similar to a real state level analysis as possible.

Another very generous educational researcher who shared her expertise in growth models was Dr. Katherine Castellano with Educational Testing Service. It was her paper with Dr. Andrew Ho on growth models that captured my interest on growth models in general, but really was my first glimpse at the Student Growth Percentile methodology.
She, too, was very willing to share with me her knowledge on growth modeling and R so that I could move forward with my research.

Krissy Johnson, SGP Data Analyst with the State of Washington (Office of Superintendent of Public Instruction (OSPI), was kind enough to share with me how SGPs were utilized in her state. She provided ground level look at growth modeling.

I also want to thank all of my colleagues and friends who would frequently ask me how my paper was progressing. Many friends, not really familiar with the research, but showing interest, encouraged me because they knew it was important to me. Many colleagues shared their dissertation experiences as well as words of support as they knew the importance of adding to the body of research in education. Many fellow workers I worked with on a regular basis deserve mentioning. Showing interest and asking me how my research was developing was extremely motivating to me and kept me focused on the end goal.

Finally, thank you to my family. I appreciate so much their willingness to let me take family time for my research. To Kate, my six-year old daughter, asking me if I had homework tonight and Hayden, my ten-year old son, asking me how my paper was coming and when will it ever be done – thank you. My wife, Allison, thank you so much for not only believing in me but also having the patience and understanding to allow me to complete this dissertation.

Finally, a number of talented caring individuals in my life have greatly influenced me and made the completion of this dissertation possible -- thank you!
TABLE OF CONTENTS

Chapter 1 – INTRODUCTION ........................................................................................................ 1
  Context ........................................................................................................................................ 1
  Purpose of Study .......................................................................................................................... 2
  Research Questions .................................................................................................................... 3
  Assumptions ................................................................................................................................. 3
  Definition of Terms .................................................................................................................... 4
  Delimitations ................................................................................................................................ 12
  Limitations .................................................................................................................................... 12
  Significance of Study .................................................................................................................. 13

Chapter II -- LITERATURE REVIEW .......................................................................................... 14
  Measuring Growth ........................................................................................................................ 14
  Defining Growth Models ............................................................................................................. 18
  Value-Added Models (VAM) ......................................................................................................... 19
  Purpose of Growth Models .......................................................................................................... 19
  Review of Growth Models ............................................................................................................ 20
    Gain-Based Models .................................................................................................................... 21
    Conditional Status Models ........................................................................................................ 21
    Multivariate Models .................................................................................................................. 21
  Limitations of Growth Models .................................................................................................... 22
  Student Growth Percentile (SGP) Methodology .......................................................................... 23
    Student Growth Trajectories or Projections ............................................................................. 29
    Benefits of SGP Methodology .................................................................................................. 31
Chapter III – METHODS ......................................................................................................................... 36

Research Design ................................................................................................................................. 36
Instrumentation ................................................................................................................................. 38
Sampling Procedures ......................................................................................................................... 41
Data Collection ............................................................................................................................... 46
IRB process ......................................................................................................................................... 49
Data Analysis ..................................................................................................................................... 49

Chapter IV – RESULTS ....................................................................................................................... 53

Introduction ........................................................................................................................................ 53
R / SGP Package Coding for Analyses ............................................................................................ 53
SGP Analysis Steps ............................................................................................................................ 53
Sampling Information and Demographic Summaries ..................................................................... 55
SGP Summary Tables on 2014 NeSA-M Analyses .......................................................................... 55
SGP Model Assumptions .................................................................................................................. 55
Research Q1 Summaries .................................................................................................................... 59
Research Q2 Summaries .................................................................................................................... 62
Gender Differences ........................................................................................................................... 62
Ethnicity Differences ........................................................................................................................ 63
Special Education Differences .......................................................................................................... 64
English Language Learner Differences ........................................................................................... 65
Building Differences ......................................................................................................................... 65
Size – Student Count ........................................................................................................................ 65
Student Performance ......................................................................................................................... 67
LIST OF FIGURES & TABLES

Figure 1.1 Catch-up, Keep-up, Move-up, and Stay-up Definitions ........................................ 5
Table 1.2 Growth Proficiency Targets ......................................................................................... 6
Table 1.3 Cohort Groupings By Grade and Year ........................................................................ 13
Figure 2.1 Sample Score Distributions for Two Consecutive Years ........................................ 25
Figure 2.2 Illustration of Calculating SGPs for Two Hypothetical Students ......................... 27
Table 2.3 Scale Scores for a Student Over Five Years On a State Reading Test ......................... 28
Table 2.4 Percentile Cut Scores for Grade 7 Reading Based Upon Grades 3, 4, 5, and 6 Scale Scores .................................................................................................................. 28
Table 2.5 Assumptions of Linear and Quantile Regression ......................................................... 33
Table 3.1 NeSA-M Cut Scores (Scale Scores) By Grade Level ................................................. 37
Table 3.2 Describing Various Levels of Student Growth .......................................................... 38
Table 3.3 NeSA-M Test Configuration ......................................................................................... 39
Table 3.4 NeSA-M Number of Items Per Content Strand Per Grade Level ............................... 40
Table 3.5 Longitudinal Distributions of Reasons Why Students Did Not Have a Valid Test Score ................................................................................................................................. 45
Table 3.6 Distribution of the Sample .......................................................................................... 46
Table 3.7 Defining the Types of Buildings Per Grouping Variable ............................................. 51
Figure 3.8 Percent of ELL Students per School Building ............................................................ 52
Table 4.1 Number of Valid Student Records By Grade Level and Year .................................... 55
Table 4.2 Number of Students Receiving a SGP Score By Grade Level for 2014 .................... 55
Table 4.3 Median SGPs By Grade Level for 2014 .................................................................... 56
Table 4.4 Correlations (r) Between Most Recent Prior Standardized Scale Score and SGPs and p-values By Grade Level for 2012-2014 .................................................. 57
Table 4.6 Growth Levels (Baseline SGP) By School Level for 2012-2014
Table 4.7 Growth Levels (Baseline SGP) By Grade Level Over Time
2012-2014
Table 4.8 Growth Levels (Baseline SGP) By Grade Level for Cohorts From
2012-2014
Table 4.9 Median Growth (Baseline SGP) By Gender for 2012-2014
Table 4.10 Pairwise Comparisons Growth of Ethnicity Groups
Table 4.11 Follow-up Test Results for Different Types of Buildings –
Size (Average Student Count)
Table 4.12 Building Median SGP By Building Size (Average Grade level)
By Year
Table 4.13 Kruskal-Wallis Tests for SGP Differences in Building Size in
2012, 2013, and 2014
Table 4.14 Pairwise Comparisons for SGP Differences for Building Size in
2012, 2013 and 2014
Table 4.15 Follow-up Test Results for Different Types of Buildings –
Performance
Table 4.16 Follow-up Test Results for Different Types of Buildings –
SPED Levels
Table 4.17 Follow-up Test Results for Different Types of Buildings –
ELL Levels
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>IRB Approval Letter .................................................................................. 90</td>
<td></td>
</tr>
<tr>
<td>Appendix B</td>
<td>Request for Data from Nebraska Department of Education ............................. 92</td>
<td></td>
</tr>
<tr>
<td>Appendix C</td>
<td>Steps for RStudio Install and R Code for SGP Package Install .................. 93</td>
<td></td>
</tr>
<tr>
<td>Appendix D</td>
<td>R Code for SGP Analysis and Report Generation ......................................... 94</td>
<td></td>
</tr>
<tr>
<td>Appendix E</td>
<td>SGP Technical Overview &amp; Documentation .................................................... 95</td>
<td></td>
</tr>
<tr>
<td>Appendix F</td>
<td>NeSA-M Table of Specifications ................................................................... 96</td>
<td></td>
</tr>
<tr>
<td>Appendix G</td>
<td>NeSA-M Student Reference Sheets .................................................................. 97</td>
<td></td>
</tr>
<tr>
<td>Appendix H</td>
<td>NeSA-M Technical Reports ............................................................................ 98</td>
<td></td>
</tr>
<tr>
<td>Appendix I</td>
<td>NeSA-M Raw Score to Scale Score Conversion Tables 2011-14 ....................... 99</td>
<td></td>
</tr>
<tr>
<td>Appendix J</td>
<td>Sample SGP Package Reports ......................................................................... 100</td>
<td></td>
</tr>
<tr>
<td>Appendix K</td>
<td>SGP Goodness-of-Fit Descriptives Reports -- Grades 4-8 ............................ 105</td>
<td></td>
</tr>
<tr>
<td>Appendix L</td>
<td>SGP Goodness-of-Fit Decile Reports -- Grades 4-8 ...................................... 108</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Context

An ongoing fundamental purpose of educational systems is to maximize each student’s skillset. Measuring student achievement provides school systems with valuable information. These data points provide local districts a way of evaluating curricular programs, instructional programs, district effectiveness, building effectiveness, classroom effectiveness, professional development needs of staff members, and individual student progress.

State departments of education are interested in measuring student growth to incorporate into their federal and state accountability systems. The Nebraska Department of Education (NDE) and State Board of Education developed an accountability system as required by state law 79-760.06, entitled the Nebraska Performance Accountability Systems (NePAS). A description of this system is found later in this study.

At the time of this study, the Nebraska State Board of Education was overseeing the development of a new accountability framework, “A Quality Education System, Today and Tomorrow” (AQuESTT), that defined the vision of the State Board. One portion focused on assessment and included a different method of measuring growth within the state accountability system. This has not yet been determined. Research by Goldschmidt, Choi, and Beaudoin (2012) compared a variety of growth models within accountability systems and came to the overarching conclusion that the model implemented does make a difference (p. 54). In classifying school buildings, the models performed differently across states (Goldschmidt et al., 2012, p.56). Even though
Goldschmidt et al. (2012) found that the various models performed fairly consistently, it is evident that there is a need to investigate how the outcomes from potential growth models might be displayed from the Nebraska mathematics test data. As Goldschmidt et al. stated “no one single model can unequivocally be assumed to provide the best results” (Goldschmidt et al, 2012, p. 54). This growth study focused on the Student Growth Percentile (SGP) methodology on the state mathematics test for the state of Nebraska.

The current study was approved the by the Institutional Review Board (IRB) at the University of Nebraska - Lincoln (UNL) which indicated appropriate precautions were taken for the safety and wellbeing of the participants (Appendix A). Nebraska state mathematics test data were obtained from the Data, Research and Evaluation Department of the NDE by complying to the Data Access and Use Policy and Procedures and completing the Data Request document (Appendix B). The form required studies to be in alignment with the Nebraska State Board of Education’s data policies, Nebraska law and protect the confidentiality of the data.

**Purpose of Study**

The purpose of this growth study was to investigate the growth by Nebraska students in mathematical achievement as measured by the statewide mathematics test. The Student Growth Percentile (SGP) analyses were applied to the state data and student, as well as group, level results were produced. Two three-year student cohort groups were formed for both elementary (grades 3-4-5) and middle school (grades 6-7-8) aged children. Aggregated results were also investigated for school buildings and other essential student groupings such as grade levels, gender, students receiving special
education services (SPED), English Language learners eligible for support (ELL), and the seven ethnicities utilized by the Nebraska Department of Education (NDE).

**Research Questions**

This study of mathematical growth addressed one main question with three additional related questions. Each of the questions utilized the Nebraska state mathematics test (NeSA-M) as the instrument to quantify student achievement. The growth model called Student Growth Percentile (SGP) was the chosen model. The focus of this study was to examine the following research questions:

- To what extent did students grow in their mathematical understandings based on the Nebraska state mathematics tests (NeSA-M)?
  - To what extent did elementary and middle school aged students grow in their mathematical understandings based on the Nebraska state mathematics test (NeSA-M) utilizing the SGP Methodology?

- To what extent can the results from the SGP Methodology be utilized to describe growth trends in various groups of Nebraska students?

**Assumptions**

The study could not assume pure randomization of the students across the state. School building assignment and classifications were used as appeared in the data file provided by the Nebraska Department of Education (NDE). The classifications followed guidelines by the state and provided to the NDE by the school districts.

The underlying statistical foundation for the SGP approach was quantile regression. These are outlined later in this study. Koenker (2005) provided a detailed explanation on quantile regression.
Definition of Terms

*Academic Peers* -- students who have the same academic history for a given assessment. These peer groupings formed the comparison groups for each student.

*Achievement* -- Ballou (2008) stated that many social scientists defined achievement in a way that it “represents the student’s mastery of the domain of the test” (p. 2). He added that much of the literature uses the term ‘ability’ in this manner and that the “knowledge and skills acquired through education” should not be confused with “innate ability” (p. 2).

*Adequate Yearly Progress (AYP)* -- under No Child Left Behind, states created annual targets in reading and mathematics in efforts to have all students reach mastery by Spring, 2014. These targets were state tests results in grades 3-8 and one determined grade level in high school. If AYP was not met, there were federal sanctions imposed upon school buildings and districts.

*Adjusted MGP (MSGP or mSGP)* -- the middle, typically median, SGP for a particular unit (i.e., classroom, building or district) based upon the academic peer groupings and English language learner status, students with disabilities status, and economic disadvantage status. The state of New York is an example that utilizes an adjusted MGP (mean) statistic for teachers and principals. Refer to the engageNY website for detailed definitions (p. 4).

*AQuESTT* -- “Accountability for a Quality Education System, Today and Tomorrow” – the framework currently being created by stakeholders and approved by the Nebraska State Board of Education will provide the foundational vision for “a next generation accountability system” (Nebraska Department of Education [NDE], 2012).
One tenet, assessment, is of most importance to this study as NDE develops a system to measure student achievement growth with the state assessments. Detailed information can be found on the NDE’s AQuESTT website (AQuESTT for Nebraska).

_Catch-up Growth_ -- one of the four Performance Level Growth Targets (Catch-Up, Keep-up, Move-up and Stay-Up) consists of students who have experienced a normal grade progression and a student growth percentile statistic in which they were at the performance level of “Below the Standards” in the previous year and earned a proficient performance level for the current year. Additionally, students in this category are those that experienced growth at or above the growth score needed to be at these performance levels within three years or by grade 11, whichever comes first. This growth projection predicts future student performance based on recent experienced growth. This concept was adapted from Betebenner’s (2011) definition in conjunction with the state of Nebraska’s testing configuration (p. 7). Figure 1.1 illustrates the definitions of this method honoring growth of students across and within achievement levels.

Figure 1.1

**Catch-up, Keep-up, Move-up, and Stay-up Definitions**

<table>
<thead>
<tr>
<th>Exceeds the Standards</th>
<th>Proficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meets the Standards</td>
<td></td>
</tr>
<tr>
<td>Below the Standards</td>
<td>Not Proficient</td>
</tr>
</tbody>
</table>

*Note: Adapted from documents from the Colorado and Washington Departments of Education*

Table 1.2 associates the growth categories to the performance levels of previous performance and needed performance.
Table 1.2

_Growth Proficiency Targets_

<table>
<thead>
<tr>
<th>Category</th>
<th>Previous Year’s Performance Level</th>
<th>Performance Level Needed in 3 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catching Up</td>
<td>Below the Standards</td>
<td>Meets the Standards</td>
</tr>
<tr>
<td>Keeping Up</td>
<td>Meets the Standards</td>
<td>Meets the Standards</td>
</tr>
<tr>
<td>Moving Up</td>
<td>Meets the Standards</td>
<td>Exceeds the Standards</td>
</tr>
<tr>
<td>Staying Up</td>
<td>Exceeds the Standards</td>
<td>Exceeds the Standards</td>
</tr>
</tbody>
</table>

_Note: Adapted from a document from Washington Department of Education_

_Economically Disadvantaged Students_ -- students whose families participate in an economic assistance program such as the Free- or Reduced Lunch Program. These determinations are established at the district level and reported to the Nebraska Department of Education (NDE).

_English Language Learners (ELL) --_ NDE defines ELL as:

“...students who have a native language other than English, OR who come from an environment where a language other than English has had a significant impact on their level of English proficiency, AND whose difficulties in speaking reading, writing, or understanding the English language may be sufficient to deny the individual (i) the ability to meet the state’s proficient level of achievement on state assessments, (ii) the ability to successfully achieve in classrooms where the language of instruction is English, or, (iii) the opportunity to participate fully in society” (2013, p. 3).

This designation is provided by the school districts to the Nebraska Department of Education.

_Elimentary and Secondary Education Act of 1965 (ESEA) --_ a federal act of Congress to provide access to high standards for all students and establishes a high level of accountability. Within this act are several programs that assist states in their focus on student poverty, students with disabilities, mobility, and struggling students. The current reauthorization of ESEA (2001), is called No Child Left Behind.
Growth -- change in student(s) achievement over time, typically from one school year to another school year. This study focused on Spring, 2011 to Spring, 2012 to Spring, 2013 to Spring, 2014.

Growth Models -- “a collection of definitions, calculations or rules that summarizes student performance over two or more time points and supports interpretations about students, their classrooms, their educators or their schools” (Castellano & Ho, 2013a, p.16).

High Growth -- for this study, a student or group of students where the SGP or mSGP exceeds 65. This threshold does vary across states but this was the most common.

Keep-up Growth -- one of the four Performance Level Growth Targets (Catch-Up, Keep-up, Move-up and Stay-Up). This category consists of students who have experienced a normal grade progression and a student growth percentile statistic in which they were at the “Meets the Standards” level in the previous year and earned a “Meets the Standards” level for the current year. Also included are those that experienced growth at or above the growth score needed to be at this level in each of the next three years or by grade 11, whichever comes first. The growth projection predicts future student performance based on recent experienced growth. This concept was adapted from Betebenner’s (2011) definition for use with the state of Nebraska’s testing configuration (p. 7). Figure 1.1 and Table 1.3 (refer to Catch-up Growth in this chapter) provide visual descriptions of the definitions.

Low Growth -- for this study, a student or group of students where the SGP or mSGP is less than 36. This threshold varies among states, but was the most common.
Median Growth Percentile (MGP or MSGP or mSGP) -- the middle, typically median, SGP for a particular unit (i.e., classroom, building or district) based upon the academic peers. This measure does not take into account individual student differences such as ethnicity, poverty, or disabilities and is sometimes referred to as an “unadjusted” median growth percentile.

Move-up Growth -- one of the four Performance Level Growth Targets (Catch-Up, Keep-up, Move-up and Stay-Up). This category consists of students who have experienced a normal grade progression and a student growth percentile statistic in which they were at the “Meets the Standards” performance level in the previous year and earned the “Exceeds the Standards” performance level for the current year. Students in this category included those that experienced growth at or above the growth score needed to be at the “Exceeds the Standards” level within the next three years or by grade 11, whichever comes first. The growth projection predicts future student performance based on recent experienced growth. This concept was adapted from Betebenner’s (2011) definition for use with the state of Nebraska’s testing configuration (p. 7). Figure 1.2 and Table 1.3 (refer to Catch-up Growth in this chapter) provide visual descriptions of the definitions.

Nebraska State Accountability -- Mathematics (NeSA-M) -- state test administered to all students in grades three through eight and eleven. The assessments were required in the amended Quality Education Accountability Act, 79-760, for the purpose of comparative accountability (SAA-14, 2014). These assessments were to measure student achievement on the Nebraska State Mathematics Standards (adopted by the Nebraska State Board of Education in 2009). NDE provided guidance to scheduling the NeSA-M
tests on two consecutive days (within the same week and preferably not Monday) or for older students, two sessions within the same day with a break in between. Special considerations were specific to each student’s Individual Education Plan (IEP) or for ELL students (SAA-14, 2014, p. 14). In 2011 and 2012, local school districts were given the option to test paper/pencil or online. Based on Moon’s (2013) comparability study of paper / pencil and online testing, policy makers in Nebraska required school districts to administer the NeSA-M tests online beginning in 2013 and for future test administrations. There were few exceptions in administering online testing for Nebraska students. NDE outlined exempt groups (listed below are those that are pertinent to this growth study):

- Students taking the alternate math test.
- Students with disabilities whose IEPs require paper/pencil testing.
- Students with 504 plans that require paper/pencil testing.
- English Language Learners whose accommodations require paper/pencil testing.
- Students contracted to institutions where online access is not allowed (SAA-14, 2014, p. 25).

Further administration and security procedures are referenced in the NDE provided document, Update: Standards, Assessment, and Accountability (SAA) -- Volume 14 (SAA-14).

*Nebraska Performance Accountability System (NePAS) --* In August 2012, the Nebraska State Board of Education adopted NePAS as the state accountability system required by state law 79-760.06. This model provided numerous average scale scores and state rankings for all state tests and graduation rates. Student achievement was categorized within three categories: status, improvement and growth. “Status” was
defined by the average scale score associated with the appropriate student grouping. “Improvement” was the difference of average scale score for two consecutive testing years. Improvement scores compare different students over time within the same grade level. NDE defines “growth” as the average of differences in scale scores for the same students over two consecutive testing seasons. Improvement and growth defined in this manner would be categorized in the simple gain growth model discussed later in this study. This growth model component is the focus for this study. A comprehensive account of all components of this state accountability system is found in the Nebraska Performance Accountability System (NePAS) document provided by the NDE (Nebraska Performance Accountability System (NePAS), 2012).

*No Child Left Behind (NCLB)* -- The No Child Left Behind Act of 2001 requires all states to develop statewide standardized tests in key areas for all kids in designated grade levels. This allows states to receive federal funding. Each state determines their own level of rigor and is charged with having all students at proficiency in reading and mathematics by Spring 2014.

*Not Proficient* -- student performance within the “Below the Standards” as measured by the Nebraska statewide test.

*Proficiency* -- student performance within the “Meets the Standards” or “Exceeds the Standards” as measured by the Nebraska statewide test.

*Scale Score* -- a transformed score that can be used to compare scores from an assessment from one year to another (refer to Appendix I).
Status -- academic achievement at one particular point in time for a student or group of students. In this study, it is measured by the NeSA-M scale score or average scale score for the appropriate group.

Student Growth Percentiles (SGP) -- describes a student’s growth from one year to the next, relative to other students with the same test history (see Academic Peers). For example, a SGP of 62 means that a student’s growth was the same or greater than 62% of similar students. The comparative group of students to which this definition refers is those tested that school year. The Office of Superintendent of Public Instruction (OSPI) in Washington refers to this as a cohort-based SGP.

Student Growth Percentiles (SGP) Baseline -- describes a student’s growth from one year to the next, relative to all students in the same grade level over previous years. This grouping creates one super-cohort. OSPI refers to this as a baseline-referenced SGP.

Stay-up Growth -- one of the four Performance Level Growth Targets (Catch-Up, Keep-up, Move-up and Stay-Up). This category consists of students who have experienced a normal grade progression and a student growth percentile statistic in which they were at the “Exceeds the Standards” performance level in the previous year and earned an “Exceeds the Standards” level for the current year or experienced growth at or above the growth score needed to be at the “Exceeds the Standards” level in each of the next three years or by grade 11, whichever comes first. Figure 1.1 and Table 1.3 (refer to Catch-up Growth in this chapter) provide visual descriptions of the definitions.

Students with Disabilities (SPED) -- students that meet the verification guidelines as outlined by Rule 51 (Rule 51, 2014). These determinations are made at the district level and reported to NDE.
Title I -- a program administered by the United States Department of Education (USDE) within the Elementary and Secondary Education Act to provide funding to schools and districts that have a high percentage of students living in poverty.

Typical Growth -- for this study, a student or group of students where the SGP or MGP falls between 36 and 65, inclusively. These boundaries vary by stated, but range is most common.

Delimitations

This study will not address how the SGP methodology might or might not be used for teacher and / or principal evaluations. Also, there are variations available to reporting SGPs at the student level and many more as data is aggregated. Unless otherwise noted, the configurations and setup were the default settings within the SGP analysis package. The researcher considered many options states currently employ for this study. The common results greatly influenced this initial analysis using the SGP methodology of measuring growth on Nebraska state math results.

Limitations

To ensure student data were not identifiable, NDE did not release poverty or district associations. Therefore, this study did not analyze potential student achievement differences between school districts or poverty levels of student groups.

Individual school districts reported all student level classifications to the NDE. There are specific guidelines for each variable but one should note they were self-reported.

The study included four years of statewide test data for grades three through eight. This allowed for two three-year cohorts of students per educational level to be used to
establish a baseline growth matrix. As the test is administered in 2015 and beyond, future students will be compared to this norm group. The SGP package creates a “SGP Baseline” statistic (refer to Student Growth Percentiles Baseline within this chapter).

Table 1.3 outlines how the cohorts are configured.

Table 1.3

<table>
<thead>
<tr>
<th>Cohort Groupings By Grade and Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

**Significance of Study**

The growth study applied the SGP methodology to the state mathematics test data for all administered years (2010-11, 2011-12, 2012-13, and 2013-14). This approach has not been applied to Nebraska’s state mathematics test data to date. Due to the fact that Goldschmidt et al. (2012) found that models provide different results from state to state, there was a need to run the model on non-simulated state test data. The results from the study can provide policy makers needed findings and reports to make an informed decision on growth analyses that fits the philosophies and needs of the state of Nebraska.
CHAPTER II
LITERATURE REVIEW

The literature review first discusses quantifying student learning growth and differentiates the difference between growth models and value-added models. The review outlines and describes various categories of growth models following Castelleno and Ho’s (2013a) classification system. General limitations of growth modeling are covered (Hull, 2007). Finally, a detailed description and key benefits of the Betebenner’s (2011) Student Growth Percentile Methodology are provided.

Measuring Growth

Many stakeholders have an interest in quantifying student achievement over time. Student growth reports provide school districts much needed information to assess program evaluation, professional development needs of staff members, teacher effectiveness, building effectiveness and overall district effectiveness. Policymakers are interested in growth models to quantify growth of students for teacher and principal evaluations and accountability systems to meet federal and state requirements.

Betebenner and Lin (2009) asserted a very fundamental perspective of measuring student growth in which the construct of interest must be defined, or as they frame it, “achievement in what?” (p. 4). “Growth is thought to occur along some continuum associated with the relevant construct.” (p. 5). As Betebenner and Lin (2009) pointed out that it is difficult to contend that a mathematical construct on a grade 8 test, perhaps containing algebraic concepts, lies on the same unidimensional continuum that a grade 3 test that focuses on basic arithmetic (p. 4).
One technique to report growth in student achievement is by changes across performance levels, or bands, from one year to the next. This is very common for state assessments as it is a requirement by NCLB. Performance bands tend to span a wide range of scale points and to demonstrate students moving from one band to the next is challenging. Normally, states have three or four performance bands for a statewide assessment. (Betebenner & Linn, 2009, p. 5). The current accountability system for the state of Nebraska reports three performance levels, “Below Standards,” “Meets Standards,” and “Exceeds Standards.” It is quite reasonable to assume students could demonstrate a great amount of growth but not detected by this method of measuring growth through performance level change.

This leads to quantifying growth utilizing the scale score metric. Even though this allows one to measure growth at a more specific manner it introduces a few other challenges. First, it tends to be additionally difficult for some stakeholders to understand, as compared to performance levels descriptions (i.e., beginning, mastering, or advanced). Secondly, to calculate growth across grade levels a vertically aligned scale is needed. This condition allows for scale scores to be subtracted arriving at a scale score gain statistic (refer to Gain-Based models later in this chapter). For very high achieving students, ceiling effects could cause negative scale score gains (Betebenner & Linn, 2009, p. 7). Gain score values depend upon a metric that possesses properties of an interval scale. These properties assume that the scale is that of equal intervals. This follows Ballou (2008) explains that a five-point difference should constitute equal gains, independent of the location of the scale (p. 1). Harris (2009) also noted that a “one-point
increase should represent that same amount of increase in learning regardless of the students’ initial level of achievement or the test year” (p. 328).

Furthermore, Betebenner and Linn (2009) conclude that even if a scale that is vertically linked and assumptions on an interval scale is acceptable it still will not make interpreting growth results more clear (p. 7). The example that was presented was that of a pediatrician’s office (detailed account appears later in this chapter). Knowing a child grew three inches in length (magnitude of growth) does not help parents in understanding the observed growth. Not until it is put in context or relative to other similar children is it understood by the parents. Parents are shown growth charts to visualize growth patterns. Ultimately, even if a perfectly designed assessment with a vertically linked scale is devised it still might not meet the desired answers from many of the stakeholders (Betebenner & Linn, 2009, p. 7).

Yen (2007) conducted a survey of various stakeholders that are closest to the students -- parents, teachers and administrators. They were informed about pros and cons of various growth measures and were asked about their areas of concern. These themes were grouped by stakeholder and summarized in the following questions:

Parents:

- Did my child make a year’s worth of progress in a year?
- Is my child growing appropriately toward meeting state standards?
- Is my child growing as much in Math as Reading?
- Did my child grow as much this year as last year?

Teachers:

- Did my students make a year’s worth of progress in a year?
- Did my students grow appropriately toward meeting state standards?
- How close are my students to becoming Proficient?
- Are there students with unusually low growth who need special attention?

Administrators:
- Did the students in our district/school make a year’s worth of progress in all content areas?
- Are our students growing appropriately toward meeting state standards?
- Does this school / program show as much growth as that one?
- Can I measure student growth even for students who do not change proficiency categories?
- Can I pool together results from different grades to draw summary conclusions?

Yen (2007) noted that most of these issues revolve around if the observed student(s) growth was “reasonable or appropriate” (p. 281).

There are three approaches to analyzing student learning data -- status, improvement and growth. Status is a single snapshot of a student or group of students at a particular point of time. Typically the focus is on grade level or at performance level (i.e., proficiency or mastery). Improvement is looking at different groups of students at a particular unit (i.e., grade level, course, etc.) over time. Districts can evaluate curricular programs by reviewing a grade level status scores over time. The desired results would be an upward trend. Growth measures improvement over time with a matched pair design. Examining data as the same students move through a district can reveal strengths or weaknesses.
Growth models can assist with evaluating particular interventions. These interventions could be professional development of staff members, a mathematics intervention program, a new mathematics curriculum program, or a new instructional model.

**Defining Growth Models**

O’Malley, Murphy Larsen McClarty, Murphy, and McBride (2011) noted that even though there are numerous methods to associate student test score over time an accurate use of a student growth model required scores that:

- can be mathematically compared from one occasion to another;
- can be connected for the same students over two or more occasions;
- show changes that indicate trait changes.

O’Malley et al. (2011), also stated that growth models “capture a student’s score changes over more than one occasion and focuses on the change itself” (p. 3).

Growth models have increased in popularity with state policymakers due to the United States Department of Education allowing the use of growth models as a way to comply with the AYP initiatives under NCLB and Race to the Top grant requirements. The models are also appearing in ways to measure teacher impact on student learning (refer to value-added models later in the chapter).

Castellano and Ho (2013a) defined growth models as “a collection of definitions, calculations, or rules that summarizes student performance over two or more time points and supports interpretations about students, their classrooms, or their schools” (p. 14). This overarching definition does capture many of the models currently employed. Others
have more simply termed growth models as a set of methods used to connect student scores over time (O’Malley, et al., 2011, p. 3).

**Value-Added Models (VAM)**

The distinction between Value-Added Models (VAM) and growth models should be noted. Auty (2008) reported that a VAM uses a variety of data (i.e., student characteristics, prior achievement, etc.) as control variables in order to isolate the impact of a teacher, program, or school on student achievement (p. 1). Implementers of VAMs intend to quantify the amount of student growth beyond that was controlled for by the model. This additional growth or “value added” is contributed to the teacher, school or district (Ligon, 2008, p. 4). Green and Fellers (2014) conceptually described the value added as the difference between observed student growth from expected growth (Green & Fellers, 2014, slide 4). Another perspective of VAM is offered by Castellano and Ho (2013a), whereas they pose the question, “What caused growth?” (p. 17). This growth metric may be used to attribute the growth experienced to teachers, principals, classrooms or buildings. Castellano and Ho (2013a) stressed that value-added is really “an average student status beyond expectations” (p. 24).

**Purpose of Growth Models**

O’Malley et al. (2011) noted three main reasons for growth models: they “conceptually align well with one of the fundamental goals of education—student learning, provide richer information on student learning than a single score at one point in time because they connect scores from multiple assessments, and focus on the educational development of individual students” (pp. 4-5). Rather than focusing on a status model like NCLB, that assesses if students have met a particular level of
proficiency, growth models measure achievement and rate of progress. This model can be used to calculate the rate of growth needed to meet a goal in a given timeframe or predict the likelihood of future growth patterns.

State Longitudinal Data System, (2012) provided a similar approach on general uses of growth models: NCLB compliance, state accountability, teacher evaluation, individual student comparison, and AYP determination (p. 6).

Growth models add another layer to the accountability systems. Not only is the number or percent of students that met expectations (status accountability model) of interest but progress towards those goals are important.

Review of Growth Models

Growth models can be categorized by a variety of criteria. This study followed the grouping schema by Castellano and Ho (2013), in which they used the model’s statistical foundation as the main determining factor. They consisted of three main classifications:

- **Gain-Based** -- based on score gains and trajectories on a vertical scale over time. Examples are gain-score, trajectory, categorical or transition / value table, and gains / slopes as outcomes.

- **Conditional Status** -- expresses scores in terms of expectations based on past scores. Examples include residual gain, projection / prediction / regression, covariate adjustment, and student growth percentile.

- **Multivariate** -- Uses entire student score histories as an outcome to associate higher-than-expected scores with particular educators. This category includes Educational Value-Added Assessment System, cross-classified and persistence models. (Castellano & Ho, 2013a, p.18).
**Gain-Based Models**

This model simply is the difference between two test scores at two different time points. To calculate the gain one would subtract the previous score from the current score. To accommodate several years of data, averaging growth over the years is common practice. The assessments used to measure student achievement should have a vertical scale (Castellano & Ho, 2013a, p. 21).

**Conditional Status Models**

These models calculated growth based on additional information. A two-step process is utilized. The model first established current score expectations based upon previous achievement. Then actual student scores are compared to this prediction (Castellano & Ho, 2013a, p. 21). It is important to note that predictions will be influenced when expectations change. For example, predictions will vary if expectations now use student demographic information or if expectations are now based upon an additional year of achievement history (Castellano & Ho, 2013a, p. 19). Student Growth Percentile methodology is one example of this model and was the focus of this student growth study.

**Multivariate Models**

This modeling procedure utilizes high-level statistical procedures to measure value-added that can be associated with teachers, schools or districts. This model is generally not appropriate for describing or predicting growth. Multivariate models often require specific software and training on interpretation of the output (Castellano & Ho, 2013a, p. 21).
Limitations of Growth Models

Hull (2007) stated that like any measurement model there are limitations of growth models. He listed the following:

- measures of achievement can be good, but none are perfect;
- there is no data for untested subjects;
- there can be missing or incomplete student data;
- experts dispute how completely “Value-Added” growth models capture teacher, program or school effectiveness;
- a single growth model does not serve all purposes;
- measuring growth in high schools is difficult.

A detailed account of each limitation was listed within his post. Hull (2007) also added that “growth models, especially value-added models, are relatively new in the education realm and their limitations are still being debated within the research community”.

O’Malley et al. (2011) noted a few limitations of growth models. Measurement error exists at each testing time. These challenges in accurately determining a student’s true score of a construct at each time point are magnified when determining growth over time. Growth models varied on the approach to handle missing student data (p. 5). The changing nature of the constructs across grade levels could lead to “imprecise longitudinal interpretations” (p. 5). Another limitation described was that most growth models assume all students progress in a similar manner. For example, a model that assumed linear growth for all students. Finally, when compared to status measures, growth models tend to be complicated for most to understand (p. 5).
**Student Growth Percentile (SGP) Methodology**

SGP methodology describes how typical a student’s growth is relative to other students with the same achievement history (academic peers). This model can provide two statistics -- a student growth percentile and a percentile growth trajectory. The prior provides that normative perspective of growth as compared to the academic peers. The latter provides the likelihood to reach a predetermined target. In most cases this is a level of “proficiency” or “grade level achievement.”

This methodology moves past a simple magnitude of growth and to a metric, percentiles, in which one can draw conclusions from. Yen (2007) summarized his survey results (listed later in this study) into a general assertion that stakeholders want to understand if the growth observed for their child was “reasonable or appropriate” (p. 281).

An example of this desire occurs daily in the medical area of pediatrics. Doctors routinely measure the height, weight and head circumference of infants. A commonly understood unit of measurement is the inch. At an annual checkup visit, a parent learns that from his/her 4-year-old son or daughter’s height increased by three inches from the previous year. Even with the unit of measurement being understood, the parents are not as interested in the magnitude of observed growth but more intrigued by the following questions: What does the growth observed (three inches) mean? How (a)typical was this experienced growth? With the use of growth charts, a pediatrician can express the growth in a normative manner. Growth charts can be found at the National Center for Chronic Disease website (http://www.cdc.gov/growthcharts). Your child’s growth of three inches
was in the 52nd percentile when compared to other similar infants (Betebenner, 2011, p. 3).

Betebenner (2011) provided a graphical explanation (Refer to Figure 2.1 below) of the SGP methodology (Betebenner, 2011, pp. 4-5). For simplicity’s purpose the illustrations below only included two consecutive scores, whereas the SGP model includes all assessment history. The bivariate distribution indicated on the left three-dimensional graph associates the prior year’s score (axis to the left) with the current score (axis to the right). The height of the surface represents the frequency of the pair of scores. The graph on the right shows the conditional distribution (indicated by the bell-shaped curve generated by slicing the plane vertically through the point of interest) for all students scoring a scale score of 600 in 2010. The specific example is a student who scored a 600 in 2010 and a 650 in 2011. In order to determine the SGP for this child one would find the percentile rank associated with a scale score of 650 on this conditional distribution. The value of 650 lies at the 75th percentile. This translates into the student grew as much or more than 75% of the students that had the same test score last year (600).
Figure 2.1. Sample Score Distributions for Two Consecutive Years. This figure displays the distribution between the prior year (2010) scale score and current year (2011) scale score.

There are multiple methods to calculate the SGPs. The simplest way is to determine the percentile rank for each observed student test score based upon the conditional distribution. This is simple to calculate; however, with more than one prior test score, the sample size of the conditional distribution is drastically reduced. The percentile corresponding to the \( n^{th} \) ranked score within the conditional distribution is calculated by:

\[
PR_n = \frac{100}{N} \left(n - \frac{1}{2}\right),
\]

where \( n \) is the rank for a current test score conditional on the prior test score and \( N \) is the total number of scores in the conditional distribution or the number of students in this academic peer group. This percentile rank is the SGP for that student (Grady, Lewis, and Gao, 2010, p. 8).

In situations where sample size might be a concern, the work of Grady et al. (2010) can be referenced. They described and implemented a “Rolling Average” technique that creates a conditional distribution of academic peers that were within one standard error of measurement of the prior year test score. This range of values was deemed the candidate
interval. This “enhanced” conditional distribution creates a larger distribution for situations where a limited set of scores are available (p. 8).

Castellano and Ho (2013a) illustrated a similar approach with the following example. The percentile rank formula for small discrete variables is:

Percentile Rank =

\[
\frac{\text{Number of Students below Score} + (0.5 \times \text{Number of Students at Score})}{\text{Number of Students in the Academic Peer Group}} \times 100
\]

For this example (Refer to Figure 2.2 below) there were six third grade students who all scored 220 on last year’s test and six third grade students who all scored 280 on last year’s test. This formed two academic peer groups (same academic test history). The SGP for the student in the first peer group that scored a 310 in fourth grade would be a 75. The percentile rank = \( \frac{4 + (0.5 \times 1)}{6} \times 100 = 75 \). This results is found by counting the number of students below the current score of 310 (4), counting the number of students that scored 310 (1) and counting the total number of students in the academic peer group (6). This student grew as well or better than 75% of his / her academic peers. The student that scored a 310 from the other peer group would have a SGP of 42. From this peer group one would count those students that scored below 310 (2), those students that scored exactly 310 (1) and total number of students in this peer group (6). This student grew as well or better than 42% of his / her academic peers.
Figure 2.2. Illustration of Calculating SGPs for Two Hypothetical Students. This figure was taken from Castellano and Ho (2013, p.94).

Betebenner (2011) describes the most common SGP calculation (the approach used for this growth study) as using quantile regression (Koenker, 2005) to “establish curvilinear functional relationships between the cohort’s prior score and the cohort’s current scores” (Betebenner, 2011, p. 5). “Quantile is a general term for dividing a distribution of scores into parts” (Slaughter, 2008, p. 8). Slaughter (2008) also noted that the ability to divide up a distribution in any desired manner allows for great levels of flexibility (p. 8). This was the methodology for this growth study.

Economists have historically used quantile regression to explore relationships involving income. This is due to relationships varying differently at either end of the income continuum. Physicians have also benefitted from quantile regressions when studying associations of the weight of patients to other variables. It is often of interest patients that are far above (obese) or below (anorexic tendencies) healthy weights.
For SGP, the quantile regression model (more in depth explanation follows) performs 100 analyses, one for each percentile, which produces a coefficient matrix that could be used as a lookup table. This chart provides the ability to enter any combination of prior assessment history to obtain the percentile cutpoints associated with a student’s current achievement. Betebenner (2011) provided a specific example that clarified this technique. For a situation with the following reading test scores (Table 2.3):

Table 2.3

<table>
<thead>
<tr>
<th>Scale Scores for a Student Over Five Years on a State Reading Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>2004</td>
</tr>
<tr>
<td>519</td>
</tr>
</tbody>
</table>

The table above was one hypothetical student’s scale scores on the state reading test in grades three (2004) through seven (2008). The current test score (dependent variable) is the student’s 7th grade test score of 601. While prior tests (independent variables) being in grades three through six (p. 6).

Using all students with the same testing history, the quantile regression analysis yields the following coefficient matrix (Table 2.4):

Table 2.4

<table>
<thead>
<tr>
<th>Percentile Cut Scores for Grade 7 Reading Based Upon Grades 3, 4, 5, and 6 Scale Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>514.8</td>
</tr>
</tbody>
</table>

In this example, the student’s current test score of 601 is greater than the 50th percentile and below the 51st percentile cutpoints, resulting in SGP of 50. Thus, the growth experienced between 6th and 7th grade for this child exceeded that of 50 percent of his/her academic peers (Betebenner, 2011, p. 6). Betebenner (2011) further explained that
the SGP percentile “reflects the likelihood of such an outcome given the student’s prior achievement” (Betebenner, 2011, p. 16). This in essence was a way to quantify the probability of attaining a particular level of student achievement.

Koenker and Bassett (1978) introduced quantile regression as an extension of ordinary least squares (OLS) model. OLS estimates the relationship between the predictor variables with the mean value of the dependent variable. Quantile regression focuses on the conditional median and allows for analysis at various locations on the conditional distribution. \( Q(\tau) \) represents the \( \tau \)-th quantile, where \( 0 < \tau < 1 \). The \( \tau \)-th quantile is where \( 100(\tau)\% \) of the distribution is equal to or below. For example, \( Q(0.5) \) is the 0.5th quantile which the location on the conditional distribution where 50% of the data are equal to or below. This particular example is of special interest due to being the conditional median.

Both Kroenker (2005) and Hao and Naiman (2007) provided in-depth explanations of quantile regression.

Betebenner (2010b) created the “SGP” package to run within the statistical software program R (R Development Core Team, 2014). Castellano and Ho (2013b) explained that by default this package estimates “100 quantile regression lines, surfaces or hypersurfaces (for 1, 2, or more than 2 prior years of data, respectively) from 0.005 to 0.995 in increments of 0.01” (p.192). Betebenner et. al. (2014) provided a detailed description of the default and possible configurations within the SGP package for using quantile regression within R.

**Student Growth Trajectories or Projections**

An extension of this one-year growth measure was a way to project out multiple years to determine the growth rate needed to meet an achievement target. Predicting
future levels of performance based upon previous history was the intent (Slaughter, 2008, p. 6). Naturally stakeholders desired to monitor growth for a very important group of students -- those that were not yet proficient. It addressed the question, “Was this student on track to be proficient?” The target (often proficiency) and time frame were established and remain fixed throughout. These projections or trajectories are calculated by using multiple coefficient matrices: Grade 4 -- Using grade 3 prior achievement. Grade 5 -- Using grade 4 and grades 3 & 4 prior achievement. And so on as tested grade levels were available (Betebenner, 2011, p. 8).

In efforts to better understand the relationship between the numerous coefficient matrices Betebenner (2011) offered a scenario of a particular student with an associated timeline. Refer to Betebenner’s (2011) technical overview for a complete explanation (pp. 8-11).

Focusing on non-proficient students is a common practice by school districts. It is also an area of focus due to accountability requirements. SGP allows for estimation of amount of time until proficient and adequate incremental growth towards proficiency. These calculations allow schools to be rewarded for AYP purposes. For example, if a student was not proficient on the statewide assessment but was making pre-determined “appropriate” growth towards the target the school would be rewarded the same as for a student that was at the proficient level (Slaughter, 2008, p. 6). This growth-to-standard approach is defined by Betebenner (2011) in which pre-determined achievement targets (performance levels for this study) and a time-frame to reach the targets are established for each student. The terms Catching-up, Keeping-up, Moving-up and Staying-up were used to assess adequate growth. These statistics allow stakeholders to determine if
incremental growth along the way were “enough”. Again, the destination and time frame to reach it remained fixed (pp. 7-8).

**Benefits of SGP Methodology**

The underlying statistical modeling of quantile regression from the SGP methodology is complex; however, the concept of a percentile is a widely known statistic in educational settings. It is a fairly new concept when associated to growth and not status. However yet, the concept remains consistent. Conceptually, creating a conditional distribution with students that possess the same prior test history is one in which is understandable by invested stakeholders (Slaughter, 2008, p. 9). Comparing growth relative to others that have the same achievement history is also an understandable approach by educational stakeholders.

Mean and median have characteristics that often make one more appropriate than the other as a measure of centralness. The mean uses each number within the data set and is influenced by outliers. The median is less sensitive to extreme values and skewed data. Similarly, the SGPs, which utilize conditional median estimation, are more robust to outliers than with traditional conditional mean estimation (Betebenner, 2011, p. 18). Kroenker (2005) explains that for the $\tau$-th quantile regression plane the influence of the data point is not dependent on the distance from the plane, but simply if it is above or below the plane (p. 44). Thus, the methodology is robust towards outliers.

The SGP methodology does not correlate the SGP statistic with prior test history. Betebenner (2011) noted that this was not the case for other multilevel models that measure student growth. He further described the models “requiring a vertical scale, fit lines with distinct slopes and intercepts to each student” (Betebenner, 2011, p. 18). The
slopes indicate an average rate of increase or growth associated with a particular student. It is also noted by Betebenner (2011) that within a normative approach (i.e., SGPs) to measuring growth, it was not appropriate to assume two students from different conditional distributions with equal SGPs experienced equal growth. The SGP methodology does not address the magnitude of growth but describes relative growth (p. 18).

Betebenner (2011) explained that it is often incorrectly understood that a vertical scale was a prerequisite to properly measure student growth across grades (p. 3). “Not only is a vertical scale not necessary, but its existence obscures concepts necessary to fully understand student growth” (p. 3). Although many statewide assessments do possess the characteristic of a vertical scale, it is not necessary to perform a growth analysis using SGP methodology (p. 16).

Koenker (2005) termed the property of “equivariance to monotone transformations” as an advantage to quantile regression. There were many scales in the assessment of student achievement. Monotone transformation was not an uncommon practice with test scales and that SGPs were not changed if this occurred (Betebenner, 2011, p. 18).

Quantile regression offered a more complete description of the impact of the predictor variables on the conditional distribution. Hao and Naiman (2007) cited that quantile regression allowed researchers to examine this impact not only on the central location of the distribution but also on the extremes (p. 3). Multiple quantiles also provided information on the shape change of the distribution (Hao & Naiman, 2007, p. 4). This more comprehensive understanding of growth patterns for all levels of student
achievement provided educational stakeholders an insight into measuring learning for all students.

Diana (2012) illustrated the differences between ordinary least squares regression and quantile regression features in the following table.

**Table 2.5**

**Assumptions of Linear and Quantile Regression**

<table>
<thead>
<tr>
<th>Linear Regression</th>
<th>Quantile Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates the mean of a response variable conditional on the values of the</td>
<td>Specifies the conditional quantile function (focus on quantiles)</td>
</tr>
<tr>
<td>explanatory variables (specifies the conditional mean function)</td>
<td></td>
</tr>
<tr>
<td>Determines the rate of change in the mean of the response variable</td>
<td>Defines functional relations between variables for all portions of a probability distribution</td>
</tr>
<tr>
<td>Provides a measure of the impact of explanatory variables on the central</td>
<td>Determines the effect of the explanatory variables on the central or non-central location, scale, and shape of the distribution of the response variable</td>
</tr>
<tr>
<td>location of the distribution of the response variable.</td>
<td></td>
</tr>
<tr>
<td>Does not account for full conditional distribution properties of the response</td>
<td>Permits the analysis of the full conditional distributional properties of the dependent variable</td>
</tr>
<tr>
<td>variable.</td>
<td></td>
</tr>
<tr>
<td>Normal distribution (sensitive to outliers)</td>
<td>Distribution-free (allows study of extreme quantiles)</td>
</tr>
<tr>
<td>Determines best fitting line for all data</td>
<td>Different estimates for different quantiles</td>
</tr>
<tr>
<td>Normal distribution of errors</td>
<td>No assumption about the distribution of errors</td>
</tr>
<tr>
<td>Assumption of constant variance in errors (homoscedasticity)</td>
<td>Does not assume homoscedasticity</td>
</tr>
</tbody>
</table>

Slaughter (2008) suggested the appropriateness of utilizing SGP, or quantile regression. In this situation, student achievement test scores were often not normally distributed and the variability at the either end of the distribution were higher than that of the center (p. 9).
SGP provided growth statistics at the student level. Commonly growth models only report growth at some group level. From a student, family and district perspective, this aspect is a significant advantage when the intent was to maximize each child’s potential. Currently, only Colorado provided this level of data to families.

Haertel (2009) felt that the use of multiple years of assessment data, not just one previous year, resulted in a more informative grouping of students (p. 6). Using quantile regression within the SGP package allowed for all data to be utilized. Often, if SGPs are calculated empirically, sample size will become an issue. This is true especially if the desire is to use multiple years of prior achievement history. In order to properly arrive at the SGP, there must be an adequate sample from which to draw, and each combination of score histories must be evaluated.

Access to software packages was not an issue. The R language was available on Windows and Mac computers (R Development Core Team, 2014). R was also available as an add-on in SPSS. Betebenner et al. (2014) have created a package, SGP, to be run within R to provide the needed SGP output. More generally, quantile regression can be performed on other common statistical packages such as STATA and SAS.

Lastly, this methodology has many advantages. Perhaps most importantly, it was an approach that met many current educational initiatives for school districts, those that can and cannot be controlled. One might suggest this was a very efficient methodology.

The review of the literature on measuring student learning growth indicated a variety of factors policy makers must consider as an accountability model is chosen to be implemented. The foundational assumptions and several advantages of the SGP methodology provided a reasonable theoretical approach for the state of Nebraska’s
mathematics test. It is important to note that there is not one model that solves all of the questions that various policy makers have. The approach for this study was based upon the research outlined in this chapter and the work of many state agencies utilizing a similar growth model.
CHAPTER III

METHODS

The structure of this study is outlined in the following section. The characteristics of the instrument used are described by grade level, number of items, content assessed, and other factors associated with the assessing of students. The data collection approach is also explained. A detailed description of the use of the R software and SGP package is outlined. The design of the study was configured to take into account the current assessment structure in place for Nebraska, utilize the benefits of the SGP methodology and it’s underlying statistical foundation - quantile regression, benefit from other states implementing a similar approach, assist in informing stakeholders on their questions or particular interest in student growth, and ultimately to address the research questions of this study.

Research Design

This study applied a normative growth methodology (SGP) to the Nebraska statewide mathematics test (NeSA-M) in grades 3-8. Juniors were omitted from this study due to the fact that two consecutive years of tests results were needed to measure growth (9th and 10th graders were not administered the NeSA-M in Nebraska). Students were tested in their typical educational setting. Randomization was neither possible nor appropriate. The school assignment was maintained for a more realistic analysis of aggregated results at a school building level.

SGPs are estimated using the statistical application R (R Development Core Team, 2014). Specifically, Betebenner’s (2014) SGP package, within the R software, was
used to compute SGPs for each appropriate student. Default settings were implemented unless otherwise noted in this study.

State specific information was provided to Dr. Damian Betebenner, an associate at The National Center for the Improvement of Educational Assessment, to be included in the SGPstateData object so that a full analysis could be ran specific to the state of Nebraska, thus replicating as close to a real situation that might be implemented by the policymakers within Nebraska. During the analyses, the package referenced the look-up files to customize the reports appropriately by state. This information can be found within the folder, Variable_Name_Lookup, on the SGPstateData page of the Center for Assessment website. The following performance levels or achievement levels (as the SGP package refers to them as) were provided: Below the Standards, Meets the Standards and Exceeds the Standards. The cut scores (scale score metric) are provided as Table 3.1 illustrates below.

Table 3.1

<table>
<thead>
<tr>
<th>NeSA-M Cut Scores (Scale Scores) By Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

The performance levels of “Meets the Standards” and “Exceeds the Standards” were also grouped at a higher level and termed, “Proficient.” “Below the Standards” level was also mapped to a proficiency level and named, “Not Proficient”. These naming
conventions allowed for a variety of reports to be generated and match the Nebraska verbiage.

Levels of growth were entitled, “Low Growth,” “Typical Growth” and “High Growth.” This wording was the default setting and matched many of the current state agencies utilizing the SGP methodology. The following Table 3.2 displays the SGP ranges in percentiles that correspond to each growth descriptor.

Table 3.2

<table>
<thead>
<tr>
<th>SGP Range</th>
<th>Low (1st – 34th)</th>
<th>Typical (35th – 65th)</th>
<th>High (66th – 99th)</th>
</tr>
</thead>
</table>

Finally, a variety of Nebraska assessment information was provided for information to appear on the numerous reports. This information included the state test name, NDE contact information, test abbreviation, content areas, grades tested, test season and test vendor.

**Instrumentation**

To ensure a level of quality for the NeSA-M tests, items were field-tested prior being an operational item. Data Recognition Corporation (DRC) produced the technical report each year and noted the psychometric targets for the items:

- item mean value (p-value) -- a mean of approximately 0.65 while maintaining each item p-value falling within the range of 0.30 to 0.90,
- point-biserial correlations -- greater than 0.25, based on previous field tested results (Nebraska State Accountability (NeSA) Reading, Mathematics, and Science Technical Report, 2014, p. 20).
Each grade level has a Table of Specifications (TOS), or test blueprint, that defined the number of items by standard and by depth of knowledge. The TOSs for each grade level can be found within Appendix F of the NeSA technical report (Nebraska State Accountability (NeSA) Reading, Mathematics, and Science Technical Report Appendices, 2014, pp. 95-118). The number of multiple choice items per test varied by grade level. Table 3.3 lists the length of each grade level test.

Table 3.3

<table>
<thead>
<tr>
<th>Grade</th>
<th>Total Number of Items</th>
<th>Number of Field Tested Items</th>
<th>Total Items on NeSA-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>50</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>10</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>10</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>58</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>10</td>
<td>70</td>
</tr>
</tbody>
</table>

In a broader sense, the NeSA-M assessment measured four mathematical content strands at all tested grade levels: number sense, geometric / measurement, algebraic, and data analysis / probability. However, the focus, or weighting, did differ at the various grade levels. Table 3.4 lists the number of items aligned to each content strand by grade level by year of administration.
Table 3.4

**NeSA-M Number of Items Per Content Strand Per Grade Level**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Content Strand</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Number Sense</td>
<td>24</td>
<td>22</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Number Sense</td>
<td>26</td>
<td>21</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>9</td>
<td>11</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Number Sense</td>
<td>25</td>
<td>22</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>10</td>
<td>13</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Number Sense</td>
<td>20</td>
<td>20</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>Number Sense</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>14</td>
<td>11</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>15</td>
<td>17</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>Number Sense</td>
<td>16</td>
<td>20</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Geometric / Measurement</td>
<td>15</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Algebraic</td>
<td>18</td>
<td>16</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Data Analysis / Probability</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

Students were not allowed to use calculators, but in grades 4-8, were provided a reference sheet (refer to Appendix G) of commonly used grade level specific formulae and conversions. Each item had only one correct answer and was scored correct or incorrect. Each question had a value of one point and the total raw score was the sum of all correct answers for a student. Annually, experts create a raw score to scale score conversion chart. These charts were included in Appendix I. For this study, the scale score was the measure used to quantify a student’s mathematical achievement. Also, the
2014 technical report for the NeSA-M test can be referenced in Appendix H. Years 2011-2013 can be accessed on assessment division of the NDE website (NDE, n.d.).

The dependent variable for this study was the NeSA-M scale score for the most current test administration (2014). The independent variables are the previous scale scores within previous tested years (2011-2013).

**Sampling Procedures**

A data request was submitted to the NDE requesting the following data:

- Unique student ID number -- in order to associate growth over the years of the study to the appropriate student.
- Gender -- grouping variable to monitor any gaps between gender groups.
- Ethnicity -- grouping variable to monitor any gaps between the various ethnic groups.
- English Language Learner -- grouping variable for those students qualifying for support services as they learn a second language.
- Student Individual Education Plan (SPED) -- grouping variables for those students qualifying for support services from the special education department.
- Assessment Year -- school years associated to the testing season. Years included in this study were 2011, 2012, 2013 and 2014.
- Grade Level -- the grade level of the student at time of testing. Grades requested were 3rd - 8th grades. 11th grade was omitted due to the fact the scope of this study was elementary and middle school aged students.
● NeSA-M Scale Score -- the scale score for each student. This was the main metric to measure growth. The scale score metric for all grade level assessments were from 0 to 200.

● Student Performance Level -- NeSA-M had three performance levels: Below the Standards, Meets the Standards and Exceeds the Standards. The SGP package referred to the performance level as the “achievement level.”

● School ID # -- a grouping variable that allowed student level SGPs to be aggregated to measure the growth of a building.

● District ID # -- a grouping variable that allowed student level SGPs to be aggregated to measure the growth of a school district. Due to student identifiable data with the combination of variables, this data element was not provided by NDE.

● Economically Disadvantaged Students -- a grouping variable that allowed for comparisons of those students of poverty to non-poverty students. Due to data security this data element was not provided by NDE.

The request shared the scope of this growth study, a brief background on the SGP methodology, and the beneficial impact on such a study to a variety of stakeholders (refer to Appendix E).

Nearly all public school students in Nebraska in grades 3-8 and 11 were expected to complete the NeSA-M test. The exception was those students who qualified for the alternate statewide math assessment (NeSA-AAM). In order to participate in the alternate assessment, local school districts had to ensure that students met all of the guidelines that
the Nebraska Department of Education has outlined on the Alternate Assessment Participation Criteria (2014):

1) The student has:
   A) A severe cognitive disability, and
   B) Significant deficits in
      a. communication / language, and
      b. adaptive behavior

2) The student requires a highly specialized educational program that facilitates the acquisition, application and transfer of skills across natural environments (home, school, community and / or workplace).

3) The student requires educational support systems, such as:
   A) assistive technology,
   B) personal care issues, and / or
   C) health / medical services (p. 1).

SGPs were calculated for all students who had completed a minimum of two consecutive administrations of the NeSA-M test. Due to this methodology needing two data points, SGPs were reported beginning in 4th grade and continue through 8th grade. Students that skipped or repeated a grade level during tested grade levels were included in the study but a SGP will not be provided. Students in 3rd grade were included as a beginning data point. In addition to these cases were situations where a valid test score was not obtained. A detailed account of these situations, as described by NDE, in which the student was not tested:

- By the district
Because of an emergency medical waiver

Due to invalidation of assessment

Due to relocation from the district / school after data submission but prior to state testing

For reasons not covered by other descriptions (i.e., Nebraska student attending another country as a foreign exchange student.)

Removed from the district data file

Due to a formal request from parent or guardian

Due to the student met the recently arrived classification

Due to student being absent from school throughout the entire testing window (SAA-14, 2014, pp. 23-24).

Finally, students who had a scale score of zero were also excluded from this study.

Beyond these exceptions, any student who had at least two valid consecutive data points (at least one prior score history) received a SGP calculation. Additionally, all students with a valid scale score were included in the study for additional information for the model. Table 3.5 below displays the distributions of the reasons why students did not have a valid test score.
Table 3.5

**Longitudinal Distributions of Reasons Why Students Did Not Have a Valid Test Score**

<table>
<thead>
<tr>
<th>Year / Grade</th>
<th>EMW</th>
<th>INV</th>
<th>NLE</th>
<th>NT</th>
<th>OTH</th>
<th>PAR</th>
<th>SAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>10</td>
<td>302</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>10</td>
<td>361</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>9</td>
<td>307</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>11</td>
<td>283</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>8</td>
<td>302</td>
<td>7</td>
<td>1</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>15</td>
<td>298</td>
<td>7</td>
<td>3</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>7</td>
<td>393</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>10</td>
<td>348</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>7</td>
<td>298</td>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>15</td>
<td>320</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>8</td>
<td>305</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>9</td>
<td>321</td>
<td></td>
<td>1</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>7</td>
<td>377</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>8</td>
<td>350</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>8</td>
<td>337</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>5</td>
<td>285</td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>9</td>
<td>326</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>7</td>
<td>303</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2013-14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>4</td>
<td>457</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>7</td>
<td>365</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>8</td>
<td>318</td>
<td></td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>8</td>
<td>316</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>7</td>
<td>319</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>17</td>
<td>312</td>
<td>7</td>
<td>3</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: EMW = emergency medical waiver; INV = invalidation of assessment; NLE = no longer enrolled; NT = not tested; OTH = other; PAR = parental refusal; SAE = student absent.

Table 3.6 indicates the number of records from the raw data file and the number of records that was used for the analyses.
Table 3.6

Distribution of the Sample

<table>
<thead>
<tr>
<th>Year / Grade</th>
<th>Number of Students</th>
<th>Records Excluded</th>
<th>Records Used in Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11</td>
<td>130,637</td>
<td>1,989</td>
<td>128,648</td>
</tr>
<tr>
<td>3rd</td>
<td>22,500</td>
<td>320</td>
<td>22,180</td>
</tr>
<tr>
<td>4th</td>
<td>22,263</td>
<td>376</td>
<td>21,887</td>
</tr>
<tr>
<td>5th</td>
<td>21,990</td>
<td>324</td>
<td>21,666</td>
</tr>
<tr>
<td>6th</td>
<td>21,418</td>
<td>299</td>
<td>21,119</td>
</tr>
<tr>
<td>7th</td>
<td>21,319</td>
<td>332</td>
<td>20,987</td>
</tr>
<tr>
<td>8th</td>
<td>21,147</td>
<td>338</td>
<td>20,809</td>
</tr>
<tr>
<td>2011-12</td>
<td>132,998</td>
<td>2,110</td>
<td>130,888</td>
</tr>
<tr>
<td>3rd</td>
<td>22,890</td>
<td>408</td>
<td>22,482</td>
</tr>
<tr>
<td>4th</td>
<td>22,672</td>
<td>368</td>
<td>22,304</td>
</tr>
<tr>
<td>5th</td>
<td>22,304</td>
<td>313</td>
<td>21,991</td>
</tr>
<tr>
<td>6th</td>
<td>22,070</td>
<td>344</td>
<td>21,726</td>
</tr>
<tr>
<td>7th</td>
<td>21,610</td>
<td>328</td>
<td>21,282</td>
</tr>
<tr>
<td>8th</td>
<td>21,452</td>
<td>349</td>
<td>21,103</td>
</tr>
<tr>
<td>2012-13</td>
<td>135,484</td>
<td>2,084</td>
<td>133,400</td>
</tr>
<tr>
<td>3rd</td>
<td>23,460</td>
<td>395</td>
<td>23065</td>
</tr>
<tr>
<td>4th</td>
<td>22,929</td>
<td>361</td>
<td>22568</td>
</tr>
<tr>
<td>5th</td>
<td>22,716</td>
<td>353</td>
<td>22363</td>
</tr>
<tr>
<td>6th</td>
<td>22,342</td>
<td>295</td>
<td>22047</td>
</tr>
<tr>
<td>7th</td>
<td>22,275</td>
<td>352</td>
<td>21923</td>
</tr>
<tr>
<td>8th</td>
<td>21,762</td>
<td>328</td>
<td>21434</td>
</tr>
<tr>
<td>2013-14</td>
<td>137,763</td>
<td>2,211</td>
<td>13,552</td>
</tr>
<tr>
<td>3rd</td>
<td>23,360</td>
<td>468</td>
<td>22,892</td>
</tr>
<tr>
<td>4th</td>
<td>23,543</td>
<td>383</td>
<td>23,160</td>
</tr>
<tr>
<td>5th</td>
<td>23,010</td>
<td>332</td>
<td>22,678</td>
</tr>
<tr>
<td>6th</td>
<td>22,805</td>
<td>332</td>
<td>22,463</td>
</tr>
<tr>
<td>7th</td>
<td>22,569</td>
<td>342</td>
<td>22,227</td>
</tr>
<tr>
<td>8th</td>
<td>22,476</td>
<td>354</td>
<td>22,122</td>
</tr>
<tr>
<td>TOTAL</td>
<td>536,882</td>
<td>8,394</td>
<td>528,488</td>
</tr>
</tbody>
</table>

Data Collection

The raw data file obtained from NDE through a secure, password protected, ftp transfer site was reshaped to meet the requirements of the SGP package within the R software and also to address the focus of this study. The data were assembled in a long format, indicating that each row would consist of a single student and a single test result,
as opposed to a wide format consisting of a single student with all historic assessment results in a single row. When the raw data file were assembled, special consideration was given to the case sensitivity of the data and to match the expectations of the Nebraska test info (explained later in this section). The required variables for the data file were student ID, content area, school year, grade level, scale score and valid case. The student ID was a unique generated number by NDE so that student performance can be tracked but still maintain privacy of the student data. For this study, only the mathematics content area was considered. The school years obtained were 2010-11, 2011-12, 2012-13 and 2013-14. Grade levels included in the study were grades three through eight. The scale score was the assigned test score by NDE, which ranged from a 0 to 200 scale score. Those records with a zero were excluded from the study. Finally, the valid case variable indicated which records were to be included in the analyses. Any student without a valid test score or performance level, not tested or a scale score of zero were deemed, “INVALID_CASE” and left out of the study.

Optional variables were also included to allow the reports to be ran as completely as possible and mirror a fully implemented state level growth system within an accountability system. Achievement level, or Nebraska’s term of “performance level,” was added as Below the Standards, Meets the Standards, and Exceeds the Standards. A first name variable was included and listed as “first” for all records. In order to create an unique and consistent variable, the last name variable was populated by the student ID information. School number was included and was a number generated by NDE so that appropriate aggregate reports could be generated. School name was left blank for the data file. District number and name were not provided by NDE due to data privacy concerns.
In order to generate reports the district number was set to a value of 1 and District Name to “NE.” Another demographic variable not able to be obtained from NDE was the variable of student poverty. SocioEconomic Status (SES) was the variable often used to identify poverty within the educational world but not for this study. Gifted and talented variable is available by the SGP package but not available for this study.

The state enrollment status variable indicated which students were continuously enrolled in the state and should be included in state summary statistics. The state, district and school enrollment status variables allowed for a variety of reports and summary statistics. The format needed for school enrollment status was “Enrolled School: No” or “Enrolled School: Yes.” The other similar variables followed the same pattern. Another grouping variable that was populated was the level of the building associated to each student: elementary, middle or high. The grade configurations followed the state of Nebraska’s definition of each for tested grades: elementary (grades 3-5), middle school (grades 6-8) and high (grade 11). This study only focused on elementary and middle.

Finally, three other demographic variables were included: ethnicity categories and if the student qualified for special education or English language learner services. The ethnicity categories were state recognized: American Indian / Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Two or More Races and Hispanic. For a detailed explanation of all aspects of the data preparation phase, one should consult the SGP Data Preparation page within the SGP Package Wiki (Betebenner, 2014).
IRB process

An official request to move forward with this study was submitted to the University of Nebraska - Lincoln Institutional Review Board. This study received IRB approval on October 21, 2013. This ensured that the subjects of this study were not placed at undue risk and their rights and welfare were protected. The complete document can be found in Appendix A.

Data Analysis

After the data were properly prepared, the calculations or analyses steps were to take place. The SGP package was developed for the user to perform all six steps in one through the abcSGP wrapper function. If one were to add an additional year of data the updateSGP wrapper function could be utilized. However, to assist in troubleshooting and since this was an initial analysis, the researcher performed the analyses through six major phases or steps as outlined on Betebenner’s (2014) SGP Data Analysis page of the SGP Package Wiki.

The first function, prepareSGP, prepared the properly formatted data to assure the data file had the appropriate variables in the correct format. This step converted variables that were not in the correct format for upcoming steps (i.e., student ID numbers should be in a character format). The function created a Nebraska_SGP object to be used in future steps. This step also provided warnings and error messages if issues existed in the data file outside of formatting.

Once prepared, the object generated was incorporated into the next function, analyzeSGP. This function ran analyses for all years and content areas provided in the data file. Betebenner (2014) listed the following output at this point within the analysis:
student growth percentiles, baseline student growth percentiles, student growth projections, baseline student growth projections, lagged student growth projections and lagged baselines student growth projections.

The third step, combineSGP, merged the results from the prior analysis within the master record, Nebraska_SGP@Data, and calculated scale scores that were associated with the SGP targets. In order to have projected target scores generated (not the default setting) the argument, sgp.target.scale.scores, was set to TRUE (Betebenner, 2014).

The summarizeSGP function followed and was one in which summary tables were created. These results were stored into the Nebraska_SGP@Summary slot (Betebenner, 2014).

Next, visualizeSGP, created a variety of student and group level graphical displays. Betebenner (2014) listed the outputted visualizations as summary bubble plots, growth achievement plots (wave charts), and student growth plots. Betebenner (2014, December) provided all possible features to be included in all steps of the analyses, but most importantly for this study the possible arguments for creating customized visualizations (pp. 69-76).

The last function, outputSGP, was used to export student data and SGP results. Again, there was a method to accomplish all six steps into a single step. The abcSGP or updateSGP function sequentially performs all six functions into a single executable call.

For a detailed account of this single process refer to the Betebenner’s (2014) SGP Package Wiki.
Characteristics of a building were defined to analyze possible growth differences in types of buildings. Table 3.7 displays the cut points to create the four levels of a building. The descriptors are High, Mid-High, Mid-Low and Low.

Table 3.7

<table>
<thead>
<tr>
<th>Levels</th>
<th>Average Grade Size</th>
<th>Proficiency %</th>
<th>SPED %</th>
<th>ELL %</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>$\bar{x} \geq 63.0$</td>
<td>$84.3 \leq p \leq 100$</td>
<td>$20.6 \leq p \leq 100$</td>
<td>$9 \leq p \leq 100$</td>
</tr>
<tr>
<td>Mid-High</td>
<td>$37.3 \leq \bar{x} &lt; 63.0$</td>
<td>$75.6 \leq p &lt; 84.3$</td>
<td>$16.4 \leq p &lt; 20.6$</td>
<td>$4 \leq p &lt; 9$</td>
</tr>
<tr>
<td>Mid-Low</td>
<td>$18.0 \leq \bar{x} &lt; 37.3$</td>
<td>$65.9 \leq p &lt; 75.6$</td>
<td>$12.1 \leq p &lt; 16.4$</td>
<td>$0 &lt; p &lt; 4$</td>
</tr>
<tr>
<td>Low</td>
<td>$\bar{x} &lt; 18.0$</td>
<td>$0 &lt; p &lt; 65.9$</td>
<td>$0 \leq p &lt; 12.1$</td>
<td>$p = 0$</td>
</tr>
</tbody>
</table>

In order to compare growth based on building size, the average number of students per grade level was used to classify the buildings. The cut points were corresponded to the quartiles of the data. Student achievement levels across buildings were defined by the percent of students “meeting” or “exceeding” the standards on the NeSA-M test. Quartile cut points were calculated to be used to create the performance bands. For students with learning disabilities the percent of students in a building that were identified to qualify for services. These cut points were also the quartiles of the distribution. Finally, English language learners were the percent of students that qualified to receive services as well. However, due to the spread of the distribution (illustrated by a histogram and boxplot in Figure 3.8) it was determined that the most appropriate groupings for the buildings were those without an ELL program (0%) to those that have low, moderate, and high percentages of students that are learning English as a second language.
This study applied a growth model that has not yet been applied to the state mathematics test results for Nebraska. The purpose was to describe student learning in mathematics in a different manner, investigate any differences in growth patterns at an aggregated level, and examine the model fit properties for the NeSA-M tests. This initial first set of analyses and reports of mathematical growth in Nebraska provides state policy makers a glimpse at what this model might provide and it could be incorporated into a state accountability system.
CHAPTER IV

RESULTS

Introduction

Quantile regression analysis provides a more complete account of the impact of the independent variables (previous statewide mathematics test scores) on the dependent variable (most recent statewide mathematics test score). The main advantage is that these results describe the impact on a variety of locations related to the conditional distribution. For this study, the analysis describes the current test score based upon the previous test scores at each point on the continuum -- one quantile that corresponds to each percentile. The results associate students with other similar performing students.

R / SGP Package Coding for Analyses

In order to run the quantile regression analysis on the state data, the SGP package (Betebenner, Van Iwaarden, Domingue, & Shang, 2014) was selected. The SGP Package runs within the R statistical software. Refer to Appendix C for specific directions on where to obtain the free software (R Studio was used for this study on a Mac laptop) and Appendix D for the location to download and code to install the SGP Package within R.

SGP Analysis Steps

For this study, the analysis was carried out in six steps. Prior to the analysis, the raw data file received from NDE were cleansed based upon the business rules as described earlier in this paper. The data input file was also configured to meet the needs of the SGP Package code and saved at as a text delimited file (.txt). This file was read into R and saved as the R object “NeSA”:

NeSA <- read.delim("~/Desktop/NeSA.txt").
This allowed for further analysis steps to be implemented. The first step was to prepare the data for analysis associated with Nebraska information and was carried out through the following code:

```
Nebraska_SGP <- prepareSGP(NeSA, state="NE").
```

The previous step also converted any variables to the needed format for further use and saved the results into the object, “Nebraska_SGP.” The next major step was to begin analyzing the data. Code used for this was:

```
Nebraska_SGP <- analyzeSGP(Nebraska_SGP).
```

The analyzeSGP function calculated the variety of matrices, created the numerous grade progressions, calculated student growth percentiles, and student growth projections. The next step was to link the results from previous analyses steps using the combineSGP function. The code needed was:

```
Nebraska_SGP <- combineSGP(Nebraska_SGP).
```

CombineSGP function created those needed associations and saved the results to the Nebraska_SGP R object. In order to create summary statistics the next function, summarizeSGP was coded appropriately as:

```
Nebraska_SGP <- summarizeSGP(Nebraska_SGP).
```

Also, the output of the summary scores was saved to be use in aggregated charts and reports. To achieve this it was saved as a file using the following code:

```
save(Nebraska_SGP, file="Nebraska_SGP.Rdata").
```

The final step was to output a variety of visualizations and files. It was accomplished by utilizing the following code:

```
outputSGP(Nebraska_SGP).
```
Two key files for this study generated were a wide and long formatted output files. These files can be used to further summarize the results from the analyses.

**Sampling Information and Demographic Summaries**

All valid student scores were used as information to generate the appropriate analyses. Table 4.1 shows the number of valid student records by year available for SGP analyses after applying the business rules discussed in the previous chapter.

Table 4.1

<table>
<thead>
<tr>
<th>Year</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>22,180</td>
<td>21,887</td>
<td>21,666</td>
<td>21,119</td>
<td>20,987</td>
<td>20,809</td>
</tr>
<tr>
<td>2012</td>
<td>22,482</td>
<td>22,304</td>
<td>21,991</td>
<td>21,726</td>
<td>21,282</td>
<td>21,103</td>
</tr>
<tr>
<td>2013</td>
<td>23,065</td>
<td>22,568</td>
<td>22,363</td>
<td>22,047</td>
<td>21,923</td>
<td>21,434</td>
</tr>
<tr>
<td>2014</td>
<td>22,892</td>
<td>23,160</td>
<td>22,678</td>
<td>22,473</td>
<td>22,227</td>
<td>22,122</td>
</tr>
</tbody>
</table>

**SGP Summary Tables on 2014 NeSA-M Analyses**

All student scores from 2011 - 2014 were used to generate a SGP statistic for all students that had at least one prior test score. Table 4.2 illustrates the number of students by grade level that received a SGP score for 2014.

Table 4.2

<table>
<thead>
<tr>
<th>Grades</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid SGP Score</td>
<td>0</td>
<td>22,146</td>
<td>21,734</td>
<td>21,440</td>
<td>21,220</td>
<td>21,131</td>
<td>107,671</td>
</tr>
</tbody>
</table>

**SGP Model Assumptions**

SGP scores are normed on the most current test result based upon prior test history. The median is used to represent the center of the distribution. For the state data set, a score of 50 is expected. This indicates that 50% of the distribution falls below the
score and 50% lies above the score. To display the normative nature of this distribution, lists the median SGP for each grade level for the 2014 administration of NeSA-M. Deviations from 50 indicate a model that does not fit and can result from a variety of reasons. In third grade, SGPs are not calculated (indicated by “*”) and all other grades are very close to or exactly 50.

Table 4.3

<table>
<thead>
<tr>
<th>Median SGPs By Grade Level for 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grades</td>
</tr>
<tr>
<td>Median SGP</td>
</tr>
</tbody>
</table>

Goodness-of-fit reports were run to further examine how well the model performed with the Nebraska data. The first series of charts inspected were the charts that listed the percent of students per achievement level on the previous year’s assessment and the median SGP per level. The SGP should not be associated to status – current year or prior year. The goodness-of-fit report examines this at a deeper level. Not simply at the grade level but at each performance level within the grade level. Refer to Appendix K for the median SGP at each performance level for each grade level. All medians were extremely close to the expected value of 50. Within grade 4 the “exceeds” level had a slightly lower median SGP.

Another study on the relationship between SGP and student status was performed. For the 2014 year, the SGPs were divided into ten categories (1-9, 10-19, …90-99) - deciles. The prior scale scores were divided similarly, by deciles. The scale scores
associated to these bands were [1, 63), [63, 79), …[155, 200]. With an infinite population of test takers and perfect model fit one would expect 10% of SGPs to be in each scale score band (Betebenner & Van Iwaarden, 2014, January, p. 3). Values that deviate from ten are shaded with varying dark levels of red and blue. The further above a cell value of ten was the darker the red shade. Similarly, the further below a cell value of ten was the darker the blue shade. Cells across grade levels were very close to ten. There was one cell in grade 4 that did have a dark blue cell. The 9th decile for prior scale score [141, 155) and SGP range of 70-79. These results further indicated a model that fit very well with the Nebraska data.

Model fit was also analyzed at the student level. Prior student achievement and growth should be uncorrelated. The most recent scale score (standardized) was correlated with the SGP at each grade level. Table 4.4 illustrates the respective relationships.

Table 4.4

<table>
<thead>
<tr>
<th>Year</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>-.013</td>
<td>.066</td>
<td>-.022*</td>
<td>.001</td>
<td>-.009</td>
</tr>
<tr>
<td>2013</td>
<td>.008</td>
<td>.268</td>
<td>-.004</td>
<td>.574</td>
<td>-.009</td>
</tr>
<tr>
<td>2014</td>
<td>-.012</td>
<td>.085</td>
<td>.006</td>
<td>.396</td>
<td>-.004</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level (2-tailed).

4th and 7th grades were the only two levels in which the correlations were significantly different from zero. They occurred in the year in which there was only one year of prior test history.

1 A square bracket [ ] is used to include the end value and a parenthesis ( ) is used to not include the end value.
Growth models often have a difficult time expressing very high or very low growth, ceiling or floor effects. Figure 4.5 plots students’ SGPs for the most recent year against their prior achievement (scale score) for each grade level. These plots demonstrate the ability of the model to allow for all levels of growth regardless of achievement level.

Figure 4.5
Growth (SGP) and Prior Achievement (Scale Score) – Grades 4-8
Research Q1 Summaries

To what extent did students grow in their mathematical understandings based on the Nebraska state mathematics tests (NeSA-M)?

- To what extent did elementary and middle school aged students grow in their mathematical understandings based on the Nebraska state mathematics test (NeSA-M) utilizing the SGP Methodology?

Growth can be described in many ways. Table 4.6 shows varying levels (low, typical and high) of growth by the level of buildings, elementary or middle, over the years of 2012-2014. Due to the norming process for the baseline SGP statistic it is not surprising that the levels of school structure do not vary when all years of growth are aggregated. The Typical growth range has not varied across or within school levels across the years. The middle school level has experiences that reflect a slight decrease in High growth from 2012-2014, while the Low growth level has increased over time at this school level. At the elementary level, the Low growth level has increased, then decreased, while the High growth has experienced the opposite pattern -- decreased, then increased.

Table 4.6

| Growth Levels (Baseline SGP) by School Level for 2012-2014 |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Low 1-34 Percentile | Typical 35-65 Percentile | High 66-99 Percentile |
|                 | Frequency | Percent | Frequency | Percent | Frequency | Percent |
| Elementary      |           |         |           |         |           |         |
| 2012            | 14,401    | 34%     | 12,750    | 30%     | 15,091    | 36%     |
| 2013            | 15,482    | 36%     | 13,156    | 31%     | 14,334    | 33%     |
| 2014            | 14,501    | 33%     | 13,745    | 31%     | 15,634    | 36%     |
| Middle          | 63,388    | 34%     | 58,060    | 31%     | 65,830    | 35%     |
| 2012            | 19,810    | 32%     | 18,811    | 31%     | 22,462    | 37%     |
| 2013            | 21,975    | 35%     | 19,057    | 31%     | 21,372    | 34%     |
| 2014            | 21,603    | 34%     | 20,192    | 32%     | 21,996    | 34%     |
| Total           | 107,772   | 34%     | 97,711    | 31%     | 110,889   | 35%     |
To investigate this growth pattern at a more granular level the following Table 4.7 displays the percent of students by growth levels by grade level over the years 2012-2014. This table is independent of school structure and allows for trend analysis within each grade level. Most grade levels have remained fairly consistent in growth level distributions. Fourth grade had the Low growth category decrease by 3% and High growth category increase by 3% over the three years of data. Sixth grade experienced a similar situation but only in reverse. The Low growth category increased by 3% and the High growth category decreased by 4%. The High growth categories in 5th and 8th grades experienced a decrease of 4% and 3% respectively from 2012 to 2014.

Table 4.7

<table>
<thead>
<tr>
<th>Grade</th>
<th>Low 1-34 Percentile</th>
<th>Typical 35-65 Percentile</th>
<th>High 66-99 Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>4th</td>
<td>22551</td>
<td>35%</td>
<td>20112</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>7482</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>8017</td>
<td>37%</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7052</td>
<td>32%</td>
</tr>
<tr>
<td>5th</td>
<td>21833</td>
<td>34%</td>
<td>19539</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>6919</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>7465</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7449</td>
<td>34%</td>
</tr>
<tr>
<td>6th</td>
<td>20929</td>
<td>33%</td>
<td>19511</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>6232</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>7623</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7074</td>
<td>33%</td>
</tr>
<tr>
<td>7th</td>
<td>21368</td>
<td>34%</td>
<td>19401</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>6868</td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>7429</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7071</td>
<td>33%</td>
</tr>
<tr>
<td>8th</td>
<td>21091</td>
<td>34%</td>
<td>19148</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>6710</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>6923</td>
<td>34%</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7458</td>
<td>35%</td>
</tr>
</tbody>
</table>
Table 4.8 displays the former table in a manner that lends itself to a cohort analysis. Although not a matched-pair design model, it has potential to provide some meaningful insights into group growth. The middle school level cohort that was 6th graders in 2012, 7th graders in 2013 and 8th graders in 2014 experienced an increase in the low growth category from 30% to 35%. This cohort lost students in the high growth category over these same years, a loss from 39% to 33%. There were no other noticeable trends. It will be important to analyze the cohort data after another year of data is collected.

Table 4.8

<table>
<thead>
<tr>
<th>Growth Levels (Baseline SGP) By Grade Level for Cohorts From 2012-2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Growth</td>
</tr>
<tr>
<td>4th    35%  37%  32%</td>
</tr>
<tr>
<td>5th    33%  35%  34%</td>
</tr>
<tr>
<td>6th    30%  36%  33%</td>
</tr>
<tr>
<td>7th    34%  36%  33%</td>
</tr>
<tr>
<td>8th    33%  34%  35%</td>
</tr>
<tr>
<td>Typical Growth</td>
</tr>
<tr>
<td>4th    31%  31%  31%</td>
</tr>
<tr>
<td>5th    29%  31%  31%</td>
</tr>
<tr>
<td>6th    30%  30%  32%</td>
</tr>
<tr>
<td>7th    31%  31%  32%</td>
</tr>
<tr>
<td>8th    31%  31%  31%</td>
</tr>
<tr>
<td>High Growth</td>
</tr>
<tr>
<td>4th    34%  32%  37%</td>
</tr>
<tr>
<td>5th    38%  35%  34%</td>
</tr>
<tr>
<td>6th    39%  34%  35%</td>
</tr>
<tr>
<td>7th    35%  34%  35%</td>
</tr>
<tr>
<td>8th    36%  36%  33%</td>
</tr>
</tbody>
</table>

Growth to standard is another feature of the SGP Methodology. Catch-up and Keep-up growth aggregate values provide a glimpse into projected performance for groups of students. These statistics are also available by student so that stakeholders can gain a sense of likelihood of skill attainment. This allows those stakeholders a targeted approach to provide student level support in order to achieve mastery or maintain
mastery. In order to meet the needs of all students the Move-up and Stay-up values should be considered. Move-up growth projects out those students that have already demonstrated mastery and are expected to move to an advanced performance level. Stay-up are those students already at the advanced performance level and expected to remain at this advanced level.

**Research Q2 Summaries**

To what extent can the results from the SGP Methodology be utilized to describe growth trends in various groups of Nebraska students?

**Gender Differences**

Table 4.9 lists the median values for both genders of the years available for growth statistics. A Mann-Whitney test indicated that the Baseline SGP was significantly higher in 2013 for female students ($Mdn = 50$) than for male students ($Mdn = 49$), $U = 1,361,604,479.5$, $p < .001$. However, the effect size (Cohen, 1988) was small ($r = -.02$), indicating small practical significance. Similarly in 2014, female students ($Mdn = 53$) were significantly higher than male students ($Mdn = 49$), $U = 1,370,340,077$, $p < .001$. However, the effect size was small $r = -.05$. The females students ($Mdn = 51$), independent of years, experiences significantly higher growth than the male students ($Mdn = 50$), $U = 12,171,986,897$, $p < .001$. The effect size was small $r = -.02$. 
Table 4.9

Median Growth (Baseline SGP) by Gender for 2012-2014

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>275,634</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>52,950</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>54,042</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>55,210</td>
</tr>
<tr>
<td>Female</td>
<td>261,248</td>
<td>51.0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>50,375</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>51,334</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>52,461</td>
</tr>
</tbody>
</table>

Ethnicity Differences

A Kruskal-Wallis test was conducted to determine differences among the groups of ethnicity in growth. The test was significant $\chi^2 (6, N = 316,372) = 2,412.95, p < .001$. A follow-up test, Mann-Whitney, was conducted to identify the groups that were significantly different. Table 4.10 below lists the pairwise comparisons and displays the significance and effect sizes.
Table 4.10

*Pairwise Comparisons Growth of Ethnicity Groups*

<table>
<thead>
<tr>
<th></th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL – AM</td>
<td>3,750.63</td>
<td>0.015</td>
<td>0.309</td>
<td></td>
</tr>
<tr>
<td>BL – H I</td>
<td>-13,662.02</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.05</td>
</tr>
<tr>
<td>BL – MU</td>
<td>-13,807.78</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.05</td>
</tr>
<tr>
<td>BL – PI</td>
<td>-21,204.47</td>
<td>0.000*</td>
<td>0.001*</td>
<td>-0.02</td>
</tr>
<tr>
<td>BL – WH</td>
<td>-25,587.79</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.06</td>
</tr>
<tr>
<td>BL – AS</td>
<td>38,212.37</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.14</td>
</tr>
<tr>
<td>AM – HI</td>
<td>-9,911.38</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.02</td>
</tr>
<tr>
<td>AM – MU</td>
<td>-10,057.15</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.04</td>
</tr>
<tr>
<td>AM – PI</td>
<td>-17,453.83</td>
<td>0.001*</td>
<td>0.019*</td>
<td>-0.04</td>
</tr>
<tr>
<td>AM – WH</td>
<td>-21,837.16</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.03</td>
</tr>
<tr>
<td>AM – AS</td>
<td>-34,461.74</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.14</td>
</tr>
<tr>
<td>HI – MU</td>
<td>-145.76</td>
<td>0.883</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>HI – PI</td>
<td>-7,542.45</td>
<td>0.139</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>HI – WH</td>
<td>-11,925.77</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.04</td>
</tr>
<tr>
<td>HI – AS</td>
<td>24,550.35</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.06</td>
</tr>
<tr>
<td>MU – PI</td>
<td>-7,396.69</td>
<td>0.152</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>MU – WH</td>
<td>-11,780.01</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.02</td>
</tr>
<tr>
<td>MU – AS</td>
<td>24,404.59</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.10</td>
</tr>
<tr>
<td>PI – WH</td>
<td>-4,383.32</td>
<td>0.389</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>PI – AS</td>
<td>17,007.90</td>
<td>0.001*</td>
<td>0.023*</td>
<td>0.03</td>
</tr>
<tr>
<td>WH – AS</td>
<td>12,624.58</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.02</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level

**NOTE:** Ethnicity: AM: American Indian, AS: Asian American, BL: Black, HI: Hispanic, MU: Multiple Races, PI: Pacific Islander, WH: White

Although the majority of the comparisons did have statistical significance, all of the effect sizes were trivial or small in nature.

**Special Education Differences**

A Mann-Whitney test was performed to determine overall growth differences between those that have qualified for SPED services and those that have not. There were significant ($U = 5,823,592,021, p < .001$) median differences for those qualifying ($Mdn = 42$) and those that do not ($Mdn = 52$). Those that did not qualify for SPED services did display more growth. However, the effect size was extremely minimal $r = -.09$. 
**English Language Learner Differences**

To determine if median growth differences were experienced in ELL students and non-ELL students, a Mann-Whitney test was conducted. There were significant differences in median growth ($U = 1,906,213,993$, $p < .001$). Those that were not identified as English Language Learners (ELL) had a higher median growth ($Mdn = 51$) than those that were identified ($Mdn = 46$) as ELL students. Cohen’s effect size ($r = -.02$) suggested no practical significance.

**Building Differences**

Kruskal-Wallis tests were conducted to determine median growth differences between buildings. The characteristics of the buildings that were examined were size (average grade size for the building), performance level (percent of students meeting or exceeding standards within the building), SPED enrollment (percent of students receiving SPED services within the building), and ELL enrollment (percent of students receiving ELL services within the building).

**Size – Student Count**

The Kruskal-Wallis test was significant $\chi^2(3, N = 315,101) = 275.05, p < .001$ for varying sizes of school buildings. This significance indicated that not all sizes of buildings experienced similar growth. A follow-up test, Mann-Whitney, was conducted to identify the groups that were significantly different. The pairwise comparisons are listed in Table 4.11. The medians for the largest, mid-large, mid-small, and smallest were, 50, 52, 51, and 52, respectively.
Table 4.11

*Follow-up Test Results for Different Types of Buildings – Size (Average Student Count)*

<table>
<thead>
<tr>
<th>Groups</th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>High – Low</td>
<td>8.959</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.01</td>
</tr>
<tr>
<td>High – Mid-Low</td>
<td>7.990</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.02</td>
</tr>
<tr>
<td>High – Mid-High</td>
<td>14.605</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.02</td>
</tr>
<tr>
<td>Mid-Low – Low</td>
<td>1.601</td>
<td>0.109</td>
<td>0.656</td>
<td></td>
</tr>
<tr>
<td>Mid-Low – Mid-High</td>
<td>-2.551</td>
<td>0.011</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>Low – Mid-High</td>
<td>.179</td>
<td>0.865</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level

The largest student count buildings (High) experienced significantly less growth than the other sizes of buildings. The effect sizes indicated that the practical significance were very small.

Growth scores were also analyzed for each year in which growth scores were available. Table 4.12 displays the building median SGP scores for the varying sizes of buildings over the 2012-2014 years.

Table 4.12

**Building Median SGP by Building Size (Average Grade Level) by Year**

<table>
<thead>
<tr>
<th>Building Size (Average Grade Level)</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi (4) $\bar{x} \geq 63.0$</td>
<td>48</td>
<td>46</td>
<td>49</td>
</tr>
<tr>
<td>Mid-High (3) $37.3 \leq \bar{x} &lt; 63.0$</td>
<td>51</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>Mid-Lo (2) $18.0 \leq \bar{x} &lt; 37.3$</td>
<td>49</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Lo (1) $\bar{x} &lt; 18.0$</td>
<td>49</td>
<td>53</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4.13 lists all three Kruskal-Wallis tests for SGP differences between the four building sizes for each of the three years in which growth scores were available.
Table 4.13

*Kruskal-Wallis Tests for SGP Differences in Building Size in 2012, 2013 and 2014*

<table>
<thead>
<tr>
<th>Year</th>
<th>Degrees of Freedom</th>
<th>N</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>3</td>
<td>102,389</td>
<td>82.725</td>
<td>0.000</td>
</tr>
<tr>
<td>2013</td>
<td>3</td>
<td>105,041</td>
<td>183.723</td>
<td>0.000</td>
</tr>
<tr>
<td>2014</td>
<td>3</td>
<td>107,671</td>
<td>50.442</td>
<td>0.000</td>
</tr>
</tbody>
</table>

There were significant differences across building size in each of the three years. Follow-up tests for each year were conducted to determine specifically the sizes of building that experienced growth differences. Table 4.14 outlines those differences.

Table 4.14

*Pairwise Comparisons for SGP Differences for Building Size in 2012, 2013 and 2014*

<table>
<thead>
<tr>
<th>Year</th>
<th>Group</th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>4 – 2</td>
<td>3.398</td>
<td>0.001</td>
<td>0.004</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>4 – 1</td>
<td>3.698</td>
<td>0.000</td>
<td>0.001</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>4 – 3</td>
<td>8.653</td>
<td>0.000</td>
<td>0.000</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>2 – 3</td>
<td>-3.209</td>
<td>0.001</td>
<td>0.008</td>
<td>-0.02</td>
</tr>
<tr>
<td>2013</td>
<td>4 – 3</td>
<td>11.203</td>
<td>0.000</td>
<td>0.000</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>4 – 2</td>
<td>6.789</td>
<td>0.000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>4 – 1</td>
<td>8.061</td>
<td>0.000</td>
<td>0.000</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>2 – 1</td>
<td>2.939</td>
<td>0.000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
<tr>
<td>2014</td>
<td>4 – 3</td>
<td>5.518</td>
<td>0.000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>4 – 2</td>
<td>5.399</td>
<td>0.000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
</tbody>
</table>

In all years the largest buildings experienced significantly less growth than at least one other sized building. The effect sizes in all pairwise comparisons suggested minimal practical significance.

**Student Performance**

The Kruskal-Wallis test was significant $\chi^2 (3, N = 315,101) = 5,904.81, p < .001$ for the different performing schools. The medians, from lowest performing buildings to highest performing, were 42, 49, 53, and 57, respectively. A follow-up test, Mann-
Whitney, was conducted to identify the groups that were significantly different. The pairwise comparisons are listed in Table 4.15.

Table 4.15

<table>
<thead>
<tr>
<th>Groups</th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low – Mid-Low</td>
<td>-32.37</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.06</td>
</tr>
<tr>
<td>Low – Mid-High</td>
<td>-55.16</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.10</td>
</tr>
<tr>
<td>Low – High</td>
<td>-71.61</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.14</td>
</tr>
<tr>
<td>Mid-Low – Mid-High</td>
<td>-22.19</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.04</td>
</tr>
<tr>
<td>Mid-Low – High</td>
<td>-40.21</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.08</td>
</tr>
<tr>
<td>Mid-High – High</td>
<td>-19.25</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level

Although all comparisons yielded significant results, the effect sizes suggest there were no practical significance between the varying levels of student performances between the buildings.

Special Education Services

The Kruskal-Wallis test was significant $\chi^2 (3, N = 315,101) = 1420.73, p < .001$. Not all building levels of SPED experienced similar growth. The medians, from lowest percentages of SPED students to highest percentages, were 52, 52, 46 and 47, respectively. Mann-Whitney post-hoc tests were conducted to identify which groups were significantly different. The pairwise comparisons are listed in Table 4.16.
Table 4.16

Follow-up Test Results for Different Types of Buildings – SPED Levels

<table>
<thead>
<tr>
<th>Groups</th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-High – High</td>
<td>-7.109</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.01</td>
</tr>
<tr>
<td>Mid-High – Mid-Low</td>
<td>29.587</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.05</td>
</tr>
<tr>
<td>Mid-High – Low</td>
<td>32.178</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.06</td>
</tr>
<tr>
<td>High – Mid-Low</td>
<td>18.762</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.04</td>
</tr>
<tr>
<td>High – Low</td>
<td>21.903</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.05</td>
</tr>
<tr>
<td>Mid-Low – Low</td>
<td>4.835</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.01</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level

Each grouping yielded statistically significant differences. However, the effect sizes suggested that there were no practical significance between varying sizes of SPED populations in school buildings.

English Language Learner Services

The Kruskal-Wallis test was significant $\chi^2 (3, N = 315,101) = 868.62, p < .001$.

Not all building levels of ELL services experienced similar growth. The medians, from no ELL students to highest levels of ELL students, were 51, 51, 47 and 46, respectively. Mann-Whitney post-hoc tests were conducted to identify which groups were significantly different. The pairwise comparisons are listed in Table 4.17.

Table 4.17

Follow-up Test Results for Different Types of Buildings – ELL Levels

<table>
<thead>
<tr>
<th>Groups</th>
<th>Std. Test Statistic</th>
<th>Significance</th>
<th>Adjusted Significance</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-High – High</td>
<td>-3.020</td>
<td>0.003*</td>
<td>0.015*</td>
<td>-0.01</td>
</tr>
<tr>
<td>Mid-High – Mid-Low</td>
<td>20.523</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.07</td>
</tr>
<tr>
<td>Mid-High – Low</td>
<td>23.941</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.09</td>
</tr>
<tr>
<td>High – Mid-Low</td>
<td>17.165</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.06</td>
</tr>
<tr>
<td>High – Low</td>
<td>20.827</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.08</td>
</tr>
<tr>
<td>Mid-Low – Low</td>
<td>6.138</td>
<td>0.000*</td>
<td>0.000*</td>
<td>0.02</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level
Even though all groups had significant differences when considering the effect sizes none of the group differences in ELL levels displayed practical significance.

**Summary**

Through the various analyses of growth patterns, there were not notable unique patterns that emerged. First, the growth model did display many encouraging properties. This model performed similarly to theoretical expectations. The SGP medians were very close to the expected value of 50. The SGP statistics were not influenced by student achievement and were confirmed by more than one perspective. This lack of relationship assumption was confirmed by performance level displays, scatterplots of growth versus both current and past performance, and correlations. Ceiling and floor effects did not appear in the visualizations at each grade level. Also, student groups were examined for atypical growth patterns. Quite often, there was statistical significance. However, the effect sizes showed miniscule or no practical effects.
CHAPTER V
DISCUSSION AND CONCLUSIONS

Introduction

In this study, the Student Growth Percentile (SGP) methodology was used to examine the growth patterns in elementary and middle school aged students on performance on the Nebraska State Accountability mathematics assessment (NeSA-M). Also investigated were differences in growth between gender, ethnicity, special education students, English language learners and various building characteristics -- size, student performance, levels of SPED students and levels of ELL students. The data used were Nebraska results from grades three through eight from the years of 2011, 2012, 2013 and 2014. This chapter provides a brief summary of findings and concludes with limitations and future considerations for research.

Review of Results and Meaning for Practice

The focus of this study was to quantify mathematical growth in a new perspective for the state of Nebraska. First, the assumptions to the growth model were also verified to ensure it performed as expected with Nebraska test results. The intent was to explore mathematical growth for cohorts of students at various age ranges. Also important was identifying any unusual growth patterns in various groups of students. The groups that were examined for group differences are: gender, ethnicity, SPED, ELL and building level differences in size, student performance, SPED services, and ELL services.

Analyses were performed to determine if the results were normed appropriately. The median growth percentile for all grade levels per year were nearly as expected -- 50. Only 4th grade had a value differing from 50, and its value was 49. Goodness-of-fit
reports were run for each grade level. The intent was to confirm that the nature of the distribution matches theory -- uniformly distributed and uncorrelated to performance (current and prior). The goodness-of-fit displays for all grades showed, the growth statistics were in fact uniformly distributed and not dependent on performance. The deciles were divided within each test performance level per grade level. The vertical lines were similar when comparing across performance levels, which indicated that the distribution throughout were similar. Also, the “Student Growth Percentile Range vs. Prior Scale Score Decile / Range” display illustrated the distribution between growth and prior performance per grade level. In theory, each cell should consist of 10% of the percent of each respective distributions -- growth and status. All visualizations showed a tight alignment to the expected results. In a few cases at lower grade levels, there were slight deviations from the theoretical values in a few cells. Finally, scatterplots by grade level were used to examine the relationship between growth (SGP) and prior achievement (scale score). No apparent relationships emerged in any of the grade levels. Also, there were growth values throughout the status continuum, which indicated that ceiling, and floor effects did not exist in the grade levels.

When the growth of cohorts were reviewed, there were slight differences across a few groups of students. However, no trends were noticeable in cohorts. Levels of growth were reviewed by grade level and by the level of the building. Neither produced consistent differences throughout the years.

8th (2014) grade decreased in growth from 39% to 34% to 33%. Both cohorts of students in this highest growth category experienced a decreasing percentage of students over the years of available data. Two other cohorts of students from the “Low” growth category experienced atypical growth patterns. For 6th grade (2012) – 7th grade (2013) – 8th grade (2014), their growth changed from 30% to 36% to 35%. The growth pattern indicated an overall increasing trend in this lowest growth category. The cohort of students within 5th grade (2012) – 6th grade (2013) – 7th grade (2014) yielded growth results from 33% to 36% to 33%. The testing year of 2013 marked the inaugural year that school districts were required to test students online in NeSA-M.

Medians from the Baseline SGP statistic were used to determine group differences between student groups of gender, ethnicity, SPED, ELL and building types. Even though the Mann-Whitney test did reveal significant differences in gender, the effect size was extremely minimal. A Kruskal-Wallis test and follow-up comparisons were utilized to examine differences in growth across ethnic groups. Sixteen of the twenty-one comparisons did yield significant differences. All of the effect sizes were very small and suggested little practical significance. Mann-Whitney tests were conducted for measuring SPED and ELL differences. Both had significant differences in medians, but the effect sizes were of no practical significance.

Finally, Kruskal-Wallis tests were computed to determine differences in various types of buildings -- size (average number of students per grade), student performance (percent of students meeting or exceeding), and levels of SPED and ELL services. The size of the buildings were categorized into four levels -- High, Mid-High, Mid-Low and Low. The largest buildings (High: average student count per grade level greater than or
equal to 63 students) displayed significantly lower levels of growth from the other sizes of buildings (Mid-High: average grade level student count of [37.3, 63) students, Mid-Low: average student count per grade level of [18, 37.3) students and Low: average grade level student count less than 18 students), as indicated by significance in median differences. Although, significant results were found the effect sizes suggested no practical significance. All other building types (performance, SPED services, ELL services) and pairwise comparisons either had non-significant results or no practical significance.

Summary of Discussion

It is evident through many perspectives of the study that the growth model performed as intended with Nebraska mathematics data. Assumptions and distributions were verified. The results that supported the need for a more detailed investigation were the pockets of trends in which growth categories, “High” and “Low”, experienced a decrease or increase in percent of students in the respective category. Additional years of data would provide further cohorts.

Limitations of Study

This study focused on the impact of the SGP growth model on Nebraska statewide math assessments and Nebraska students. The results are very applicable to Nebraska stakeholders. However, one should be cautious in generalizing results to other assessments or states. Goldschimdt et al. (2012) stressed that policymakers need to identify and prioritize the intent of an accountability system when developing and / or considering a growth model. The study did incorporate the most typical configurations from other states that used a similar growth model.
The major limitation of this study was the lack of access to the demographic variable of poverty. The impact of poverty on student performance has been documented and continues to be a challenge of school districts. One purpose of such a growth model would be to investigate not only the status of impoverished students, but also the growth of this group of students. It could show students or groups of students that were closing the achievement gap.

Another limiting factor was the availability of having only four years of achievement data. Four years allowed analysis of two three-year cohorts of growth to be analyzed. With additional years of data, subsequent cohorts results will be available. Additional cohort data would be critical if it was the desire of Nebraska policymakers to develop a criterion approach to growth expectations for future students.

There are a great number of configurations and business rules associated with any growth model. This is the case for the SGP methodology as well. Default settings within the SGP package loaded in R Studio along with the more commonly utilized state approaches to this growth model were incorporated into this study. This methodology does offer many other variations for consideration. Also, of note is the recognition that there are several aspects to an accountability system other than just the growth model.

Further Research

There is a large volume of business rules within an accountability system. Growth models are typically a subset of an accountability system. This study focused on measuring mathematical growth in a manner in which it can be most efficient to all stakeholders. The intent was to incorporate the most common configurations of the SGP model seen in the states already utilizing this approach. However, more analysis would
prove beneficial in regards to the statewide reading assessment, incorporation of the 11th
grade results, and evaluating if this approach of growth is worthy within an accountability
system.

Follow-up studies would be valuable as more state test results become available. The lower grades lack a wealth of previous scores. It would be advantageous if more years of data were included to provide balance. Also, more years would allow for more cohorts to analyze. This is especially true if a baseline criterion is to be established for future results. Having three three-year cohorts of student results to develop the targets would provide a better chance for a stable model. In addition, learning from other states that have implemented this approach would be very beneficial.

SGP methodology has been implemented in varying degrees by many states but still is somewhat a new approach for the educational world. Further research on the statistical properties of this approach, as well as other approaches, would help states improve on measuring growth for all learners.

As mentioned in the limitations section of this study, the impact of poverty on student learning would be of interest to many education stakeholders. This group of learners is a common focus of school districts across the state. Often, when achievement gaps are referenced, it is between those students who qualify for free or reduced priced lunches and those students paying full price lunches. If impoverished students are performing lower than their counterparts, it would be of value to monitor their growth and ensure they are growing at a higher rate. Another beneficial factor of such a study is to identify groups of unusually high growth among these impoverished students. If
decision-makers could learn from classrooms, where students of poverty excel, other buildings and districts could possibly replicate learning strategies.

A more in depth look at ways in which one could measure growth to a target would be appropriate. Analyzing “Catch-up,” “Stay-up,” “Move-up,” and “Keep-up” growth patterns could prove a more comprehensive way to acknowledge the growth of more types of students. This aspect of the SGP methodology is a way to monitor growth of more student groups. In a system that emphasizes grade level expectations or minimum expectations, stakeholders focus on the “floor.” However, if these other categories were included in a growth model, it would honor all groups of students. The catch-up growth would focus on those students who are not currently proficient, but are on track to meet the predetermined expectation. Keep-up growth would focus on those students at the expected level and who are projected to stay at this level. Move-up growth would quantify students moving from “meets” to “exceeds.” Finally, the Stay-up category would measure those students at the highest performance level and displayed growth that would maintain that level of achievement.

A related aspect of the projections is the assumption that students will continue to experience similar growth patterns moving forward. Also, more studies analyzing the appropriateness of this assumption could improve the accuracy of the projections. Perhaps this assumption is not appropriate for students that have recently experienced atypical growth. More studies could determine if there is a more effective approach to project future growth.

Based on the results of the growth patterns of cohorts of students within the categories of “High” vs. “Low” growth, more investigation on if transitioning to an
online test made a difference would be warranted. With additional results the future study could exclude the year in which the majority of students were tested in the paper / pencil format.

The Goodness of Fit (Appendix L) density plot for grade 4 displayed some variation from the theoretical cell value of ten. These cell values only appeared in the upper deciles for achievement and growth. Further examination of the characteristics of these learners would be beneficial.

Another possible study is to examine the impact teacher experience has on student growth. In examining only the large school buildings across the state of Nebraska, one could associate teacher experience to student growth experienced. Teacher experience could be categorized in those with three years or less experience as compared to those with more than three years.

A possible future endeavor could be to develop a qualitative perspective on student growth from key assessment experts across the state of Nebraska. The focus on that study could be through an interview panel to determine the impact on such a growth model, how student growth data are implemented at the district level, and other aspects that would emerge from conversing with practitioners. In order to truly understand the value of such a growth model, obtaining district viewpoints would be meaningful.

**Conclusions and Implications**

The purpose of this study was to examine the appropriateness of the SGP methodology with Nebraska mathematics results and explore growth patterns between a variety of student groupings. The study has the potential to assist policymakers by not only providing additional information regarding this particular growth model, but
providing a detailed look at results, reports and visualizations with actual Nebraska mathematics data.

Such a growth model could be a component of a state accountability system in which the model aids in several initiatives such as local organizational improvement, program evaluation, quantifying student learning growth, and federal and / or state regulations. If desired, further study on detailed business rules would need to take place.

Additionally, if such a growth model proved to meet the policy makers’ needs, it would be worthwhile to examine the statistical properties of quantile regression as related to the educational world. Statisticians could offer input as to the appropriateness of such a model.

The SGP package offered several visualization options. Figures 1 and 2 in Appendix J displayed scatterplots, which the Median Student Growth Percentiles for each building were plotted against the percent of proficient students within the building. A proportionately sized circle illustrated the school student count. The visualizations listed a vertical line at the 50th growth percentile and labeled the four quadrants. The quadrants were Higher Achievement / Higher Growth, Higher Achievement / Lower Growth, Higher Growth / Lower Achievement and Lower Growth / Lower Achievement. This visualization provided a perspective of status and growth for school buildings. Figure 1 used prior student achievement, whereas Figure 2, current student achievement.

The “Likelihood Graph For Future Performance” visualization (Appendix J – Figure 3) offered the likelihood of future student performance based upon current achievement at the lowest scale score for proficiency (scale score of 85). The shading indicated the three performance levels for NeSA-M. Figure 4 displayed the same
information for a student with current achievement level at the lowest scale score of the highest performance level, “Exceeds the Standards” (scale score of 135).

Sample school building visualizations were generated as Figures 5 and 6 demonstrated. Both were examples of school buildings in which students are plotted. Each student’s SGP and scale score were plotted. A vertical line noted the building’s Median SGP and horizontal lines partitioned the display into the three performance levels.

Also generated, were sample individual student growth reports (Appendix J – Figures 5 and 6). Individual student reports displayed all historical data, such as scale scores, achievement levels, SGPs and growth category. It also visualized the predicted student achievement for the upcoming year and the likelihood based on the growth categories using a fan display.

As reflected in this study, growth models do not follow a “one size fits all” approach. No one model is appropriate in all situations. Decision makers must prioritize their intent so that educated decisions can be made. Growth models have many varied approaches for implementation. It is important to make as well of an informed decision as possible. This study offers a first look at the SGP methodology in Nebraska.
REFERENCES


Betebenner and Linn Presenter Session 1.pdf


WashingtonGrowthModel.pdf


http://ir.uiowa.edu/etd/931.


October 21, 2013
Darin Kelberlau
Department of Educational Psychology
1228 Maplewood Dr Fremont, NE 68025

Delwyn Harnisch
Teaching, Learning and Teacher Education
125A HECO, UNL, 68588-0800

IRB Number: 20131013906EX
Project ID: 13906
Project Title: Measuring mathematical growth using the Student Growth Percentile (SGP) methodology

Dear Darin:

This letter is to officially notify you of the certification of exemption of your project. Your proposal is in compliance with this institution's Federal Wide Assurance 00002258 and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46) and has been classified as exempt, category 4.

We wish to remind you that the principal investigator is responsible for reporting to this Board any of the following events within 48 hours of the event:

* Any serious event (including on-site and off-site adverse events, injuries, side effects, deaths, or other problems) which in the opinion of the local investigator was unanticipated, involved risk to subjects or others, and was possibly related to the research procedures;

* Any serious accidental or unintentional change to the IRB-approved protocol that involves risk or has the potential to recur;

* Any publication in the literature, safety monitoring report, interim result or other finding that indicates an unexpected change to the risk/benefit ratio of the research;

* Any breach in confidentiality or compromise in data privacy related to the subject or others; or
*Any complaint of a subject that indicates an unanticipated risk or that cannot be resolved by the research staff.

This project should be conducted in full accordance with all applicable sections of the IRB Guidelines and you should notify the IRB immediately of any proposed changes that may affect the exempt status of your research project. You should report any unanticipated problems involving risks to the participants or others to the Board.

If you have any questions, please contact the IRB office at 472-6965.

Sincerely,

Becky R. Freeman, CIP for the IRB
APPENDIX B

REQUEST FOR DATA
FROM NEBRASKA DEPARTMENT OF EDUCATION

Data Access and Use Policy and Procedures

URL:
http://www.education.ne.gov/nssrs/docs/Nebraska_Data_Policy_December_2010.pdf
APPENDIX C

STEPS FOR RSTUDIO INSTALL AND R CODE FOR SGP PACKAGE INSTALL

Download R Studio for any of the following operating systems: Mac, Windows, Ubuntu, and openSUSE.

URL: http://www.rstudio.com/products/rstudio/download/

After install R Studio, launch the program.

Loading the SGP package (within the R Studio console):

    require(devtools)

    install_github("dbetebenner/SGP")
library(SGP)

NeSA <- read.delim("~/Desktop/NeSA.txt")

Nebraska_SGP <- prepareSGP(NeSA, state="NE")

Nebraska_SGP <- analyzeSGP(Nebraska_SGP)

Nebraska_SGP <- combineSGP(Nebraska_SGP)

Nebraska_SGP <- summarizeSGP(Nebraska_SGP)

Nebraska_SGP <- visualizeSGP(Nebraska_SGP)
APPENDIX E

SGP TECHNICAL OVERVIEW & PACKAGE ‘SGP’ DOCUMENTATION

URL:


http://cran.r-project.org/web/packages/SGP/SGP.pdf
APPENDIX F

NESAM TABLE OF SPECIFICATIONS

NeSA-M Website:
http://www.education.ne.gov/Assessment/NeSa_Math.htm

Grade 3: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade3.pdf
Grade 4: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade4.pdf
Grade 5: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade5.pdf
Grade 6: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade6.pdf
Grade 7: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade7.pdf
Grade 8: http://www.education.ne.gov/Assessment/pdfs/Math_TOS_Grade8.pdf
APPENDIX G

NESAM STUDENT REFERENCE SHEETS

NeSA-M Website
http://www.education.ne.gov/Assessment/NeSa_Math.htm

Note: Grade 3 has no reference sheet.

Grade 4

Spanish:
http://www.education.ne.gov/Assessment/pdfs/2013_GR04_MATH_REFERENCE_SHEET_SP.pdf

Grade 5

Spanish:
http://www.education.ne.gov/Assessment/pdfs/2013_GR05_MATH_REFERENCE_SHEET_SP.pdf

Grade 6

Spanish:
http://www.education.ne.gov/Assessment/pdfs/2013_GR06_MATH_REFERENCE_SHEET_SP.pdf

Grade 7:

Spanish:
http://www.education.ne.gov/Assessment/pdfs/2013_GR07_MATH_REFERENCE_SHEET_SP.pdf

Grade 8:

Spanish:
http://www.education.ne.gov/Assessment/pdfs/2013_GR08_MATH_REFERENCE_SHEET_SP.pdf
APPENDIX H

NESA-M TECHNICAL REPORTS

2014 Technical Report:


2014 Appendices:

http://www.education.ne.gov/Assessment/pdfs/Final_Appendices_2014_NeSA_Tech_Report.pdf

2013 Technical Report:


2013 Appendices:


2012 Technical Report:


2011 Technical Report:

APPENDIX I

NESAM RAW SCORE TO SCALE SCORE CONVERSION TABLES 2011-14

URL:

http://www.education.ne.gov/assessment/NeSA_Scoring.htm
APPENDIX J

SAMPLE SGP PACKAGE REPORTS

Figure 1: SGP Growth vs. Percent of Students at / above Proficiency from Prior School Year: Each bubble represents a school. The size of the bubble is representative of the size of the school building.

Figure 2: SGP Growth vs. Percent of Students at / above Proficiency from Current School Year: Each bubble represents a school. The size of the bubble is representative of the size of the school building.
APPENDIX J

SAMPLE SGP PACKAGE REPORTS

Figure 3: Likelihood Graph for Future Performance: starting at the “Meets Standards” performance level cut point.
Figure 4: Likelihood Graph on Future Performance: starting at the “Exceeds Standards” performance level cut point.
APPENDIX J

SAMPLE SGP PACKAGE REPORTS

Figure 5: School Building Growth vs. Achievement Report: Each circle represents a student.

Figure 6: School Building Growth vs Achievement Report: A larger building with more students.
APPENDIX J
SAMPLE SGP PACKAGE REPORTS

Figure 7: Student Growth Report: Shows prior and current achievement. It also illustrates likelihood of performance in the upcoming year. This sample student is nearing proficiency and shows how likely it is to achieve this level in 2015.

Figure 8: Student Growth Report: Another sample student.
APPENDIX K

SGP GOODNESS-OF-FIT DESCRIPTIVES GRADES 4 & 5 REPORTS

Student Growth Percentile Goodness–of–Fit Descriptives
2014 Mathematics SGP, Grade 4 (N = 22,146)

SGP Deciles by Prior Achievement Level

Exceeds: (25.7%)
Meets: (48.8%)
Below: (25.5%)

Median Student Growth Percentile

1  10  20  30  40  50  60  70  80  90  99

Student Growth Percentile Goodness–of–Fit Descriptives
2014 Mathematics SGP, Grade 5 (N = 21,734)

SGP Deciles by Prior Achievement Level

Exceeds: (22.5%)
Meets: (30.5%)
Below: (37.1%)

Median Student Growth Percentile

1  10  20  30  40  50  60  70  80  90  99
APPENDIX K

SGP GOODNESS-OF-FIT DESCRIPTIVES_grades 6 & 7 REPORTS

Student Growth Percentile Goodness–of–Fit Descriptives
2014 Mathematics SGP, Grade 6 (N = 21,440)

SGP Deciles by Prior Achievement Level

Prior Achievement Level (Mathematics)

<table>
<thead>
<tr>
<th>Decile</th>
<th>Median Student Growth Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds</td>
<td>99</td>
</tr>
<tr>
<td>Meets</td>
<td>50</td>
</tr>
<tr>
<td>Below</td>
<td>1</td>
</tr>
</tbody>
</table>

(Student Growth Percentile Goodness–of–Fit Descriptives
2014 Mathematics SGP, Grade 7 (N = 21,220)

SGP Deciles by Prior Achievement Level

Prior Achievement Level (Mathematics)

<table>
<thead>
<tr>
<th>Decile</th>
<th>Median Student Growth Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds</td>
<td>99</td>
</tr>
<tr>
<td>Meets</td>
<td>50</td>
</tr>
<tr>
<td>Below</td>
<td>1</td>
</tr>
</tbody>
</table>
APPENDIX K

SGP GOODNESS-OF-FIT DESCRIPTIVES GRADE 8 REPORT

Grade 8

Student Growth Percentile Goodness-of-Fit Descriptives
2014 Mathematics SGP, Grade 8 (N = 21,131)

SGP Deciles by Prior Achievement Level

- Exceeds (23.1%)
- Meets (47.1%)
- Below (29.9%)

Median Student Growth Percentile

1 10 20 30 40 50 60 70 80 90 99
## APPENDIX L

### SGP GOODNESS-OF-FIT DECILE GRADE 4 REPORT

#### Grade 4

<table>
<thead>
<tr>
<th>Prior Scale Score Decile/Range</th>
<th>1 to 9</th>
<th>10 to 19</th>
<th>20 to 29</th>
<th>30 to 39</th>
<th>40 to 49</th>
<th>50 to 59</th>
<th>60 to 69</th>
<th>70 to 79</th>
<th>80 to 89</th>
<th>90 to 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st/[1,63)</td>
<td>10.74</td>
<td>10.73</td>
<td>10.34</td>
<td>10.00</td>
<td>8.73</td>
<td>10.05</td>
<td>9.61</td>
<td>9.46</td>
<td>9.90</td>
<td>10.39</td>
</tr>
<tr>
<td>2nd/[63,79)</td>
<td>9.73</td>
<td>8.51</td>
<td>9.86</td>
<td>10.52</td>
<td>9.78</td>
<td>10.04</td>
<td>10.74</td>
<td>10.39</td>
<td>9.38</td>
<td>11.05</td>
</tr>
<tr>
<td>6th/[109,119)</td>
<td>9.93</td>
<td>8.77</td>
<td>11.82</td>
<td>8.51</td>
<td>10.74</td>
<td>10.14</td>
<td>8.08</td>
<td>9.57</td>
<td>11.52</td>
<td>10.83</td>
</tr>
<tr>
<td>7th/[119,131)</td>
<td>8.71</td>
<td>10.55</td>
<td>8.74</td>
<td>11.21</td>
<td>8.94</td>
<td>11.21</td>
<td>11.21</td>
<td>10.90</td>
<td>9.13</td>
<td>9.40</td>
</tr>
<tr>
<td>8th/[131,141)</td>
<td>9.96</td>
<td>10.59</td>
<td>10.27</td>
<td>12.86</td>
<td>7.96</td>
<td>7.80</td>
<td>12.75</td>
<td>10.91</td>
<td>7.22</td>
<td>9.69</td>
</tr>
<tr>
<td>9th/[141,155)</td>
<td>9.28</td>
<td>11.03</td>
<td>10.70</td>
<td>11.68</td>
<td>8.73</td>
<td>13.97</td>
<td>9.39</td>
<td>4.53</td>
<td>12.83</td>
<td>7.86</td>
</tr>
<tr>
<td>10th/[155,200]</td>
<td>10.61</td>
<td>9.55</td>
<td>9.89</td>
<td>12.91</td>
<td>8.83</td>
<td>10.41</td>
<td>10.75</td>
<td>10.10</td>
<td>6.85</td>
<td>10.10</td>
</tr>
</tbody>
</table>

*Prior scale score deciles can be uneven depending upon the prior score distribution.
# APPENDIX L

## SGP GOODNESS-OF-FIT DECILE GRADE 5 REPORT

**Grade 5**

<table>
<thead>
<tr>
<th>Prior Scale Score Decile/Range</th>
<th>1 to 9</th>
<th>10 to 19</th>
<th>20 to 29</th>
<th>30 to 39</th>
<th>40 to 49</th>
<th>50 to 59</th>
<th>60 to 69</th>
<th>70 to 79</th>
<th>80 to 89</th>
<th>90 to 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd ([63,77])</td>
<td>9.11</td>
<td>10.32</td>
<td>9.11</td>
<td>9.64</td>
<td>10.22</td>
<td>9.59</td>
<td>10.27</td>
<td>10.85</td>
<td>10.17</td>
<td>10.71</td>
</tr>
<tr>
<td>3rd ([77,89])</td>
<td>9.44</td>
<td>9.35</td>
<td>9.58</td>
<td>9.94</td>
<td>9.94</td>
<td>10.07</td>
<td>10.03</td>
<td>9.76</td>
<td>11.52</td>
<td>10.39</td>
</tr>
<tr>
<td>4th ([89,99])</td>
<td>9.87</td>
<td>10.15</td>
<td>9.78</td>
<td>9.87</td>
<td>10.15</td>
<td>10.43</td>
<td>9.78</td>
<td>10.48</td>
<td>9.31</td>
<td>10.20</td>
</tr>
<tr>
<td>7th ([115,125])</td>
<td>9.59</td>
<td>10.28</td>
<td>9.73</td>
<td>10.51</td>
<td>9.23</td>
<td>10.00</td>
<td>10.28</td>
<td>9.91</td>
<td>10.05</td>
<td>10.42</td>
</tr>
<tr>
<td>8th ([125,138])</td>
<td>10.16</td>
<td>10.30</td>
<td>10.48</td>
<td>9.81</td>
<td>9.90</td>
<td>9.54</td>
<td>9.14</td>
<td>9.72</td>
<td>10.74</td>
<td>10.21</td>
</tr>
<tr>
<td>9th ([138,156])</td>
<td>9.17</td>
<td>9.39</td>
<td>10.23</td>
<td>9.83</td>
<td>10.49</td>
<td>10.05</td>
<td>10.49</td>
<td>10.14</td>
<td>10.23</td>
<td>9.96</td>
</tr>
<tr>
<td>10th ([156,200])</td>
<td>10.12</td>
<td>9.97</td>
<td>10.20</td>
<td>9.89</td>
<td>9.62</td>
<td>10.35</td>
<td>10.05</td>
<td>10.24</td>
<td>10.59</td>
<td>8.97</td>
</tr>
</tbody>
</table>

*Prior score deciles can be uneven depending upon the prior score distribution*
APPENDIX L

SGP GOODNESS-OF-FIT DECILE GRADE 6 REPORT

Grade 6

<table>
<thead>
<tr>
<th>Prior Scale Score Decile/Range *</th>
<th>1 to 9</th>
<th>10 to 19</th>
<th>20 to 29</th>
<th>30 to 39</th>
<th>40 to 49</th>
<th>50 to 59</th>
<th>60 to 69</th>
<th>70 to 79</th>
<th>80 to 89</th>
<th>90 to 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd/[66,78)</td>
<td>9.82</td>
<td>8.67</td>
<td>8.89</td>
<td>10.10</td>
<td>10.43</td>
<td>9.88</td>
<td>10.32</td>
<td>10.54</td>
<td>9.88</td>
<td>11.47</td>
</tr>
<tr>
<td>3rd/[78,90)</td>
<td>9.11</td>
<td>10.01</td>
<td>10.30</td>
<td>9.93</td>
<td>9.60</td>
<td>10.05</td>
<td>10.47</td>
<td>9.81</td>
<td>10.05</td>
<td>10.67</td>
</tr>
<tr>
<td>4th/[90,97)</td>
<td>9.42</td>
<td>9.73</td>
<td>9.24</td>
<td>10.52</td>
<td>10.21</td>
<td>10.27</td>
<td>9.85</td>
<td>8.81</td>
<td>11.79</td>
<td>10.15</td>
</tr>
<tr>
<td>9th/[138,154)</td>
<td>9.39</td>
<td>9.98</td>
<td>10.21</td>
<td>9.84</td>
<td>10.62</td>
<td>10.03</td>
<td>9.34</td>
<td>10.07</td>
<td>10.21</td>
<td>10.30</td>
</tr>
</tbody>
</table>

*Prior score deciles can be uneven depending upon the prior score distribution.
Appendix L

SGP Goodness-of-Fit Decile Grade 7 Report

Grade 7

<table>
<thead>
<tr>
<th>Prior Scale Score Decile/Range</th>
<th>1st (1.54)</th>
<th>2nd (54,70)</th>
<th>3rd (70,82)</th>
<th>4th (82,94)</th>
<th>5th (94,105)</th>
<th>6th (105,117)</th>
<th>7th (117,128)</th>
<th>8th (128,143)</th>
<th>9th (143,166)</th>
<th>10th (166,200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.63</td>
<td>9.26</td>
<td>9.84</td>
<td>9.34</td>
<td>10.16</td>
<td>10.05</td>
<td>9.84</td>
<td>9.84</td>
<td>10.21</td>
<td>10.32</td>
</tr>
<tr>
<td></td>
<td>(8.9%)</td>
<td>(10.1%)</td>
<td>(9.3%)</td>
<td>(10.7%)</td>
<td>(9.6%)</td>
<td>(11.1%)</td>
<td>(9.4%)</td>
<td>(9.7%)</td>
<td>(10.6%)</td>
<td>(10.6%)</td>
</tr>
</tbody>
</table>

*Prior score deciles can be uneven depending upon the prior score distribution.*
APPENDIX L

SGP GOODNESS-OF-FIT DECILE GRADE 8 REPORT

Grade 8

<table>
<thead>
<tr>
<th>Prior Score Decile/Range</th>
<th>1 to 9</th>
<th>10 to 19</th>
<th>20 to 29</th>
<th>30 to 39</th>
<th>40 to 49</th>
<th>50 to 59</th>
<th>60 to 69</th>
<th>70 to 79</th>
<th>80 to 89</th>
<th>90 to 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st ([1, 57])</td>
<td>10.38</td>
<td>10.43</td>
<td>11.00</td>
<td>9.61</td>
<td>9.95</td>
<td>9.66</td>
<td>9.61</td>
<td>9.76</td>
<td>9.61</td>
<td>10.00</td>
</tr>
<tr>
<td>2nd ([57, 71])</td>
<td>9.79</td>
<td>9.14</td>
<td>8.92</td>
<td>9.73</td>
<td>9.46</td>
<td>10.60</td>
<td>9.79</td>
<td>11.06</td>
<td>9.95</td>
<td>11.52</td>
</tr>
<tr>
<td>3rd ([71, 85])</td>
<td>8.40</td>
<td>9.80</td>
<td>9.76</td>
<td>9.80</td>
<td>10.39</td>
<td>10.27</td>
<td>10.39</td>
<td>10.01</td>
<td>10.86</td>
<td>10.31</td>
</tr>
<tr>
<td>4th ([85, 95])</td>
<td>11.11</td>
<td>9.84</td>
<td>9.99</td>
<td>9.73</td>
<td>9.23</td>
<td>10.45</td>
<td>9.73</td>
<td>9.89</td>
<td>9.33</td>
<td>10.70</td>
</tr>
<tr>
<td>5th ([95, 105])</td>
<td>9.23</td>
<td>10.12</td>
<td>10.30</td>
<td>10.35</td>
<td>10.49</td>
<td>9.33</td>
<td>9.51</td>
<td>9.98</td>
<td>10.07</td>
<td>10.63</td>
</tr>
<tr>
<td>6th ([105, 114])</td>
<td>8.99</td>
<td>10.84</td>
<td>10.10</td>
<td>10.15</td>
<td>9.99</td>
<td>9.25</td>
<td>11.63</td>
<td>9.78</td>
<td>8.72</td>
<td>10.57</td>
</tr>
<tr>
<td>7th ([114, 124])</td>
<td>9.05</td>
<td>9.79</td>
<td>9.37</td>
<td>10.00</td>
<td>10.16</td>
<td>10.11</td>
<td>9.79</td>
<td>10.21</td>
<td>10.11</td>
<td>11.43</td>
</tr>
<tr>
<td>8th ([124, 136])</td>
<td>10.53</td>
<td>8.99</td>
<td>10.10</td>
<td>9.71</td>
<td>9.23</td>
<td>11.26</td>
<td>8.55</td>
<td>10.53</td>
<td>11.16</td>
<td>9.95</td>
</tr>
<tr>
<td>9th ([136, 160])</td>
<td>9.03</td>
<td>10.44</td>
<td>9.82</td>
<td>10.29</td>
<td>10.13</td>
<td>9.42</td>
<td>10.13</td>
<td>10.05</td>
<td>10.09</td>
<td>10.60</td>
</tr>
<tr>
<td>10th ([160, 200])</td>
<td>9.58</td>
<td>9.84</td>
<td>10.18</td>
<td>10.01</td>
<td>10.05</td>
<td>10.44</td>
<td>10.57</td>
<td>9.88</td>
<td>8.93</td>
<td>10.53</td>
</tr>
</tbody>
</table>

*Prior score deciles can be uneven depending upon the prior score distribution*