ABSTRACT

REAMER, AMY CRAIG. Characterizing the Progression of Performance in Mathematics over Time: The Application of Markovian Models to an Education System. (Under the direction of Dr. Julie Simmons Ivy (co-chair) and Dr. Robert E. Young (co-chair)).

Transforming Science, Technology, Engineering, and Mathematics (STEM) education in the United States K-12 system is a topic of national concern. Many reports have documented the deficiencies in the vertical alignment of STEM learning in the K-12 education system, citing the need for young students to build a solid foundation in mathematics. Currently, many students fail to master essential mathematics concepts before advancing to more complex applications, leaving them ill-prepared to succeed in higher level STEM coursework.

In this study, we use Markovian models to characterize the evolution of students' performance in mathematics in the elementary and middle grades. This research constitutes a first step in developing a Markovian framework for modeling a large-scale, learning system as a stochastic process.

We first present a longitudinal analysis of student performance on the standardized North Carolina End-of-Grade (EOG) mathematics exam and use Markov chain models to probabilistically characterize the movement of students' scores from one grade level to the next. We conduct a descriptive statistical analysis of one cohort of students to identify demographic factors that affect performance. We use the Markov chain model to forecast a student's level of proficiency in mathematics by the eighth grade given their level of achievement in the third grade.

Next, to examine a method for optimizing the vertical alignment of STEM learning, we model the sequential process of learning mathematics over time using a Markov Decision

Process (MDP) and identify optimal learning intervention strategies. Specifically, we formulate an MDP model of EOG test performance to determine an optimal control policy for the administration of retests of the exam and conduct a numerical experiment. Many of the numerical cases suggest that it is optimal to retest students within the lowest level of mathematical proficiency, contradicting the state of North Carolina's current policy on retesting. As a result of the numerical experiment, we hypothesize that a control-limit policy can be proven given sufficient conditions.

The Markovian models described in this research demonstrate a progression of complexity in modeling an uncertain environment, an education system. The Markov chain describes the natural progression of a student's performance on a standardized test over time, while the MDP allows the educator to dynamically alter a student's performance trajectory by their selection of a learning intervention. Future work will characterize a student's accumulation of mathematical knowledge over time; assuring the development of a solid foundation in mathematics. These models offer a structure for the application of optimization modeling techniques in the K-12 STEM education system.

Characterizing the Progression of Performance in Mathematics over Time: The Application of Markovian Models to an Education System

by
Amy Craig Reamer

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2012

APPROVED BY:

_____	_____
Dr. Julie Simmons Ivy Committee Co-Chair	Dr. Robert E. Young Committee Co-Chair
_____	_____
Dr. Russell E. King	Dr. Dianne Raubenheimer
_____	
Dr. Anita Vila-Parrish	

DEDICATION

"Trust in the Lord with all your heart and lean not on your own understanding. In all your ways acknowledge Him, and He will direct your path" (Proverbs 3: 5-6). First and foremost, thank you Lord for holding my hand through the many celebrations and challenges of life.

This dissertation document is dedicated to my husband, Daniel, my family, and to my close friends. To my husband, I could not have done this without you. You have supported me in every way possible and I could never thank you enough for that. I love you with all of my heart HB. To my Mom, Dad, and sister, Meredith, I thank you for always being there for me. This dissertation is the culmination of a lifelong dream and I could not have done it without your love and abiding faith in me. You always told me to do my best, and this effort is just that.

To Lindsey, Jennifer, and Lauren, whom I started my NC State journey with so many years ago, thank you for your constant encouragement and for reminding me to wear my "thinking scarf." To the many other family members and friends who so thoughtfully listened to my "nerd work" discussions through the years, I sincerely appreciate your friendship and support.

And lastly, to our son, Jackson, this dissertation project is evidence that hard work and determination pays off. I hope that someday you will be proud of my perseverance and dedication to finishing this project.

BIOGRAPHY

Amy Craig Reamer was born in Kinston, NC and her family moved to Wilmington, NC when she was 6 months old. She attended E.A. Laney High School in Wilmington, graduating #5 in her class in 1999. She attended NC State University, majoring in Industrial and Systems Engineering, and graduated with her Bachelor of Science degree in 2003 where she was recognized as a university valedictorian. Amy proceeded to attend the NC State University graduate school and obtained her Master of Industrial Engineering degree in 2004. She accepted a position as a cost engineer in the Personal Computing Division of IBM in Research Triangle Park in 2005. After nearly two years in the corporate world, Amy decided to return to NC State University as the Coordinator of Student-Owned Computing in the College of Engineering Information Technology group. At that time she decided to begin her pursuit of a doctoral degree in Industrial and Systems Engineering and passed the written qualifying examination with concentrations in Production and Ergonomics in 2007. She accepted a position as the project manager for a large NSF funded research project (MINDSET), under the direction of Dr. Robert Young, in the Edward P. Fitts Department of Industrial and Systems Engineering in 2008 and served in that capacity until she began working on her doctoral research full-time in 2010. Amy also became a certified Pilates instructor in 2008. She married Dr. R. Daniel Reamer, D.D.S. in June 2010 and they had their first child, Jackson Davis Reamer, in September 2011. She currently resides in Wilmington, NC with her family and two beloved cats, Miss Ellie and Mr. Nacho.

ACKNOWLEDGMENTS

I would like to graciously acknowledge Dr. Julie Ivy, my committee chair, for her contributions to these research efforts. Dr. Ivy, this research is what it is because of your vision and guidance. I have endless respect for you and sincerely thank you for taking your valuable time to mentor me. To Dr. Robert Young, my committee co-chair, I thank you for your support of my career over the past decade and for giving me the opportunity to work on the MINDSET project. It was an honor to work on MINDSET, as well as this research, under your leadership. To Dr. Anita Vila-Parrish, thank you for your support and friendship throughout this process. You are a gifted advisor and I have benefited from your counsel. To Dr. Russell King, much of my academic success is a function of your support. You helped me obtain scholarships and provided just the right words of encouragement when I needed them most. As an undergraduate, you once told me that "the cream always rises to the top" after I scored lower than I would have liked on an exam. I have often thought about that remark. To Dr. Dianne Raubenheimer, you are one of my favorite colleagues and I hold your opinions in high regard. Thank you for your encouragement over the past several years. I would also like to acknowledge the many other NC State colleagues and friends with whom I have crossed paths through the years. Each of you played an integral role in leading me to this day. And finally, to Clarence Smith, my "Raleigh Dad," I want to express my gratitude for always having your door open for me. As I said in your retirement speech, you gave me faith when I had very little and you gave me guidance at every turn of my academic journey. Thanks C!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 - Understanding the Evolution of Mathematics Performance in Primary Education and the Implications for STEM Learning: A Markovian Approach

1. Introduction

Many reports have documented the growing concern that the United States education system is not adequately preparing our students to succeed in the technology-driven workforce of the 21st century (National Science Board 2007, National Academy of Science et al. 2007, National Center for Education Statistics 2006, U.S. Department of Education 2008). In 2007, the National Science Board drafted an action plan for the STEM education system. According to the report, "Almost 30% of students in their first year of college are forced to take remedial science and math classes because they are not prepared to take college-level courses" (National Science Board 2007). The National Academies Press' 2007 *Rising Above the Gathering Storm* report cites the following statistics to motivate this research, "Fewer than one-third of US eighth grade students performed at or above a level called 'proficient' in mathematics…Alarmingly, about one-fifth of the fourth graders and one-third of the eighth graders lacked the competence to perform even basic mathematical computations" (National Center for Education Statistics 2006). The National Research Council presents another discouraging statistic citing Schmidt (2011) stating that "three-quarters of eighth graders still enter high school not having reached the proficient level" (National Research Council 2011, pg. 19). Although mathematics is one pillar of STEM education, it is arguably the most important because its concepts and methodologies pervade all science, technology, and engineering disciplines (National Research Council 2001).

The National Science Board argues that there are two primary deficiencies in the K-12 STEM education system that need to be addressed (2007). First, STEM education is not aligned "horizontally," i.e., mathematics standards and course sequences vary greatly from state to state (see also Common Core Standards 2010). Second, there is little or no "vertical" alignment of STEM learning at the elementary and middle school levels, suggesting that students do not master critical concepts before moving on to more complex ideas in high school (National Science Board 2007, pg. 5). The National Council of Teachers of Mathematics (NCTM) published a report in 2006 to address the issue of vertical alignment. The report identified curriculum focal points and important mathematical topics for prekindergarten through eighth grade (NCTM 2006). These topics along with standards adopted by the states provide a framework for curriculum development at each grade level, reinforcing the creation of a STEM pipeline.

A solid foundation in mathematics in early education is critical to a student's success in higher level science and mathematics coursework in high school (Common Core Standards 2010). This is essential for college-bound students, particularly those interested in majoring in STEM disciplines (Tyson 2011, Nicholls et al. 2010). Some studies suggest that students that take Algebra I in middle school (typically in eighth grade) are more likely to transition to higher education and potentially enroll in STEM degree programs (U.S. Department of Education 2008, President's Council of Advisors on Science and Technology 2010, Loveless 2008). In North Carolina, students are assessed on their preparation to transition into Pre-Algebra or Algebra I as early as the sixth grade. A number of studies have reported the importance of "algebra-readiness" and the difficulties some students have in transitioning

from general arithmetic to algebraic thinking (National Research Council 2001, Pearson School 2008, Loveless 2008). In fact, in the second largest school district in North Carolina, enrollment in Algebra I in middle school has doubled since 2009; however the question remains, are these students truly prepared to succeed in algebra (Hui & Goldsmith 2012)? In this study we consider preparation for higher level mathematics courses, starting with Algebra I, as a function of performance on the North Carolina End-of-Grade (EOG) mathematics tests.

Our research is motivated by the need for quantitative methods to represent how students perform in mathematics over time. As a result of the No Child Left Behind (NCLB) Act of 2001, statistical models of K-12 education based on standardized testing data have been developed to assess individual student's academic progress over time (Sanders 2006, U.S. Department of Education 2012). This work is the first step in developing a Markovian framework to forecast students' development of mathematical knowledge, as assessed by the standardized EOG mathematics tests administered annually to students in grades 3-8. Future work will focus on the development of an optimal control system to maximize a student's level of proficiency by the eighth grade, but in order to achieve this; we first establish the Markov chain models and validate them against results reported by the state.

In this chapter, we model elementary and middle school students' performance in mathematics as a stochastic process to evaluate their proficiency, as measured by EOG results. We consider performance on the EOG mathematics test to be an informative measure of student learning. We develop a Markov chain model of student performance on the EOG mathematics test and evaluate the overall utility of the model in the context of education. The

Markov chain is used to improve the understanding of the progression of a student's performance in mathematics over time and forecast a student's future success in the discipline. The transition probability matrices of the Markov chain provide a novel longitudinal analysis of EOG assessment data. Further, we are interested in analyzing differences in performance by various demographic factors such as gender and ethnicity.

This chapter is organized as follows. In the remainder of Section 1 we present background information on the North Carolina public school system, a review of relevant literature, and a discussion of the data source. We present the descriptive statistical analysis of the demographic variables of interest and a Markov chain model of the problem in Section 2. In Section 3, we discuss the results of the aggregate Markov chain model followed by the Markov chains based on the demographic factors. We discuss conclusions, the limitations of the model, and areas of future research in Section 4.

1.1 Background

The North Carolina public school system consists of 115 school districts serving nearly 1.4 million students per year and is home to the 16th and 18th largest public school districts in the country (Public Schools of North Carolina 2009a, Wake County Public School System 2011, Proximity 2011). In the 1996-1997 academic year, the North Carolina State Board of Education introduced the ABCs of Accountability System as a statewide school improvement program initiative to assess the performance of North Carolina public schools (Public Schools of North Carolina 2011a, 2011b). The assessment program sets growth and performance standards for all public schools in the state of North Carolina. It was expanded in 2002-2003 to accommodate the requirements of the No Child Left Behind Act of 2001

(Public Schools of North Carolina 2011b). Integral to the Accountability System are the state-mandated End-of-Grade (EOG) tests in mathematics and reading comprehension administered to students in grades 3-8 during the last three weeks of each academic year. In the 2009-10 academic year over 655,000 students took the EOG mathematics test (Public Schools of North Carolina 2011c). Wide scale EOG testing was introduced in 1995 and these multiple choice, standardized exams "are designed to measure student performance on the goals, objectives and grade-level competencies specified in the Standard Course of Study," the statewide K-12 curriculum (Public Schools of North Carolina 2011d, 2011e). A student's raw score on the EOG test is converted into a scale score which falls into one of four categories known as *achievement levels*. "Achievement levels are used to provide an interpretation of student performance relative to pre-determined standards based on ranges of scale scores" (Public Schools of North Carolina 2006a). A descriptor for each respective achievement level is shown in Table 1-1. A student whose scale score falls within achievement levels III or IV is considered "proficient" by the state and consistently performs at or above grade-level expectations (Public Schools of North Carolina 2006c).

Table 1-1: Description of North Carolina End-of-Grade (EOG) test achievement levels (Public Schools of North Carolina 2006b)

Achievement Level	Descriptor	North Carolina Proficiency Status
I	Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.	Non-Proficient in Mathematics
II	Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.	Non-Proficient in Mathematics
III	Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.	Proficient in Mathematics
IV	Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work.	Proficient in Mathematics

Although the ranges of scale scores corresponding to each achievement level may vary from year to year, the definitions of the respective achievement levels remain consistent. Achievement levels are specifically defined for each grade level and correspond to the competencies and goals outlined for that grade level in the Standard Course of Study. "Unlike percentiles, which yield only relative comparisons, the achievement levels give common meaning throughout the state as to what is expected at various levels of competence in each subject area" (Public Schools of North Carolina 1999). The achievement levels provide a metric for comparing the relative performance of students. According to testing protocol, students at higher achievement levels are assumed to have mastered the competencies and goals associated with the lower achievement levels (Public Schools of North Carolina 2006b). We refer the reader to the following resources for additional

information on the EOG test development process and data on the test's reliability and validity: Sanford (1996), Public Schools of North Carolina (1999), Public Schools of North Carolina (2006a), Public Schools of North Carolina (2011f), and Public Schools of North Carolina (2011g).

For this study, EOG testing data was obtained from the North Carolina Education Research Data Center (NCERDC) at Duke University (NCERDC 2012). The NCERDC stores and manages data on public school districts, schools, students, and teachers collected by the North Carolina Department of Public Instruction (DPI). We refer the reader to the following examples of research efforts using NCERDC data: Muschkin et al. (2009), NCERDC (2008), Goldhaber and Anthony (2007), Miranda et al. (2007), Cook et al. (2008), and Bifulco and Ladd (2004). In contrast to these studies, we use NCERDC EOG assessment data to perform a longitudinal analysis by building a Markov chain model of student performance on the EOG mathematics test. To the best of our knowledge this is the first study to model the data in this way.

1.2 Objectives

The objectives of the research conducted in this chapter are to:

- Longitudinally track EOG mathematics test achievement level scores of various cohorts of students from third grade to eighth grade and probabilistically characterize the movement of students' achievement level scores from one grade level to the next.
- Conduct a statistical analysis of one representative cohort of students to compare test performance of various student populations and identify demographic factors that are related to performance.

- Develop a first-order Markovian model of student performance in the context of primary mathematics education. Create a Markov chain of EOG mathematics test achievement level scores to represent the progression of performance of a 'typical' student. Based on the results of the statistical analysis, develop additional Markov chains to examine the performance of specific student populations.
- Use the Markov chain to estimate the probability that a student will be "proficient" in mathematics by eighth grade.

This work is our first step in developing a framework to forecast students' development of mathematical knowledge in primary education in North Carolina. The model has the potential to provide educators with a quantitative measure of a student's preparation for higher level STEM course work.

1.3 The Benefits of a Markovian Model

Markovian models have been applied in a number of fields including machine learning, medical decision making, and behavioral ecology, but to a limited extent in education systems (Monahan 1982). A Markov chain is a discrete-time, stochastic process that can be described by a series of finite states and a set of probabilities that characterize the transitions between states (Howard 1960). With regard to the probabilistic nature of the learning process, learning trajectory literature asserts that the process of learning is sequential and dynamic and that a student faces obstacles along their learning trajectory (Confrey et al. 2009). These obstacles depict the inherent randomness in the learning pathway, suggesting that the process is stochastic in nature. We hypothesize that the process of learning can be broken down into a set of sequenced events that can be effectively modeled as a Markov

process. In this study we associate a student's cognition with their performance on the EOG test. Markov chains satisfy the Markov property, which states that the probability of occupying a future state is conditional on being in the present state, and does not depend on past states (Howard 1960). According to learning theory literature, a student's current state of cognition can be considered as independent of their collection of past knowledge (Smallwood 1971, Simon & Tzur 2004, Confrey & Maloney 2010). We submit that the "memoryless" property of the Markov chain is beneficial for modeling the sequential process of accumulating knowledge over time. As true in many time series analyses, the most recent data point collected often provides the best fit for the selected forecast model. This provides justification for using a Markovian model that uses a student's most recent state of knowledge to forecast their future knowledge; explicitly representing the dependent relationship between the present and the future.

Regression models are commonly used in the social sciences; however, Markovian models offer additional benefits in this context. Multinomial logistic regression models provide short-term predictions of the probabilities of various outcomes of a categorical dependent variable. In contrast, the one-step transition probabilities of the Markov chain can easily be used to forecast model behavior several steps or periods into the future. In this study we are interested in determining a student's level of proficiency by eighth grade based upon their level of achievement in third grade. Such questions cannot be answered as simply or as elegantly with regression analyses as their ability to forecast beyond the next immediate time period is limited. In addition, regression analyses offer a static representation of student

performance. The Markov chain model lays the foundation for a control infrastructure where we can dynamically optimize sequential decisions about a student's education.

1.4 Relevant Literature

1.4.1 Value-Added Models in K-12 Education

In the past decade, assessment models known as value-added models have been developed for K-12 education to quantitatively measure an individual student's year-to-year gain in knowledge as well as to produce a measure of teacher effectiveness (i.e., to quantify the "value" teachers "add" to a student's education). One of the most widely used value-added models, is the Education Value-Added Assessment System (EVAAS) developed by Dr. William Sanders and his colleagues at the University of Tennessee at Knoxville, and is available as a commercial software package from SAS (SAS EVAAS 2012). Several states, including North Carolina, have made EVAAS available to school districts in an effort to evaluate the effectiveness of a student's education (Public Schools of North Carolina 2012a). In North Carolina, EVAAS is populated with EOG and End-of-Course (EOC) test data and the model longitudinally tracks student performance over time. The software provides diagnostic reports, identifies at-risk students, and predicts student success on future academic milestones such as future EOG tests and the SAT (SAS EVAAS 2012).

EVAAS uses two types of statistical models: a multivariate response model (MRM) and a univariate response model (URM) (Sanders et al. 1997, Wright et al. 2010). The MRM is a longitudinal, linear mixed model that incorporates a student's entire history of performance on the EOG test. It is "essentially a multivariate repeated-measures ANOVA model," where "the entire set of observed test scores are fitted simultaneously" (Wright et al.

2010, pg. 1). By considering the entire testing history, the MRM places weight on past performance. In comparison, we use a first-order Markovian model where a student's most recent score on the EOG mathematics test is used to forecast their score on the next installment of the exam. In other words, the path to the current state is not deemed as important as the current state itself for making predictions about the future. The URM is similar to an ANCOVA and is used when the MRM is not appropriate, such as with EOC tests where a mean gain in performance does not make sense (Wright et al. 2010). A traditional ANCOVA is used to calculate student success probabilities on future tests (Wright et al. 2010). Since an ANCOVA is essentially a regression analysis, as previously mentioned, we advocate that a Markovian model can more effectively forecast performance several periods into the future.

There is controversy surrounding the practical use of value-added models, such as EVAAS, in assessing teacher effectiveness and due to the overall complexity of the statistical methods employed (Amrein-Beardsley 2008, Sanders et al. 2009). It is not our intention to engage in the debate surrounding the measurement of teacher quality or effectiveness, but rather we tout the benefits of longitudinally tracking individual student performance on standardized tests. In fact, our model could serve to address the concerns of Amrein-Beardsley (2008) regarding the validity of EVAAS by verifying the model's student success probabilities. Many school districts in North Carolina use these probabilities to make Algebra I placement decisions, such as in Wake County where students are placed into Algebra I in middle school if EVAAS indicates that the student has a "70 percent probability of passing Algebra I by eighth grade" (Hui & Goldsmith 2012). A Markovian model can be used to

evaluate the optimality of this 70 percent benchmark, as it is a control system. EVAAS, on the other hand, cannot be used as a responsive, control model; it is a descriptive statistical model based on regression techniques. As evidenced by McCaffrey et al. (2004), Raudenbush and Byrk (2002), Tekwe et al. (2004), and Ballou et al. (2004), rigorous statistical models can be used to assess student performance. In this context, we propose the use of a Markovian model for its relative transparency as compared to some of the other statistical methods employed in value-added modeling. The Markov chain maps a sequential pathway from one exam score to the next. Further research in this area is required and we aim to offer a new approach to forecasting student performance using standardized test data.

1.4.2 Markov Chains in Education Modeling

Markov chains have been used to model some aspects of teaching and learning theory related to mathematics and engineering education. McFarland (2006) used a Markov model to analyze the "structure and dynamics of participant flows across math courses in two high schools" (pg. 177). McFarland (2006) calculated a transition probability matrix of students moving from one mathematics course to another in a rural high school and in a magnet high school. McFarland (2006) used logistic regression models to explain curricular moves such as leaving school, repeating a course, or jumping to another career path. McFarland (2006) presented a robust model of the sequence of mathematics courses taken by students at two high schools; however, he did not consider a student's mathematical knowledge collected prior to entering high school. In this study, we use test performance in early education to forecast a student's preparation for high school STEM courses.

Perdikaris (1992) used a Markov chain to model mathematics teachers' decision making and their process of inquiry in the classroom. The focus of Perdikaris (1992) was modeling teacher behavior whereas our objective is to model student performance. Per Howard (1960), the state space of a Markov chain is assumed to be mutually exclusive and collectively exhaustive, representing all possible outcomes. The states in Perdikaris (1992) did not appear to be mutually exclusive and it was difficult to understand how the underlying process moved through time. Furthermore, with a sample size of two teachers, the model lacked predictive power (Young & Sweet 1982a, 1982b).

Voskoglou used Markov chains to model problem solving (1996), the process of learning (2009), and the process of mathematically modeling real world problems (2007). The state space of the model in Voskoglou (2009) appeared to be contiguous where learning occurred along a continuum of conceptual phases or 'steps.' A limitation of this study was that the state transitions and decision epochs of the model were unclear. In our study, we examine transitions over time where the decision epochs for the finite horizon Markov chain are represented by grade levels. Pinar-Yildirim et al. (2007) used direct observation and a work sampling methodology to observe teamwork on student design projects in a capstone engineering course. The state space of the Markov chain was divided into nine subtasks (e.g., project management, research, group discussion) representing the engineering design process. Again, the authors acknowledged that it was difficult to discern whether or not the states were mutually exclusive.

Markov chains have been used to model other aspects of education systems. Shah and Burke (1999) used a Markov chain to model the higher education system in Australia.

Kuravsky and Malykh (2004) developed a continuous time, discrete state Markov process of longitudinal IQ data of Russian school children ages 6-14. Fam et al. (2006) presented a Markov chain model of academic dishonesty in an undergraduate course. However, to the best of our knowledge, Markovian models of the primary mathematics education system have not been developed. The Markovian model presented herein will improve our understanding of a student's acquisition of mathematical knowledge by eighth grade.

In summary, as demonstrated by McFarland (2006), Markov chains have the potential to be used to improve the understanding of a student's progression through mathematics. However, many of the existing models appear to lack rigor. Three challenges persist: defining mutually exclusive states, describing how the states change over time according to time-dependent probability distributions, and securing access to statistically significant sample sizes. As indicated by Perdikaris (1992) and Voskoglou (2009), much of the research in this area relies on anecdotal evidence based on one or two classrooms, making it difficult to generalize the results to a larger population. While models of the undergraduate engineering education system have emerged in the past decade (Klingbeil et al. 2005, Shah & Burke 1999, Pinar-Yildirim et al. 2007), it appears that longitudinal Markovian models of the K-8 mathematics education system have not been developed. In our review of prior research, we have found no evidence of Markov chains for modeling consecutive years of elementary and middle school students' performance in mathematics. Our objective is to develop a Markovian model of student performance in mathematics and to use the model to assess a student's mathematical preparation by the eighth grade.

1.4.3 The Development of a STEM Pipeline

The development of a STEM pipeline is a topic of national interest that is well documented in the literature (National Academy of Science et al. 2007, U.S. Department of Education 2007). The purpose of this study is not to evaluate existing STEM pipeline research or K-12 engineering outreach curricula, but rather we offer a method for modeling performance in mathematics in early education as a potential measure of preparation for pathways to STEM fields. Many studies have mapped success in algebra to taking advanced courses in high school and future interest in STEM disciplines. Burkam and Lee (2003) used multivariate regression models to investigate "mathematics pipeline" measures by examining high school coursework, taking into account students' gender, ethnicity, and socioeconomic status. Nicholls et al. (2010) used various statistical techniques to predict the probability of obtaining a STEM degree based on student variables available as early as the eighth grade. Tyson et al. (2007) studied the connection between high school science and mathematics coursework and enrollment in STEM majors at Florida universities, with particular emphasis on gender and racial disparities. Tyson (2011) used multinomial logistic regression analyses to determine the effects of taking advanced physics and calculus courses in high school on engineering degree attainment.

These analyses of the STEM pipeline consider longitudinal student performance just as Markovian models consider probabilistic movements over time. The Markov chain models student performance as a stochastic process. In this study we consider preparation for higher level mathematics courses, starting with Algebra I, as a function of performance on the EOG mathematics tests. By focusing on standardized assessments administered in the third through

eighth grade, we can objectively assess a critical window of mathematical preparation and learning. This, in turn, will improve the understanding of a student's foundation of knowledge acquired in early education and will allow educators to forecast a student's future success in STEM disciplines. We propose that the results of the Markov chain model could inform existing models that link high school course taking to STEM degree attainment.

2. Methods

2.1 Data

All procedures for acquiring, storing, and analyzing the NCERDC data used in this study were approved by the Duke University Institutional Review Board (IRB) as well as the NC State University IRB (Miranda et al. 2007). Each year, the NCERDC creates a database containing students' EOG mathematics and reading test scale scores and achievement level scores with select demographic and socioeconomic data. All student data is encrypted and de-identified; names and other recognizable identifiers are replaced by a unique master identification number. These identification numbers are created by the NCERDC to identify individual test scores and cannot be traced to specific students, thus meeting state and federal human subjects research requirements. In this study, key demographic data of interest included gender, ethnicity, participation in a free/reduced price school lunch program, exceptionality status, i.e., their status as academically/intellectually gifted and/or learning disabled, and English proficiency. Statewide test results were obtained from 1995 to 2009, the most recent year of data available at the time of this study.

2.1.1 Characteristics of the Data

The EOG tests change from one academic year to the next. Thus, not only the condition of the student changes from year to year, but the assessment instrument itself changes. To account for this in our analysis, we rely on the state's annual classification of scale scores into the achievement levels. Although the test and scaling of the scores vary each year, the definitions of the achievement levels remain consistent. As is the case in any public school system, various state and local policies and procedures may change from one academic year to the next impacting the structure of the data the NCERDC receives from DPI. In 2006, the scaling of the EOG mathematics test scores was changed significantly to align with the 2003 Standard Course of Study and the new growth formulas and performance standards associated with the North Carolina ABCs of Accountability System. According to a background report, "The current ABCs formulas are different enough from the original ones that comparisons between the performance of schools from 2006 forward and prior years should be avoided" (Public Schools of North Carolina 2011b). Although achievement level descriptors were not changed, upon review of the data, achievement level scores were impacted considerably in the year of the changeover. Thus, in creating the aggregate Markov chains, we conducted two separate analyses, one prior to 2006 and another from 2006 and beyond, to ensure consistency in the longitudinal comparisons.

2.1.2 Creating Longitudinal Datasets of Cohorts of Third Grade Students

The NCERDC maintains separate EOG test score databases for each grade level for each academic year. Student records were linked longitudinally by the student's unique master identification number. We longitudinally tracked ten different cohorts of third grade

students through eighth grade from 1995 to 2009. For example, the 1995-2000 cohort included third grade students that took the EOG tests in the 1994-1995 academic year and subsequently took the eighth grade EOG tests in the 1999-2000 academic year. Based on information obtained from the NCERDC, we expected to track approximately 60-70% of third grade students through eighth grade due to attrition (i.e., if a student was retained in a grade or moved out of the state). These percentages have increased over time, as data management tools have improved since the mid-1990s. These results were validated in our analysis and the percentage of students tracked for each cohort is shown in Table 1-2. The sample sizes are quite large, indicating statistical strength. A list of variable descriptions created by the NCERDC is provided in Appendix A. An example of the data mining procedures followed for a cohort is shown in Appendix B.

Table 1-2: Sample sizes of longitudinal datasets and the percentage of North Carolina students tracked from third to eighth grade (before exclusions, see Appendix B)

Cohort	Total Records in Longitudinal Dataset	Percentage Tracked from 3rd Grade
1995-2000	51,280	58.01
1996-2001	55,414	61.41
1997-2002	62,889	67.09
1998-2003	65,088	66.29
1999-2004	67,516	67.38
2000-2005	68,695	68.21
2001-2006	69,888	70.47
2002-2007	71,308	72.53
2003-2008	74,428	73.89
2004-2009	73,238	72.53

2.1.3 Data Analysis Assumptions

We made the following assumptions in our analysis of the NCERDC data.

1) We excluded records with duplicate identification numbers and records missing EOG test scores from each raw dataset for each grade level for each academic year (fewer than 5% of total records on average).
2) If separate records were included to indicate that a student took a retest, in this analysis we only considered a student's score on the original administration of the test, thus excluding all retest scores. We refer the reader to Appendix A and Chapter 2 for a description of the state of North Carolina's policy on retesting.
3) Although there appear to be inconsistencies in the coding of the Limited English Proficiency (LEP) status variables, we found that excluding the LEP students adversely impacted our ethnicity distribution counts, thus these records were included in the analyses. We acknowledge that using standardized test scores may less accurately reflect mathematical proficiency for LEP students; however, the percentage of LEP students in each cohort is small (fewer than 4.5% of each dataset), and given the large sample sizes of the longitudinal datasets, their inclusion did not skew our results.
4) In creating the longitudinal datasets, we excluded students who were retained in a grade. We evaluated students that moved with their respective cohort from third to eighth grade. Students retained in a grade may follow a different Markov chain and will be studied in future analyses. Further discussion of this model limitation is provided in Section 3.

5) After creating a longitudinal dataset for each cohort of students, we reviewed the sample sizes of each variable category within each cohort to determine if any additional exclusion was required. Due to small sample sizes, we excluded all learning disabled (LD) students and otherwise exceptional students. However, students identified as academically/intellectually gifted (AIG) were not excluded from subsequent analyses. Many of the other exceptional student variable categories had small sample sizes ($n < 100$ students), which could compromise the confidentiality requirements of the IRB and the NCERDC data use agreement.

2.2 Descriptive Statistical Analysis

We hypothesize that performance on the EOG mathematics test may vary as a function of gender, ethnicity, socioeconomic status, and exceptionality status. To assess potential differences in the student attributes of interest, we conducted statistical hypothesis testing on select covariates to identify statistical differences in EOG mathematics achievement level scores among the various student populations. We calculated confidence intervals based on the Student's *t*-distribution and conducted standard *t*-tests to compare population means. A full discussion of the statistical methods employed and the results from one cohort of students (2000-2005) on the gender, ethnicity, school lunch program, and AIG mathematics variables are presented in Appendix C. We chose this cohort because it was the last cohort that was unaffected by the changes made to the scale scores in 2006. We also calculated the correlation between the mathematics and reading comprehension test scores for this cohort.

Select key results from the statistical hypothesis testing follow. We found that true statistical and practical differences exist between AIG and non-AIG students, where AIG students scored significantly higher than non-AIG students on the EOG mathematics test. Students that participate in a school lunch program scored significantly lower than those that do not. These findings for the AIG and school lunch program variables were consistent across all ten cohorts of students studied. As a result, separate Markov chains need to be developed for each of these types of students to further examine the differences in student performance on the EOG mathematics test over time.

With regard to ethnicity and using the NCERDC variable names defined in Appendix A, Asian and White students scored significantly higher than all other ethnicities considered in this study, while Black students consistently scored lower. Given the smaller sample sizes of the Hispanic, American Indian, and Multi-Racial populations, it is difficult to accurately determine if differences amongst students of these ethnicities are statistically significant. However, of these three ethnic groups, Multi-Racial students scored the highest. Although statistically significant differences in mean achievement level scores were found amongst the ethnic groups, these differences did not necessarily translate into true differences in achievement level scores. For example, by eighth grade, the average achievement level score of students of all ethnicities exceeded 3.0, indicating proficiency. In contrast, in third grade, Black, Hispanic, and American Indian students scored less than 3.0, on average, indicating non-proficiency in third grade. Separate Markov chains of the ethnic groups will provide a more comprehensive perspective of the ethnicity differences.

Upon review of the gender differences, male students scored higher than female students across all grade levels. Although statistical differences were found, practically the population means were nearly equivalent. Specifically by the eighth grade the difference between average male and female achievement levels scores was nominal. Nevertheless, given the gender disparities in STEM interest and enrollment, it is interesting to review separate Markov chains of male and female performance to identify potential trends. As suspected, there is significant correlation between EOG mathematics test scale scores and EOG reading comprehension test scale scores.

2.3 Markov Chain Formulation

Based on the findings from this statistical analysis, we present several Markov chains, one where we consider all types or categories of students to represent the trajectory of a 'typical' student, and others where we examine the effects of gender, ethnicity, socioeconomic status, and AIG mathematics status on EOG mathematics test performance. Recall, a Markov chain is a discrete-time stochastic process that satisfies the Markov property. A stochastic process with the Markov property implies that a future state depends only on the present state of the system and not on past states (Howard 1960). This is consistent with the assumptions about the process of learning outlined in the learning theory literature that says that a student's current state of cognition is independent of their collection of past knowledge (Smallwood 1971, Simon & Tzur 2004, Confrey & Maloney 2010). This is also consistent with North Carolina DPI testing protocol that states that students at higher achievement levels are assumed to have mastered the competencies and goals associated with the lower achievement levels (Public Schools of North Carolina 2006b). Further, a student's

past performance on the EOG tests prior to the previous grade level is not used to evaluate their proficiency at the current grade level. Consequently, we assume the model satisfies the Markov property and we assume a Markovian order of one.

A state-transition diagram associated with a student's movement between EOG achievement level scores is shown in Figure 1-1. This diagram shows the probabilities of the various movements between scores from the third to fourth grade for the pre-2006 aggregate cohort (full results are presented in Figure 1-2). In this context, we define the "state," S_t, as a student's EOG mathematics test achievement level score at the end of grade t. In the diagram, the nodes represent the states of the system and the arrows represent the possible movements in scores from one grade level to the next. For example, referring to Figure 1-1, a student that scores within achievement level I (state 1) in third grade will score within achievement level II (state 2) in fourth grade with probability 0.56. There is evidence of movement between all states in the data; therefore, we consider all possible movements without constraint. The objective is to calculate the probability that a student will score within achievement levels III or IV by eighth grade. North Carolina students that score within levels III or IV are classified as "proficient" in mathematics and we associate a student's proficiency with their preparation for higher level STEM coursework in high school and beyond.

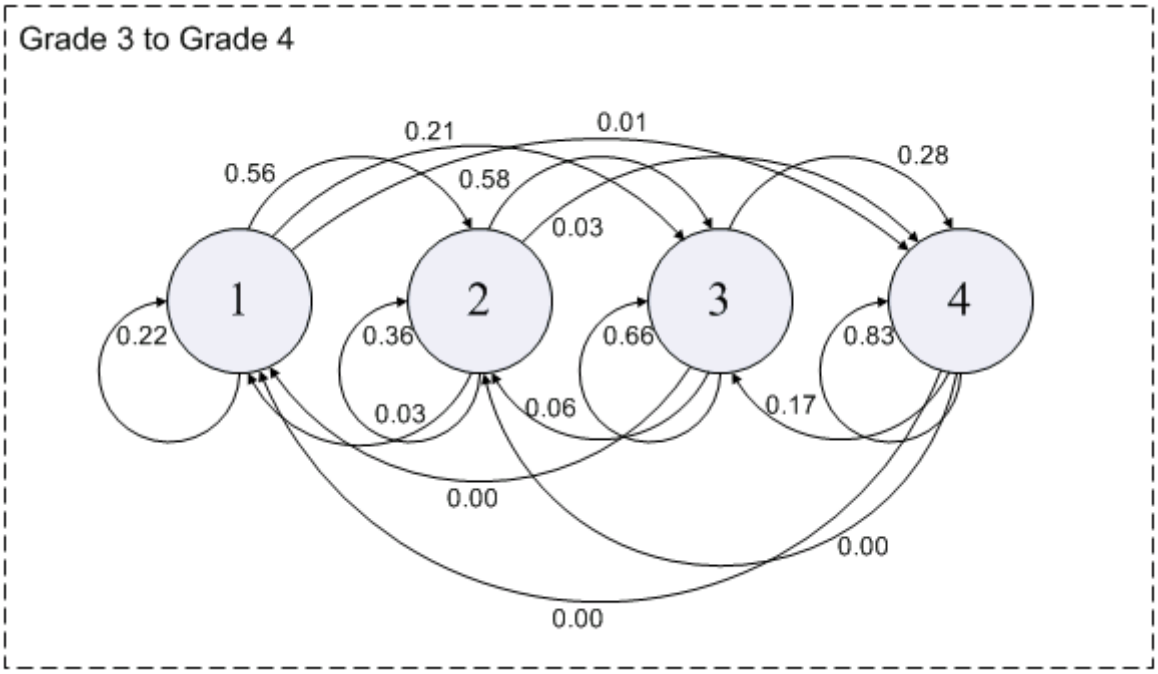


Figure 1-1: Markov chain state-transition diagram (As an example, the one-step transition probabilities from the third to fourth grade are labeled)

The structure and notation of the model is defined as follows. As mentioned previously, the states correspond to a student's EOG mathematics test achievement level score at the end of grade t, where S_t: {I, II, III, IV} ~ {1, 2, 3, 4}. The set of discrete time intervals are defined by T: {0, 1, 2, 3, 4, 5} and represent the grade levels in which the EOG mathematics tests are administered, where elements of T are denoted by t and $t = 0$ represents the third grade. The one-step transition probabilities, $P_t = [p_{ij}(t)]$, are the set of conditional probabilities that specify the likelihood of a transition from state i to state j from time period t to $t + 1$, where $\sum_{j=1}^{n} p_{ij}(t) = 1 \; \forall \; t$. In other words, given that a student performs at achievement level i at time t, the student will perform at achievement level j at time t +1 with

probability $p_{ij}(t)$. Longitudinal EOG mathematics test achievement level scores were used to calculate the various probabilities of movement from one achievement level to another. Let

$$p_{ij}(t) = n_{ij}(t)/n_i(t)$$

where $n_{ij}(t)$ = the number of students in state i at grade level t and state j at grade level $t + 1$, and $n_i(t)$ = the total number of students in state i in grade level t. Using the Chapman-Kolmogorov equations, we calculate the multi-step transition probabilities to forecast student performance in eighth grade given a specific level of achievement in third grade (i.e., five "steps"/grades into the future).

There are several model assumptions. First, we assume classification into an achievement level is sufficient to represent a student's mathematical performance. Second, as discussed previously, we assume this stochastic process satisfies the Markov property. A student's achievement level in period $t + 1$ is assumed to be related to their current achievement level in period t, but is independent of past levels. Third, we assume period $t = 0$ corresponds to the third grade. Students are not pre-classified into an achievement level prior to third grade, thus the first EOG mathematics test data point is collected in third grade. According to the Markov property, the time between transitions is constant.

For the Markov chain of a typical student, we calculated the one-step transition probability matrices for each grade level for each of the ten cohorts of students (see Appendix D). Due to the changes in the scale scores implemented in 2006, we performed two separate analyses: one prior to 2006 and another from 2006 to 2009.

1.) Pre-2006 Analysis

Six cohorts were considered in the pre-2006 analysis: the 1995-2000, 1996-2001, 1997-2002, 1998-2003, 1999-2004, and 2000-2005 cohorts. First, for each of the aforementioned cohorts, we calculated the one-step transition probability matrices for each grade level (see Appendix D). Next, we created an aggregate cohort in an effort to create one transition probability matrix per grade level for use in the Markov chain. We combined the six longitudinal datasets from the cohorts to create one large sample (n = 338,080 student records) and recalculated the one-step transition probability matrices for each grade level. In this analysis, we seek to represent the performance of a typical student on the EOG mathematics test, thus, we created the aggregate cohort to minimize any potential cohort effects.

2.) 2006-2009 Analysis (Post-2006 Analysis)

Four cohorts were considered in the post-2006 analysis: the 2001-2006, 2002-2007, 2003-2008, and 2004-2009 cohorts. As in the previous analysis, for each cohort we calculated the one-step transition probability matrices for each grade level (see Appendix D). The changes to the scale scores on the EOG mathematics test occurred during the 2005-2006 academic year. In that year, the transition probability matrices are inconsistent and may not be representative of actual transitions in test performance. Per the North Carolina DPI, longitudinal comparisons before 2006 and afterward are not advised (Public Schools of North Carolina 2011b). As a result, we had to omit various matrices from the aggregate cohort. Due to the stepwise nature of the longitudinal data and given that the 2009 scores were the most recent year of data available at the time of this study; we were unable to calculate transition

probability matrices for the transition from third grade to fourth grade and the transition from fourth grade to fifth grade.

3. Results and Discussion

3.1 Markov Chain of a Typical Student

The one-step transition probability matrices for a typical student from the aggregate cohorts are shown in Figure 1-2. For example, in the first matrix of the pre-2006 cohort, a typical student that scores within achievement level I in third grade will advance to achievement level II in fourth grade with probability 0.56 as shown in the shaded cell.

Pre-2006 Aggregate Cohort
Cohorts: 95-00, 96-01, 97-02, 98-03, 99-04, 00-05

Grade 3 to Grade 4 (n = 338,080 records)

3 -> 4	1	2	3	4
1	0.221106	0.561259	0.210457	0.007179
2	0.033340	0.356774	0.579107	0.030779
3	0.001472	0.055238	0.662346	0.280945
4	0.000092	0.001406	0.164851	0.833652

Grade 4 to Grade 5 (n = 338,080 records)

4 -> 5	1	2	3	4
1	0.305355	0.540280	0.146647	0.007718
2	0.072709	0.433237	0.464869	0.029186
3	0.003841	0.078545	0.571337	0.346278
4	0.000084	0.001407	0.085567	0.912943

Grade 5 to Grade 6 (n = 338,080 records)

5 -> 6	1	2	3	4
1	0.200677	0.560497	0.233634	0.005192
2	0.059210	0.390723	0.530610	0.019457
3	0.005047	0.093135	0.686167	0.215651
4	0.000124	0.002071	0.146399	0.851406

Grade 6 to Grade 7 (n = 338,080 records)

6 -> 7	1	2	3	4
1	0.196641	0.553893	0.239695	0.009771
2	0.071644	0.437703	0.462708	0.027944
3	0.006361	0.113450	0.609038	0.271152
4	0.000092	0.001834	0.094259	0.903815

Grade 7 to Grade 8 (n = 338,080 records)

7 -> 8	1	2	3	4
1	0.256090	0.485993	0.245432	0.012485
2	0.107527	0.392618	0.473091	0.026764
3	0.015110	0.128405	0.667487	0.188998
4	0.000363	0.004090	0.142140	0.853407

Post-2006 Aggregate Cohort
Cohorts: 01-06, 02-07, 03-08, 04-09

Grade 3 to Grade 4 (n = 0 records)

3 -> 4	1	2	3	4
1	N/A			
2				
3				
4				

Grade 4 to Grade 5 (n = 0 records)

4 -> 5	1	2	3	4
1	N/A			
2				
3				
4				

Grade 5 to Grade 6 (n = 62,948 [04-09 cohort])

5 -> 6	1	2	3	4
1	0.352003	0.567889	0.079769	0.000339
2	0.108517	0.538835	0.348380	0.004268
3	0.006460	0.140220	0.691979	0.161341
4	0.000154	0.002160	0.223225	0.774460

Grade 6 to Grade 7 (n = 127,205 [03-08, 04-09])

6 -> 7	1	2	3	4
1	0.431970	0.489217	0.078338	0.000476
2	0.144862	0.507166	0.346000	0.001972
3	0.008676	0.130613	0.732907	0.127804
4	0	0.001148	0.223209	0.775643

Grade 7 to Grade 8 (n = 188,219 [02-07, 03-08, 04-09])

7 -> 8	1	2	3	4
1	0.355657	0.498689	0.144693	0.000962
2	0.129274	0.455175	0.410338	0.005213
3	0.010934	0.121709	0.702317	0.165040
4	0.000090	0.001411	0.181826	0.816673

Figure 1-2: Pre-2006 and post-2006 one-step transition probability matrices for a typical student

By the Chapman-Kolmogorov equations, the five-step transition probabilities for the pre-2006 cohort are shown in Figure 1-3. These could not be computed for the post-2006 cohort due to the missing third to fourth grade and fourth to fifth grade one-step transition probability matrices. Instead, we compared the three-step transition probabilities for fifth to eighth grade from the pre-2006 cohort to the post-2006 cohort and these results are shown in Figure 1-4.

Pre-2006 Aggregate Cohort Cohorts: 95-00, 96-01, 97-02, 98-03, 99-04, 00-05					
Grade 3 to Grade 8 (n = 338,080 records)					% Proficient
3 -> 8	1	2	3	4	by Grade 8
1	0.035742	0.151668	0.452660	0.359930	81.26
2	0.026119	0.118461	0.407747	0.447673	85.54
3	0.016799	0.082990	0.346086	0.554126	90.02
4	0.008960	0.049754	0.275408	0.665878	94.13

Figure 1-3: Pre-2006 five-step transition probabilities for a typical student

Pre-2006 Aggregate Cohort Cohorts: 95-00, 96-01, 97-02, 98-03, 99-04, 00-05					
Grade 5 to Grade 8 (n = 338,080 records)					% Proficient
5 -> 8	1	2	3	4	by Grade 8
1	0.068787	0.247951	0.513798	0.169464	68.33
2	0.047328	0.192050	0.507139	0.253483	76.06
3	0.023613	0.117206	0.437192	0.421989	85.92
4	0.005230	0.033282	0.238119	0.723369	96.15

Post-2006 Aggregate Cohort Cohorts: 01-06, 02-07, 03-08, 04-09					
Grade 5 to Grade 8 (n > 62,948 records)					% Proficient
5 -> 8	1	2	3	4	by Grade 8
1	0.147516	0.365833	0.427739	0.058913	48.67
2	0.098520	0.288065	0.496909	0.116506	61.34
3	0.038156	0.161924	0.526644	0.273276	79.99
4	0.008603	0.057296	0.364398	0.569703	93.41

Figure 1-4: Pre-2006 and post-2006 three-step transition probability matrices for a typical student

3.1.1 Classification of States

Referring to Figure 1-2, for each grade level in the pre-2006 aggregate cohort, all one-step transition probabilities are non-zero. Thus, the states of the Markov process all communicate with each other. The states of the communicating class are all positive recurrent and aperiodic, thus the Markov chain is ergodic. The same can be said for the fifth to sixth grade and seventh to eighth grade Markov processes in the post-2006 cohort. In the post-2006 sixth to seventh grade Markov process, no students made the transition from

achievement level IV to achievement level I in one step; however, state 1 is accessible from state 4 via states 2 or 3.

3.1.2 Discussion of a Typical Student's Performance Trajectory

For the pre-2006 cohort, the one-step transition probabilities are non-stationary, in that the probability of a transition depends on time period *t*. This finding provides evidence that the acquisition of knowledge is a cumulative process, and that falling behind in mathematics will affect a student's future level of proficiency. As shown in the pre-2006 cohort matrices in Figure 1-2, although the transition probability matrices differ statistically by grade level, they have some similar characteristics. For all grade level transitions, there is only a 1% chance that a student that scores within achievement level I will score within achievement level IV in the following grade. The likelihood improves for a transition from achievement level I to III in all grade levels, with the highest probability occurring in the seventh to eighth grade transition (24.5%) and the lowest occurring in the fourth to fifth grade transition (14.6%). Students who score within achievement level II have a significant chance of improving their score to a level III within one grade level, but are unlikely to score a level IV (~3%). Across all grade level transitions, students who score a level III are likely to score a level III in the following grade level, ranging from 57.1% in the fourth to fifth grade transition to 68.6% in the fifth to sixth. Students who score a level III may improve to a level IV, with the highest chance occurring in the fourth to fifth grade transition at 34.6%. It is possible for a student who scores a level III to regress to a level II in the next grade and this probability increases over time, where 12.8% of eighth grade students will fall out of the "proficient" zone. Students who score within achievement level IV in any grade level are

most likely to remain in achievement level IV in the next grade, thus demonstrating the highest level of proficiency in mathematics. There is a 16.5% chance that a fourth grade student who scores within achievement level IV in third grade will regress to a level III and this probability fluctuates between 9% and 14% for subsequent grade level transitions. It is unlikely (< 1%) that a student who scores a level IV will regress to a level I or II within one grade level.

Referring to Figure 1-2, as is the case in the pre-2006 cohort, the post-2006 cohort one-step transition probability matrices are non-stationary and share some similar characteristics. Students that score a I are most likely to remain within achievement level I in the sixth to seventh grade transition (43.2%). Students that score a II are likely to score a II in the next grade or may advance to achievement level III, with the highest probability of occurrence in eighth grade at 41%. For all grade level transitions, students that score a III are almost equally likely to advance to a IV or regress to a II (12.1%-16.5%), but approximately 70% of the time they will remain at achievement level III in the next grade level. Students that score within achievement level IV in all grade levels are likely to remain in achievement level IV (77.4%-81.6% of the time) but will regress to a III approximately 20% of the time. It is most unlikely (< 1%) that a student who scores a IV will regress to a I or II within one grade level.

The one-step transition probability matrices of the pre-2006 cohort are significantly different from those of the post-2006 cohort for all grade levels. This suggests that the changes made to the scale scores in 2006 had a significant impact on how students were labeled within the North Carolina Accountability System. Upon review of the fifth to sixth

grade transition probability matrices in Figure 1-2, a student that scores within achievement level I in fifth grade will stay within achievement level I in sixth grade with a greater probability in the post-2006 cohort (35.2%) than in the pre-2006 cohort (20.1%); however, both cohorts have a similar probability that a student will advance to achievement level II in sixth grade, with probabilities of 56.0% in the pre-2006 cohort and 56.8% in the post-2006 cohort. In the post-2006 cohort, there is only an 8% chance that a fifth grade student that scores a I will improve their score to a III in sixth grade as opposed to a 23.3% chance of that occurring in the pre-2006 cohort. In the post-2006 cohort, for fifth grade students that score within achievement level II, over half of them will score a II in sixth grade (53.9%), but fewer will advance to a III than in the pre-2006 cohort (34.8% vs. 53.0%). The transition probabilities for fifth grade students that score within achievement level III are similar pre- and post-2006, with a slight increase in the probability of regressing to achievement level II found in the post-2006 cohort (14% vs. 9.3%). Fifth grade students who score a IV are unlikely to transition to a I or II in both the pre- and post-2006 cohorts, with a combined probability of such an event at less than 1% for both analyses. Similar trend statements can be made upon comparing the sixth to seventh grade transitions and the seventh to eighth grade transitions from the pre- and post-2006 cohorts.

3.1.3 Proficiency of a Typical Student by the Eighth Grade

Recall that the state of North Carolina classifies students that score within achievement levels III or IV on the EOG mathematics test as "proficient" in mathematics. The probabilities that a typical student will end up as "proficient" by eighth grade as classified by their initial third grade score is shown in the grey shaded column of Figure 1-3

for the pre-2006 cohort. For example, prior to 2006 a third grade student that scored within achievement level I had an 81.2% chance of advancing to achievement levels III or IV by eighth grade. As shown in Figure 1-4, the changes to the scale scores in 2006 had an impact on the percentage of "proficient" students by eighth grade. A fifth grade student that scores within achievement level I has a 68.3% chance of advancing to achievement levels III or IV by eighth grade in the pre-2006 cohort, but only a 48.6% chance of advancing in the post-2006 cohort. This likelihood increases to 76.1% for pre-2006 fifth grade students that score within achievement level II, and to 61.3% for students in the post-2006 cohort. In both the pre- and post-2006 cohorts, fifth grade students that score within the "proficiency" threshold are unlikely to regress to a I or II by eighth grade. The division between "proficient" and "non-proficient" students is more pronounced in the post-2006 cohort than in the pre-2006 cohort. In the post-2006 cohort, students that start in achievement levels I or II have more difficulty advancing into the "proficient" zone and this difficulty may propagate over time.

To validate these results, Figure 1-5 shows the percentage of "proficient" students on the eighth grade EOG mathematics test for all ten cohorts studied. Note the effects of the changes to the EOG scale scores in 2006. These results are in line with the probabilities of proficiency calculated in Figure 1-3 and Figure 1-4. Table 1-3 shows the state-wide EOG proficiency results reported by the North Carolina DPI for the years included in this study (Public Schools of North Carolina 2011h).

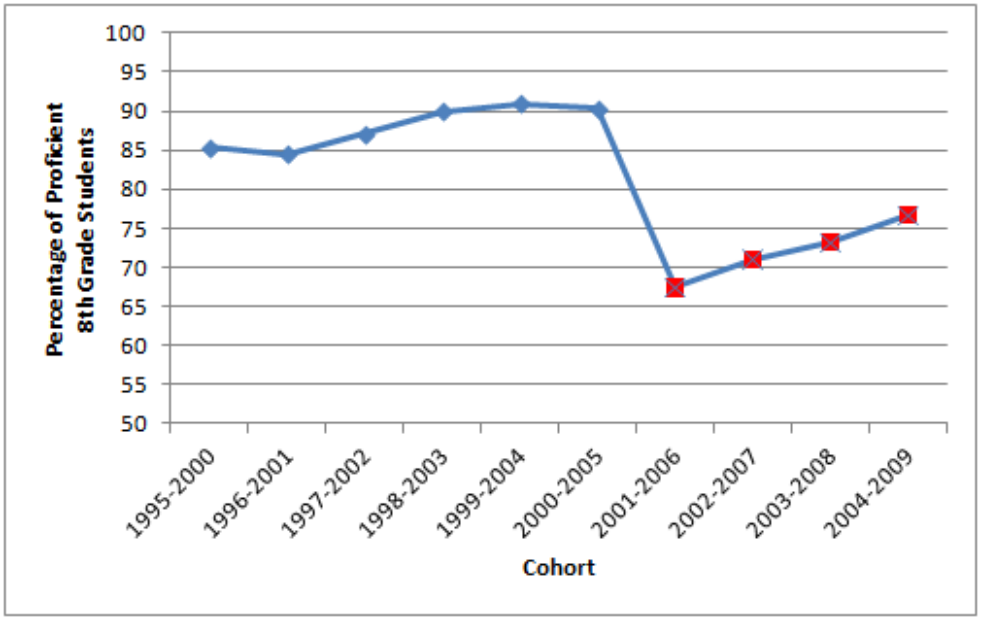


Figure 1-5: Percentage of "proficient" students on the eighth grade EOG mathematics test in the longitudinal datasets created from NCERDC data

Table 1-3: 1995-2010 percentage of "proficient" students on the original administration of the EOG mathematics test by grade level reported by the North Carolina DPI, no retests (Public Schools of North Carolina 2011h)

	Percentage of "Proficient" Students Scoring at Level III or Level IV on the EOG Mathematics Test						8th Grade Difference Between Cohort Result & Corresponding State Reported Result
Academic	Grade Level						
Year	3	4	5	6	7	8	
1995-96	67.4	71.5	70.0	72.6	68.5	67.7	
1996-97	70.2	74.6	73.1	72.7	70.8	68.9	
1997-98	68.2	79.3	78.0	78.3	76.9	76.3	
1998-99	70.0	82.7	82.4	81.1	82.5	77.6	
1999-00	71.8	84.4	82.9	81.0	80.7	80.6	4.72
2000-01	73.6	86.8	86.7	82.9	81.2	79.5	4.99
2001-02	77.3	88.9	88.4	86.4	83.3	82.3	4.83
2002-03	88.9	94.7	92.6	90.0	83.8	84.2	5.73
2003-04	89.0	94.6	93.4	90.0	84.9	85.0	5.98
2004-05	86.1	92.9	90.9	90.2	85.2	84.8	5.48
2005-06	68.8	65.9	63.9	62.5	62.4	61.3	6.10
2006-07	72.2	69.0	68.0	65.7	64.6	66.3	4.73
2007-08	74.8	74.5	71.1	69.4	68.6	69.8	3.36
2008-09	75.3	75.6	73.5	70.5	71.3	72.8	3.83
2009-10	75.8	77.8	75.1	73.8	73.5	77.7	

We present a few theories regarding why these results, in general, appear to be favorable. First, the longitudinal datasets we created from the NCERDC data contain students that could be tracked from third grade to eighth grade. Students retained in a grade were dropped as their respective identification numbers would not appear in the subsequent grade level (i.e., a student retained in third grade in 2000 would not appear in the fourth grade dataset from 2001). As a result, our longitudinal datasets could exclude a batch of potentially poor scores and admittedly only the "best of the best" students would appear in the eighth grade dataset. Upon review of the "Difference" column in Table 1-3, the effect of this limitation of our model can be assessed. The largest difference in the percentage of eighth grade proficiency between our longitudinal datasets and the state reported results occurs in 2005-06, with a delta of 6.1%. This could be related to the change in the test that occurred in that year. In total, the average potential overstatement of proficiency by our datasets is 4.98%. In relation to the preceding discussion regarding the exclusion of retained students, we argue that 4.98% is within a margin of error that is sufficient for us to conclude that the longitudinal datasets and subsequent model can produce meaningful results.

Another theory to explain the favorability of these results is with regard to the assessment instrument itself, the EOG test. In reference to Figure 1-5, notice how the total percentage of proficient eighth grade students increases over time prior to the 2006 changes to the EOG scale scores and again after the 2006 change. This could be a result of "exam creep" in that the EOG test is overstating the number of students that are proficient in mathematics. Incidentally, it may be that the state of North Carolina's definition of "proficiency" overestimates the actual capabilities of students in the public school system

with respect to preparation for more advanced mathematics. These results contrast with national reports of eighth grade proficiency in mathematics (National Center for Education Statistics 2006, National Research Council 2011). Or perhaps, compared to other states, North Carolina is doing a better job of teaching mathematics.

3.2 Markov Chains Based on Demographic Factors

The descriptive statistical analysis previously discussed was performed on the 2000-2005 cohort, thus the following Markov chains were calculated using the longitudinal dataset from that cohort.

3.2.1 Gender

The one-step transition probability matrices by gender for the 2000-2005 cohort are shown in Appendix E. The five-step transition probabilities are shown in Figure 1-6.

2000-2005 Cohort (n = 63,227 total records): Gender n-step Probabilities - Grade 3 to Grade 8

FEMALE (n_F = 33,855 records)						MALE (n_M = 29,372 records)					
3 -> 8	1	2	3	4	% Proficient by Grade 8	3 -> 8	1	2	3	4	% Proficient by Grade 8
1	0.023499	0.115769	0.443501	0.417231	86.07	1	0.030283	0.123309	0.403312	0.443096	84.64
2	0.017286	0.090727	0.391300	0.500687	89.20	2	0.022785	0.097176	0.358176	0.521862	88.00
3	0.012393	0.069740	0.343642	0.574225	91.79	3	0.016751	0.075273	0.316536	0.591440	90.80
4	0.006726	0.042717	0.274409	0.676149	95.06	4	0.009471	0.047275	0.256703	0.686551	94.33

Figure 1-6: 2000-2005 cohort – Five-step transition probability matrices by gender

From a practical standpoint, the female and male one-step transition probability matrices are similar. A few differences worth noting occur in the one-step transitions from achievement level I. Female students that score within achievement level I in sixth grade are more likely to advance to achievement level II in seventh grade than their male counterparts (57.7% vs. 50.0%), but are less likely to advance to achievement level III (18.3% vs. 30.5%). The

opposite is true for the transition from achievement level I in the seventh to eighth grade. In seventh grade, female students that score within achievement level I are less likely to advance to a II in eighth grade than male students (40.7% vs. 51.5%), but are more inclined to advance to a III (30.9% vs. 22.9%). The five-step transition probabilities in Figure 1-6 reveal that female and male students are practically equally likely to end up as "proficient" by eighth grade given each initial third grade score.

3.2.2 Ethnicity

The one-step transition probability matrices by ethnicity for the 2000-2005 cohort are shown in Appendix E. The five-step transition probabilities are shown in Figure 1-7.

2000-2005 Cohort (n = 63,227 total records): Ethnicity n-step Probabilities - Grade 3 to Grade 8

% Proficient by Grade 8 (shaded)

ASIAN [A] (n_A = 1,080 records)

3 -> 8	1	2	3	4	%
1	0.012099	0.059872	0.303171	0.624858	92.80
2	0.008767	0.042856	0.244567	0.703810	94.84
3	0.006811	0.032983	0.209612	0.750594	96.02
4	0.003356	0.015989	0.145585	0.835070	98.07

BLACK [B] (n_B = 17,874 records)

3 -> 8	1	2	3	4	%
1	0.041710	0.178827	0.509446	0.270016	77.95
2	0.034462	0.155678	0.487417	0.322443	80.99
3	0.028268	0.134726	0.464764	0.372242	83.70
4	0.019883	0.103375	0.424301	0.452441	87.67

HISPANIC [H] (n_H = 2,011 records)

3 -> 8	1	2	3	4	%
1	0.018636	0.110289	0.440718	0.430358	87.11
2	0.014437	0.090085	0.405671	0.489807	89.55
3	0.010851	0.072120	0.371415	0.545614	91.70
4	0.006911	0.050558	0.322719	0.619811	94.25

AMERICAN INDIAN [I] (n_I = 973 records)

3 -> 8	1	2	3	4	%
1	0.024357	0.138614	0.473747	0.363282	83.70
2	0.020265	0.118925	0.451049	0.409760	86.08
3	0.016516	0.100319	0.427347	0.455818	88.32
4	0.012189	0.077889	0.395035	0.514887	90.99

MULTI-RACIAL [M] (n_M = 1,163 records)

3 -> 8	1	2	3	4	%
1	0.025142	0.102544	0.477122	0.395192	87.23
2	0.016264	0.075013	0.392673	0.516051	90.87
3	0.012123	0.061379	0.349352	0.577147	92.65
4	0.006925	0.042510	0.286262	0.664303	95.06

WHITE [W] (n_W = 40,126 records)

3 -> 8	1	2	3	4	%
1	0.013200	0.067199	0.351501	0.568099	91.96
2	0.009978	0.052888	0.308073	0.629061	93.71
3	0.007287	0.040484	0.267765	0.684464	95.22
4	0.004105	0.024989	0.212637	0.758269	97.09

Figure 1-7: 2000-2005 cohort – Five-step transition probability matrices by ethnicity

Upon review of the one-step transition probabilities, a number of conclusions can be made about the differences in the probabilistic movements of students from the various ethnic groups. We present a few results here. With a few exceptions, Black students that score within achievement levels I or II are more likely to stay within these respective achievement levels in the following grade than all other ethnicities. Although the sample sizes are

somewhat small, Multi-Racial students have the highest probability of advancing from a II to a III in middle school (60.5% in sixth to seventh grade and 60.0% in seventh to eighth grade). For all ethnicities, the highest probability of regressing from a III to a II occurs in the sixth to seventh grade transition, where Black students have the highest probability at 19.4% and Asian students have the lowest probability at 7.8%. Regressing from a III or IV to a I is a rare occurrence (< 1.5%) for all ethnic groups, but is higher than the "typical student" results. Across all grade level transitions, Asian students that score a IV are most likely to remain a IV in the following grade. The five-step transition probabilities in Figure 1-7 show that Black and American Indian students that score within achievement level I in third grade have the lowest probabilities of advancing to levels III or IV by eighth grade at 77.9% and 83.7%, respectively. Asian and White students that score within achievement levels I or II in third grade have the highest probabilities of advancing to "proficiency" by eighth grade. For all third grade scores, Asian and White students have practically the same probabilities of advancing to "proficiency" by eighth grade. There is also parity between Multi-Racial and Hispanic students.

3.2.3 Participation in a School Lunch Program

The one-step transition probability matrices by participation in a school lunch assistance program for the 2000-2005 cohort are shown in Appendix E. The five-step transition probabilities are shown in Figure 1-8.

2000-2005 Cohort (n = 63,227 total records): School Lunch Program n-step Probabilities - Grade 3 to Grade 8

% Proficient by Grade 8 (shaded)

FREE LUNCH (n_{FL} = 18,460 records)

3 -> 8	1	2	3	4	%
1	0.041582	0.170832	0.500818	0.286768	78.76
2	0.034063	0.148348	0.478868	0.338721	81.76
3	0.027903	0.128787	0.457265	0.386046	84.33
4	0.019764	0.099790	0.418707	0.461738	88.04

REDUCED PRICE LUNCH (n_{RL} = 5,138 records)

3 -> 8	1	2	3	4	%
1	0.019877	0.119775	0.459970	0.400378	86.03
2	0.017001	0.103046	0.431937	0.448016	88.00
3	0.014067	0.086023	0.401428	0.498482	89.99
4	0.009859	0.061761	0.352213	0.576167	92.84

FULL PAY/SCHOOL NOT PARTICIPATING (n_{FP} = 39,629 records)

3 -> 8	1	2	3	4	%
1	0.014213	0.074014	0.364960	0.546814	91.18
2	0.009861	0.054254	0.306713	0.629173	93.59
3	0.006841	0.039852	0.260510	0.692797	95.33
4	0.003689	0.023832	0.203982	0.768497	97.25

Figure 1-8: 2000-2005 cohort – Five-step transition probability matrices by participation in a school lunch program

From the statistical analysis, we know that students in the 2000-2005 cohort that participated in the free or reduced price lunch program scored lower on average than those students that did not participate in a school lunch assistance program. These results are confirmed by the one-step transition probability matrices. For example, in the third to fourth grade transition, 31.2% of free lunch students score within achievement levels I or II in third grade as opposed to only 11.1% of full pay students. Across all grade level transitions, free lunch students that score within achievement level II are the most likely to stay in achievement level II in the following grade. Free lunch students that score a III are less likely to advance to a IV than full pay students, and are more likely to regress to a II. The likelihood of a free lunch student regressing from a III to a II roughly increases over time, with the highest instances occurring at the sixth to seventh (18.0%) and seventh to eighth grade transitions (16.7%). Comparatively, full pay students are nearly half as likely to regress from a III to a II in the transition from sixth to seventh grade, at 9.5%. The five-step transition probabilities in Figure 1-8 also reveal that free lunch students are less likely to be "proficient" by eighth grade. Note that these are pre-2006 scores so these disparities are likely to be worse following the scoring changes made in 2006.

3.2.4 AIG Mathematics Students

The one-step transition probability matrices by AIG mathematics status for the 2000-2005 cohort are shown in Appendix E. The five-step transition probabilities for non-AIG mathematics students are shown in Figure 1-9. Note, we were unable to calculate the five-step transition probabilities for the AIG mathematics students. As shown in Figure E-4, no AIG mathematics students scored within achievement level I in grades 4-7 and no AIG mathematics students scored within achievement level II in grades 5-6. As a result, transitions from these levels could not be tracked and subsequent calculations of the *n*-step transition probabilities could not be computed.

2000-2005 Cohort (n = 63,227 total records): AIG Mathematics n-step Probabilities - Grade 3 to Grade 8

NOT AIG MATH (n_{NOT} = 52,282 records)					% Proficient
3 -> 8	1	2	3	4	by Grade 8
1	0.027357	0.123636	0.454222	0.394784	84.90
2	0.020859	0.099935	0.415507	0.463698	87.92
3	0.015920	0.081032	0.381583	0.521465	90.30
4	0.010809	0.059773	0.337878	0.591541	92.94

Figure 1-9: 2000-2005 cohort – Five-step transition probability matrices by AIG mathematics status

Although it is difficult to draw conclusions from the results of the AIG mathematics students, it is safe to assume that very few students that are classified as AIG in mathematics will score a I or a II on the EOG mathematics test beyond third grade. In fact, upon inspection of the frequency counts, the majority of AIG students score within achievement level IV on the test each year. For each grade level transition in the one-step transition probability matrices, AIG

students that score a IV never regress to a I or II. They regress to a III approximately 1% of the time, with the exception of the third to fourth grade transition where a retreat to achievement level III occurs for 6.4% of AIG students. Non-AIG students from the 2000-2005 cohort follow similar trends to those students included in the pre-2006 aggregate cohort. Figure 1-9 confirms that most non-AIG students will be "proficient" by the eighth grade.

4. Conclusions

4.1 Summary

In summary, to represent student performance in mathematics over time, we: (1) longitudinally track EOG mathematics test achievement level scores of various cohorts of students from third to eighth grade and probabilistically characterize the movement of students' scores from one grade level to the next; (2) conduct a descriptive statistical analysis of one cohort of students to identify demographic factors that affect performance; and (3) develop Markovian models of student performance and use these models to estimate a student's proficiency in mathematics by eighth grade.

From the statistical analysis we find that performance on the EOG mathematics test differs significantly by a student's ethnicity, AIG mathematics classification, and socioeconomic status, as indicated by their participation in a school lunch assistance program. Although statistical differences in female and male achievement level scores exist, practically their assessment outcomes are the same. This is confirmed by the gender specific transition probability matrices of female and male students.

The aggregate Markov chains offer a broad viewpoint of a typical student's performance on the EOG mathematics test. The transition probability matrices of the pre- and post-2006 cohorts are significantly different, highlighting the impacts of the changes made to the EOG mathematics test scale scores in 2006. For both the pre- and post-2006 aggregate cohorts, the one-step transition probabilities are non-stationary indicating that performance in mathematics depends on the grade level. This result supports the belief that the acquisition of mathematical knowledge is a cumulative and sequential process. A solid foundation in mathematics in early education is critical to a student's success in higher level science and mathematics coursework. The Markov chains based on the demographic factors provide a more in depth view of the performance of one cohort of students, but exhibit similar overall trends to those of the aggregate model. For all demographic factors considered, we observe that the probability of falling out of the "proficiency" threshold is the highest in the sixth to seventh and seventh to eighth grade transitions. This result emphasizes the importance of a student's transition from general arithmetic to algebraic thinking in middle school. Surmounting this gateway is critical to a student's successful trajectory along the STEM pipeline (Loveless 2008).

The Markov chains presented herein provide new insights into modeling student performance that existing quantitative and qualitative models do not. In this paper, we model elementary and middle school students' performance in mathematics as a stochastic process to evaluate their proficiency in the discipline. Previous studies involving North Carolina EOG assessment data have not examined transitions between achievement level scores in this way. Value-added models provide a static representation of student performance based on

regression techniques, whereas the Markov chain model lays the foundation for future work where we can dynamically optimize decisions about an individual student's education. Education policy makers can use the transition probability matrices of the Markov chain to further their understanding of a student's progression of performance in mathematics over time, enabling them to forecast individual elementary students' success in higher level mathematics coursework. Many research studies link high school performance to STEM enrollment in college; we characterize the evolution of performance in mathematics in primary education, a critical time period where a student's foundation of mathematical understanding is developed. We propose that the results of the Markov chain model could inform existing models that link high school course taking to STEM degree attainment.

4.2 Limitations and Future Work

This research has raised a number of questions that warrant further study. Using the NCERDC EOC exam scores, we plan to link students' scores in Algebra I, Algebra II, and Geometry to their EOG results to evaluate if EOG "proficiency" equates to success in higher level mathematics courses. This could serve to validate the state of North Carolina's definition of "proficiency." We acknowledge that excluding students that were retained in a grade could bias the results. By tracking students that were retained in a grade separately, we will assess how retention impacts future performance on the EOG mathematics test.

The Markov chain models presented here describe a student's performance in mathematics over time as measured by the standardized EOG mathematics test in North Carolina. Although standardized testing is used across the country to establish accountability standards for public education, critics admonish the propensity for educators to have to

"teach to the test" without the flexibility to cover select topics in greater depth (K12 Academics 2011). We have assumed that performance on a standardized test is an indicator of a student's cognition and learning. In reality, a student's internal intellectual condition can never be known with certainty and a standardized test may not be the best method for ascertaining information about a student's true state of knowledge. We plan to extend this work by introducing control features to develop Markov decision process models where the states are completely observable and partially observable (Howard 1960, Monahan 1982). We recommend that these types of models be used as a baseline by which to measure the effectiveness of various curricular and teaching interventions in North Carolina. In addition, they can be used to compare the results of EVAAS decisions (i.e., enrolling students in Algebra I with a 70 percent EVAAS calculated probability of passing) to the optimal decision strategy (Hui & Goldsmith 2012).

Acknowledgements

The authors especially thank the NCERDC for granting us access to their extensive and detailed datasets. Special thanks to Dr. James Wilson for his guidance on the statistical analysis and for directing us to the references pertaining to the skewness-adjusted confidence intervals. Also, special thanks to Dr. Rebecca Brent for her valuable feedback on previous drafts.

This project has been approved by NC State University Institutional Review Board for the Use of Human Subjects in Research, IRB Approval Number 1845-11-2.

Chapter 2 - Characterizing Student Performance in Mathematics: An Application of a Markov Decision Process to an Education System

1. Introduction

Transforming Science, Technology, Engineering, and Mathematics (STEM) education in the United States K-12 system is at the forefront of the education reform conversation (National Research Council 2011). In the United States, occupations in STEM fields are expected to increase by 17.0% from 2008 to 2018, as compared to 9.8 percent in non-STEM professions (Economics & Statistics Administration 2011). Education reformers cite the importance of "constructing a strong, coordinated STEM education system" to prepare students for STEM careers (National Science Board 2007, pg. 1). Of particular emphasis is the need to 'vertically' align STEM learning such that young students build a solid foundation of mathematical knowledge (National Science Board 2007, Common Core Standards 2010).

To examine a method for optimizing the vertical alignment of STEM learning, we model the sequential process of learning mathematics over time and use a Markov Decision Process (MDP) to identify optimal intervention strategies. Specifically, we model student performance on the North Carolina End-of-Grade (EOG) mathematics test, administered annually to public school students in grades 3-8, and evaluate the utility of administering retests of the exam. The underlying Markov chain of the MDP characterizes the evolution of performance in mathematics in primary education when a student's foundation of mathematical understanding is developed. The control features of the MDP offer an optimal decision policy for administering the retests. The model provides a framework for making

sequential decisions about an individual student's education and has the potential to assist education administrators in optimally allocating valuable resources, such as teacher labor. Although the purpose of this study is to recommend an optimal decision structure for the administration of retests on a standardized exam, the framework presented can be used to assess other educational interventions, such as curriculum changes or the impacts of remedial efforts.

1.1 Background

1.1.1 United States STEM Education System

Many reports have documented deficiencies in the current United States K-12 STEM education system. The National Academies Press' 2007 *Rising Above the Gathering Storm* report cites the following statistics to motivate this research, "Fewer than one-third of US eighth grade students performed at or above a level called 'proficient' in mathematics…Alarmingly, about one-fifth of the fourth graders and one-third of the eighth graders lacked the competence to perform even basic mathematical computations" (National Center for Education Statistics 2006). Citing Schmidt (2011), the National Research Council states that "three quarters of eighth graders still enter high school not having reached the proficient level" (National Research Council 2011, pg. 19). In an effort to advance students' proficiency in STEM disciplines, the federal government spent over $3.4 billion in STEM education initiatives in 2010, which does not account for state and local STEM education expenditures (National Science and Technology Council 2011). Although mathematics is one pillar of STEM education, it is arguably the most important because its concepts and

methodologies pervade all science, technology, and engineering disciplines (National Research Council 2001).

In the National Science Board's 2007 action plan for STEM education, the Board identified two specific challenges that need to be addressed (National Science Board 2007). First, STEM education is not aligned "horizontally," i.e., mathematics standards and course sequences vary from state to state. In this regard, North Carolina recently adopted the Common Core State Standards, a set of mathematics standards that emphasize a consistent progression of learning across grade levels (Common Core Standards 2010). Second, there is "little or no (vertical) alignment of STEM learning during students' progression through school. Students do not always obtain a mastery of key concepts at the elementary and middle school levels, thus limiting academic success at the high school level" (National Science Board 2007, pg. 5). Mathematics, in particular, is learned incrementally and concepts are highly sequential in nature (President's Council of Advisors on Science and Technology 2010, pg. 17). As many mathematics educators will attest, learning interventions in high school are far too late to impact a student's mathematics performance trajectory (National Research Council 2001). As a result, the focus of this paper is on mathematics competency in the elementary and middle grades.

1.1.2 The North Carolina Public School System

The North Carolina public school system consists of 115 school districts serving nearly 1.4 million students per year and is home to the 16th and 18th largest public school districts in the country (Public Schools of North Carolina 2009a, Wake County Public School System 2011, Proximity 2011). Integral to the state's assessment and accountability system

are the state-mandated EOG tests in mathematics and reading comprehension. These standardized, multiple choice exams are administered annually in grades 3-8 and "are designed to measure student performance on the goals, objectives and grade-level competencies" outlined in the statewide K-12 curriculum (Public Schools of North Carolina 2011b, 2011g). A student's raw score on an EOG test is converted into a scale score which falls into one of four categories known as *achievement levels*. A description of the respective achievement levels was presented in Table 1-1. The achievement levels provide a metric for comparing the relative performance of North Carolina public school students. A student whose scale score falls within achievement levels III or IV is defined as "proficient" by the state, which means that the student consistently performs at or above grade-level expectations (Public Schools of North Carolina 2006a). We refer the reader to the following resources for additional information on the EOG test development process and data on the test's reliability and validity: Sanford (1996), Public Schools of North Carolina (1999), Public Schools of North Carolina (2011f), and Public Schools of North Carolina (2011g).

Since 2001 the North Carolina Department of Public Instruction (DPI) has considered grades 3, 5, and 8 as "gateways" for promotion decisions (Public Schools of North Carolina 2006d). Students in these grade levels were required to score at or above achievement level III on the EOG mathematics test to demonstrate grade level proficiency. If they failed to do so, various "safeguards" including retesting and/or other formal evaluation procedures were implemented (Public Schools of North Carolina 2006a). In 2009, the State Board of Education implemented a new policy on retesting. All public school students in grades 3-8 that score within achievement level II on the EOG mathematics test must be administered a

retest, and those students that score within achievement level I may opt-in to retest (Public Schools of North Carolina 2009b). As a result, over 24,000 third grade students (~22%) were administered at least one mathematics retest in 2009 compared to only 13,000 (~13%) in 2008. In this study we consider the administration of a retest as a control action in the MDP; allowing us to evaluate the policy change enacted in 2009.

1.2 Objectives

The objectives of the research conducted in this chapter are to:

- Introduce an optimal control system, an MDP, in the context of primary education and describe the benefits of this approach relative to other modeling techniques.
- Develop an MDP model of student performance on the EOG mathematics test to determine an optimal policy for the administration of retests and investigate the structural properties of the model.
- Conduct a preliminary sensitivity analysis to explore the impacts of modifications to the cost structure of the MDP model.

The remainder of this paper is organized as follows. In the rest of Section 1 we present a review of relevant literature. We describe the formulation of the MDP model in Section 2 and present a numerical example of the model using data from the North Carolina public school system in Section 3. In Section 4 we provide a set of optimal decision policies, a preliminary examination of the structural properties of the model, and the results of the sensitivity analysis on the cost estimates. We discuss conclusions, the limitations of the model, and areas of future work in Section 5.

1.3 Relevant Literature

Much of the literature on learning theory and learning models comes from research conducted in psychology and education. Many of these models are conceptual (Simon 1995, Simon & Tzur 2004, Confrey et al. 2009) or statistical and thus purely descriptive (Tyson et al. 2007, McCaffrey et al. 2004, Tekwe et al. 2004, Ballou et al. 2004). Many of the findings rely on anecdotal evidence based on small sample sizes, making it difficult to generalize the results to a larger population (Perdikaris 1992, Voskoglou 2009). In this review, we explore the application of Markovian models to education systems.

1.3.1 Statistical Models of Education Systems

As a result of the No Child Left Behind (NCLB) Act of 2001, states are required to administer annual standardized tests to show that students are making 'Adequate Yearly Progress' toward the state's performance goals (US Department of Education 2012). In response to NCLB and given the increased availability of longitudinal test data, researchers have developed statistical models to track student progress on standardized tests (Raudenbush & Byrk 2002, McCaffrey et al. 2004, Tekwe et al. 2004, Ballou et al. 2004). The North Carolina DPI has adopted the Education Value-Added Assessment System (EVAAS), a statistical modeling package available from SAS, to measure the impacts of a student's education on their performance outcomes (SAS EVAAS 2012, Public Schools of North Carolina 2012a). EVAAS longitudinally tracks individual student performance on EOG tests and measures the student's year-to-year gain in knowledge (Wright et al. 2010). EVAAS uses several statistical modeling techniques; the one most relevant to this research is the Analysis of Covariance (ANCOVA) model used to predict student success probabilities

on future academic milestones, such as future EOG tests or the SAT (Wright et al. 2010). In comparison, in Chapter 1 we used Markov chain models to forecast a student's likelihood of success on the EOG test in eighth grade given their level of achievement on the exam in the third grade. An ANCOVA is essentially a regression analysis, thus, the model's ability to forecast behavior beyond the next immediate time period is limited. The MDP model, in contrast, will provide long-term estimates of performance probabilities. The model can also serve to evaluate standards set as a result of EVAAS predictions, such as Wake County's decision to place students in Algebra I in middle school if EVAAS indicates a 70% probability of their success, and identify optimal intervention strategies (Hui & Goldsmith 2012). In general, EVAAS and other value-added models provide a static representation of student performance based on regression techniques, whereas the MDP will potentially allow us to dynamically optimize decisions regarding a student's education.

1.3.2 Markov Chains in Education Modeling

Upon review of the existing literature, we found few examples of the use of Markov chains to model teaching and learning in K-12 education. Perdikaris (1992) used a Markov chain to model mathematics teachers' decision making and their process of inquiry in the classroom. The focus of Perdikaris (1992) was modeling the behavior of two teachers whereas our objective is to model student performance. McFarland (2006) used a Markov model to analyze the flow patterns of mathematics course taking in two high schools; he did not consider a student's mathematical knowledge collected prior to entering high school, which is the focus of this paper. Markov chains have been used to model various aspects of teaching and learning theory at the undergraduate level. Some examples include Shah &

Burke (1999), Klingbeil et al. (2005), Fam et al. (2006), Pinar-Yildirim et al. (2007), and Voskoglou (2009).

To the best of our knowledge, Markov chain models of the primary education system (K-8) do not exist. Three challenges persist upon review of existing Markov chain literature as applied to education systems: defining mutually exclusive states, describing how the states change over time according to the time-dependent probability distributions, and ensuring access to statistically significant sample sizes. Furthermore, Markov chains do not offer decision makers the control features inherent to MDPs. We aim to extend this stream of research by not only creating a Markov chain model of the primary education system, but also by incorporating actions and rewards into the model.

1.3.3 Applications of Markov Decision Process Models

MDP models have been applied in a variety of research areas such as machine maintenance, machine learning, behavioral ecology, marketing, and medical decision making (Smallwood & Sondik 1973, Rosenfeld 1976, Jaakkola et al. 1995, Littman et al. 1995, Cassandra et al. 1996, Hu et al. 1996, Hauskrecht & Fraser 2000, Alagoz et al. 2004, Maillart et al. 2008, Chhatwal et al. 2010). Chhatwal et al. (2010) used an MDP to determine an optimal policy for making breast biopsy decisions, where the states of the model were described by a patient's breast cancer risk score. Although the context is different, we similarly map a student's test score in one time period to their score in the next and overlay control. The aforementioned papers provide examples of sequential decision making models in uncertain environments; however their state space is tied to a physical process (a machine) or a human system in the case of medical decision making models. Our model will introduce

an added layer of complexity by capturing human cognition in the state space, namely how a student performs in mathematics over time. To the best of our knowledge MDPs have not been applied to K-12 education systems.

1.3.4 Summary of Relevant Literature Reviewed

Models of the undergraduate engineering education system have emerged in the past decade (Shah & Burke 1999, Klingbeil et al. 2005, Pinar-Yildirim et al. 2007); however it appears that Markovian models of the K-12 education system do not exist. In our review of prior research, we have found no evidence of MDP models in modeling consecutive years of a child's learning process. Prior works have used Markov chains to describe the progression of learning but have not tried to control this progression optimally (Stevens et al. 2005, Voskoglou 2009). Although rigorous formulations of MDP models exist in the domains of machine maintenance, machine learning, and medical decision making, none of these involve an analysis of human cognition. We acknowledge the inherent challenges in applying this class of models to the complex issues found in education systems, but advocate that quantitative models for measuring student performance are necessary and lacking. To the best of our knowledge, this is one of the first large-scale stochastic optimization models employed in K-12 education research.

2. Methods

2.1 Markov Decision Process Formulation

We formulate a discrete-time, finite horizon MDP to characterize student performance on the EOG mathematics test over time. The underlying Markov chain is a discrete-time stochastic process that satisfies the Markov property. A stochastic process with

the Markov property implies that a future state depends only on the present state of the system and not on past states (Howard 1960). This is consistent with the assumptions about the process of learning outlined in the learning theory literature that says that a student's current state of cognition is independent of their collection of past knowledge (Smallwood 1971, Simon & Tzur 2004, Confrey & Maloney 2010). This is also consistent with North Carolina DPI testing protocol that states that students at higher achievement levels are assumed to have mastered the competencies and goals associated with the lower achievement levels (Public Schools of North Carolina 2006b). Further, a student's past performances on the EOG tests beyond the previous grade level are not used to evaluate their proficiency at the current grade level. Consequently, we assume the model satisfies the Markov property and we assume a Markovian order of one.

The state-transition diagram of the MDP model is shown in Figure 2-1. A description of the model notation follows.

Decision Epochs - The decision maker (the educator) makes a decision to either retest the student or to forgo an intervention after the original administration of the EOG mathematics test in grade level t, where $t \in T = \{0, 1, 2, 3, 4\}$ and $t = 0$ represents the third grade.

State Space - The states of the system correspond to a student's EOG mathematics test achievement level score at the end of grade t. Define s_t as the state of the system at the end of grade t and the state space $S = \{\text{I, II, III, IV}\} \sim \{1, 2, 3, 4\}$, where $s_t \in S$.

Actions - The actions represent the possible decisions available when the state of the system is $s_t \in S$. In states $s_t = 1$ and $s_t = 2$ for all t, the decision maker may select a control action from the set $A_{s_t} = \{a = 1 \text{ (Do Nothing)}, a = 2 \text{ (Retest)}\}$.

Transition Probabilities - The transition probabilities, $P_t(a) = [p_t(s_{t+1}|s_t, a)]$, are the set of conditional probabilities that specify the likelihood of a transition from state s_t to state s_{t+1} for action a, where $\sum_{s_{t+1} \in S} p_t(s_{t+1}|s_t, a) = 1 \; \forall \; t, \forall \; a$. Longitudinal EOG mathematics test achievement level scores were used to calculate the various probabilities of movement from one achievement level to another for each action alternative. In general, let $p_t(s_{t+1}|s_t) = n_t(s_t, \; s_{t+1})/n_t(s_t)$, where $n_t(s_t, \; s_{t+1})$ represents the number of students in state s_t at grade level t and state s_{t+1} at grade level $t + 1$, and $n_t(s_t)$, represents the total number of students in state s_t in grade level t. There is evidence of movement between all states in the data for each action alternative; therefore, we consider all possible movements without constraint.

Rewards - The decision maker receives a reward as a result of choosing action $a \in A_{s_t}$ in state s_t. Rewards can take on a positive value to represent an actual reward/gain or a negative value to symbolize a cost incurred to the system. There are two types of rewards considered in this analysis. Define x as the cost of retaining a student in a grade level for one year, and define y as the fixed cost of administering a retest to a student. Instead of using these costs to calculate individual rewards earned by the system for each transition from state s_t to state s_{t+1} for each action a (i.e., $R_t(a) = [r_t(s_t, a, s_{t+1})]$), let $q_t(s_t, a)$ represent the expected immediate reward earned in state s_t for action a. The calculations of $q_t(s_t, a)$ are described in Section 2.2.

The optimal decision policy can be found by solving the set of recursive equations that represent the maximum total expected reward (Howard 1960, Puterman 1994):

$$v_{t+1}(s_t) = \max_{a \in A_{s_t}} \{q_t(s_t, a) + \sum_{s_{t+1} \in S} p_t(s_{t+1}|s_t, a) \cdot v_t(s_{t+1})\}$$

The objective function of this model seeks to maximize the expected benefit to the education system whilst considering individual student's advancement to higher achievement levels on the EOG mathematics test. In this basic framework, we consider the rewards associated with the control actions (in this case negative rewards) and not the rewards associated with a state. Future work will examine both, adapting the framework such that interventions can be selected to advance an individual student's level of achievement.

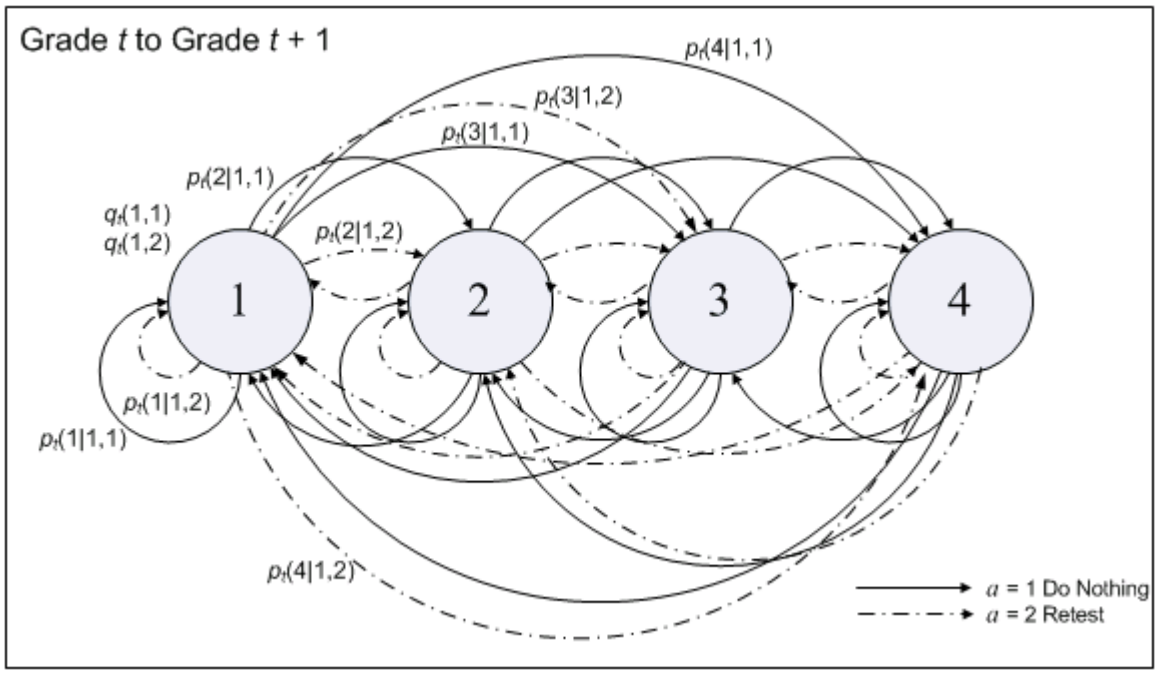


Figure 2-1: MDP state-transition diagram for each action (Although all transition arcs are shown, as an example, only transition probabilities out of state 1 are labeled in the figure)

The following additional model assumptions are made. The probability of a transition depends on time period *t*. Thus, the one-step transition probabilities are, by definition, non-stationary with respect to grade level. This result supports the belief that the acquisition of

mathematical knowledge is a sequential process and that falling behind in mathematics will affect a student's future level of proficiency (President's Council of Advisors on Science and Technology 2010). We assume that the actions are state dependent since it would not be reasonable to administer a retest to a 'proficient' student that scores within achievement levels III or IV on the EOG test.

2.2 Derivation of the Expected Immediate Rewards

To calculate the expected costs, $q_t(s_t, a)$ for $s_t = 1, 2 \; \forall \; t, \forall \; a$, we first calculated the percentages of students that score within achievement levels I or II on the original administration of the EOG mathematics test in each grade level. These 'non-proficient' students could potentially be at risk of being retained in their current grade and would certainly be candidates for intervention. For the sake of subsequent calculations, let o_t represent a student's score on the *original* administration of the EOG mathematics test in grade level *t*. For example, $P(o_t = I)$ represents the probability of scoring within achievement level I on the original administration of the EOG mathematics test in grade *t*.

The aggregate rates of non-promotion for each grade level are reported annually by DPI; however, they are not classified by EOG test achievement level performance. Thus, we estimate a percentage of level I students that are not promoted to the next grade level and a percentage of level II students that are not promoted. For example, suppose that 90% of level I students and 20% of level II students are not promoted to the next grade. Let ρ represent the event that a student is retained in their current grade level for one year. For the 'do nothing' action ($a = 1$), we define $P(\rho|o_t = I)$) and $P(\rho|o_t = II)$ to represent the respective probabilities that a student is retained in their current grade level given they score a I or II,

respectively, on the original administration of the EOG mathematics test. For the 'retest' action ($a = 2$), we first define r_t as a student's score on the first *retest* of the EOG mathematics test in grade level t. We define $P(\rho|r_t = I)$ and $P(\rho|r_t = II)$ to represent the respective probabilities that a student is retained in their current grade level given they score a level I or II, respectively, on the retest of the EOG exam. We assume these probabilities are consistent across all grade levels.

The expected cost for states $s_t = 1$ and $s_t = 2$ for the 'do nothing' action ($a = 1$) are comprised of the expected cost of retaining a student in a grade level and are calculated as follows (Recall, x represents the cost of retaining a student in a grade level for one year):

$$q_t(1,1) = (x)[(P(\rho|o_t = I) \cdot P(o_t = I)] \; \forall \; t$$

$$q_t(2,1) = (x)[(P(\rho|o_t = II) \cdot P(o_t = II)] \; \forall \; t$$

Essentially, these values represent the opportunity cost or loss for deciding to forgo an intervention and not promote specified percentages of students that score a level I or II on the original administration of the EOG mathematics test. We let $q_t(3,1) = 0$ and $q_t(4,1) = 0$ since we assume that 'proficient' students are always promoted to the next grade level.

The expected costs for states $s_t = 1$ and $s_t = 2$ for the 'retest' action ($a = 2$) are a function of the cost of retaining a student in a grade level plus the fixed cost of administering a retest to a student. To calculate these values, we first computed a set of conditional probabilities to represent a student's transition from their original score on the EOG mathematics test (o_t) to their score on the first retest (r_t) administered within their current grade level. The retest scores are essentially a set of observations collected within one time period that can update our belief about a student's level of cognition. In essence, the state of

the system changes because of the action to retest *then* the time period transition occurs. We use these intermediary probabilities to estimate the percentage of retested students that remain 'non-proficient' after the retest intervention. For state $s_t = 1$, we consider students that score within achievement level I on the retest. This includes students that regress from a level II on their first attempt of the exam to level I on the retest, i.e., $P(r_t = I|o_t = II)$, as well as students that fail to progress from level I on the retest, i.e., $P(r_t = I|o_t = I)$. As a result, from the law of total probability, the expected cost for state $s_t = 1$ for the 'retest' action ($a = 2$) is:

$$q_t(1,2) = (x)\big[\big(P(r_t = I|o_t = I) \cdot P(o_t = I) + P(r_t = I|o_t = II) \cdot P(o_t = II)\big) \cdot P(\rho|r_t = I)\big] + (y)[P(o_t = I)] \; \forall \, t$$

With regard to 'retest' alternative state $s_t = 2$, in terms of non-promotion, we not only consider those students that stay within achievement level II on the retest, but also those that improve from a level I to a level II on the retest. The expected cost for state $s_t = 2$ under the 'retest' action ($a = 2$) is calculated as follows:

$$q_t(2,2) = (x)\big[\big(P(r_t = II|o_t = II) \cdot P(o_t = II) + P(r_t = II|o_t = I) \cdot P(o_t = I)\big) \cdot P(\rho|r_t = II)\big] + (y)[P(o_t = II)] \; \forall \, t$$

Again, we let $q_t(3,2) = 0$ and $q_t(4,2) = 0$ since we assumed that 'proficient' students would not take a retest and are not at risk of being retained.

3. Numerical Example

3.1 Source & Characteristics of the Data

For this study, EOG testing data was obtained from the North Carolina Education Research Data Center (NCERDC) at Duke University (NCERDC 2012). Established in 2000, the NCERDC stores and manages data collected by the North Carolina DPI. All procedures for acquiring, storing, and analyzing the NCERDC data used in this study were approved by the Duke University Institutional Review Board and the Institutional Review Board of our university (Miranda et al. 2007). Each year, the NCERDC creates separate databases organized by grade level containing students' EOG mathematics test and reading comprehension test scale scores and achievement level scores with select demographic and socioeconomic data. All student data is encrypted and de-identified; names and other recognizable identifiers are replaced by a unique master identification number. These identification numbers are created by the NCERDC to identify individual test scores and cannot be traced to specific students, thus meeting state and federal human subjects research requirements. Statewide test results were obtained from 1995 to 2010, the most recent year of data available at the time of this study.

Using the student's master identification number, we longitudinally tracked EOG test performance of eleven different cohorts of third grade students through the eighth grade. For example, the 2005-2010 cohort included third grade students that took the EOG tests in the 2004-2005 academic year and subsequently took the eighth grade EOG tests in 2009-2010. We excluded all records with erroneous duplicate master identification numbers and all records with blank entries for the test administration variable (i.e., ADMINIST = 0 [regular

administration]; = 1 [retest #1]; = 2 [retest #2]). Due to small sample sizes, we excluded all learning disabled and otherwise exceptional students. After these exclusions (on average fewer than 5% of the total records in a dataset), each of the longitudinal datasets contained at least 50,000 student records, indicating statistical strength.

In 2006 the scaling of the EOG mathematics test scores was changed significantly to align with the revised mathematics curriculum. According to a background report, "…comparisons between the performance of schools from 2006 forward and prior years should be avoided" (Public Schools of North Carolina 2011b). In this analysis we consider student performance after 2006 to ensure consistency amongst our comparisons.

With regard to the retest data, the NCERDC creates separate records for a student's score on the original administration of the EOG test and any subsequent retest scores. Although a student can be administered a maximum of two retests, the higher of the original administration and first retest score are considered in DPI summary reports, therefore we only consider the first retest score in this analysis. Students may be retested in mathematics and/or reading comprehension. Given that the focus of this study is on mathematics education, we excluded records indicating that a student took a retest in reading comprehension. As mentioned in Section 1.1.2, the state modified their policy on retesting in 2009 requiring students in all grade levels to be retested if they scored within achievement level II on the EOG test. As a result, we only considered retest data from 2009 and 2010.

3.2 Parameter Estimates

3.2.1 Transition Probability Calculations

The results of the one-step transition probability computations are shown in Figure 2-2. The transition probabilities for the 'do nothing' action represent the natural progression of a student's performance on the EOG test without intervention, in this case, without considering performance on a retest. Due to the changes made to the scale scores of the EOG mathematics test in 2006, the transition probabilities for the 'do nothing' action consider performance of students in the post-2006 cohorts. We combined the corresponding longitudinal datasets from the post-2006 cohorts to create aggregate cohorts and calculated the one-step transition probability matrices for each grade level. The cohorts used to calculate each grade level matrix are indicated in Figure 2-2.

To calculate the transition probabilities for the 'retest' action, we linked a student's 2009 retest score to their score on the original administration of the EOG mathematics test in the subsequent grade level in 2010. Due to the policy change in 2009, retests were administered in grades 3-8; therefore we can calculate a transition probability matrix for each grade level transition.

We made two additional assumptions in calculating the one-step transition probabilities. For the 'do nothing' action, due to the stepwise nature of the longitudinal data and given that 2010 was the most recent year of data available from the NCERDC at the time of this study, we were unable to compute the one-step transition probability matrix for the transition from third grade to fourth grade. We assume that the third to fourth grade transition is equivalent to the fourth to fifth grade transition. In addition, very few students who take a

retest score within achievement level IV ($n < 34$ students). We assume that transitions from achievement level IV on the retest are not possible as these may be recording errors.

Do Nothing ($a = 1$)
Source: Post-2006 Aggregate Cohort(s)

Grade 3 to Grade 4 (n = 62,532 records [05-10 cohort])

$p_0(s_1 \mid s_0, 1)$	1	2	3	4
1	0.327961	0.579537	0.092502	0.000000
2	0.086261	0.542949	0.367663	0.003128
3	0.003661	0.134776	0.724367	0.137196
4	0.000000	0.001309	0.256304	0.742387

Grade 4 to Grade 5 (n = 62,532 [05-10 cohort])

$p_1(s_2 \mid s_1, 1)$	1	2	3	4
1	0.327961	0.579537	0.092502	0.000000
2	0.086261	0.542949	0.367663	0.003128
3	0.003661	0.134776	0.724367	0.137196
4	0.000000	0.001309	0.256304	0.742387

Grade 5 to Grade 6 (n = 125,480 [04-09, 05-10])

$p_2(s_3 \mid s_2, 1)$	1	2	3	4
1	0.330889	0.578622	0.090295	0.000193
2	0.096064	0.534455	0.364973	0.004508
3	0.006217	0.138117	0.691419	0.164247
4	0.000071	0.002243	0.213828	0.783858

Grade 6 to Grade 7 (n = 189,737 [03-08, 04-09, 05-10])

$p_3(s_4 \mid s_3, 1)$	1	2	3	4
1	0.426083	0.491411	0.081902	0.000605
2	0.137889	0.500637	0.359414	0.002061
3	0.008334	0.126878	0.736584	0.128203
4	0.000000	0.001043	0.221922	0.777035

Grade 7 to Grade 8 (n = 250,751 [02-07, 03-08, 04-09, 05-10])

$p_4(s_5 \mid s_4, 1)$	1	2	3	4
1	0.342299	0.499750	0.156596	0.001355
2	0.117417	0.443101	0.433251	0.006231
3	0.009486	0.111853	0.699219	0.179442
4	0.000080	0.001239	0.168182	0.830498

Retest ($a = 2$)
Source: 2009 Retest Data Linked to 2010

Grade 3 Retest #1 to Grade 4 Original (n = 18,635 records)

$p_0(s_1 \mid s_0, 2)$	1	2	3	4
1	0.495872	0.432110	0.071560	0.000459
2	0.191647	0.530027	0.275887	0.002439
3	0.049544	0.358055	0.581915	0.010486
4	0.000000	0.001309	0.256304	0.742387

Grade 4 Retest #1 to Grade 5 Original (n = 18,684)

$p_1(s_2 \mid s_1, 2)$	1	2	3	4
1	0.400616	0.531587	0.067411	0.000385
2	0.142099	0.619058	0.237179	0.001664
3	0.033132	0.439542	0.519895	0.007431
4	0.000000	0.001309	0.256304	0.742387

Grade 5 Retest #1 to Grade 6 Original (n = 20,545)

$p_2(s_3 \mid s_2, 2)$	1	2	3	4
1	0.367000	0.577275	0.053983	0.001741
2	0.133923	0.613734	0.250416	0.001927
3	0.031898	0.396296	0.558283	0.013523
4	0.000071	0.002243	0.213828	0.783858

Grade 6 Retest #1 to Grade 7 Original (n = 22,641)

$p_3(s_4 \mid s_3, 2)$	1	2	3	4
1	0.469365	0.464166	0.066469	0.000000
2	0.193449	0.545209	0.260073	0.001269
3	0.042955	0.351984	0.598085	0.006977
4	0.000000	0.001043	0.221922	0.777035

Grade 7 Retest #1 to Grade 8 Original (n = 21,322)

$p_4(s_5 \mid s_4, 2)$	1	2	3	4
1	0.319149	0.537098	0.142662	0.001091
2	0.121640	0.493258	0.380638	0.004464
3	0.029028	0.284906	0.665602	0.020464
4	0.000080	0.001239	0.168182	0.830498

Figure 2-2: One-step transition probabilities for each action (Note due to lack of available data, we assume that the third to fourth grade transition is equivalent to the fourth to fifth grade transition for $a = 1$ [shaded in gray on left]. Also, due to small sample sizes, for $a = 2$, we assume that transitions out of achievement level IV are not possible [shaded in gray on right])

The set of conditional probabilities that represent a student's transition from their original score on the EOG mathematics test (o_t) to their score on the first retest (r_t)

administered within their current grade level are shown in Figure 2-3. As discussed in Section 2.2, we use these intermediary probabilities to estimate the percentage of retested students that remain 'non-proficient' after the retest intervention and are, therefore, at risk of not being promoted to the next grade level.

Source: 2009 Data

Grade 3 Original to Grade 3 Retest #1 (n = 23,496 records)

n_i		1	2	3	4
3860	1	0.430829	0.525907	0.043005	0.000259
19614	2	0.087947	0.539513	0.370552	0.001988
19	3	0.473684	0.210526	0.315789	0.000000
3	4	0.666667	0.333333	0.000000	0.000000

Grade 4 Original to Grade 4 Retest #1 (n = 21,811 records)

n_i		1	2	3	4
3888	1	0.448045	0.498714	0.053241	0.000000
17902	2	0.095799	0.522288	0.380907	0.001005
21	3	0.476190	0.190476	0.333333	0.000000
0	4	-	-	-	-

Grade 5 Original to Grade 5 Retest #1 (n = 23,748 records)

n_i		1	2	3	4
3330	1	0.434835	0.528829	0.035435	0.000901
20401	2	0.075045	0.562472	0.361110	0.001372
16	3	0.562500	0.250000	0.187500	0.000000
1	4	0.000000	1.000000	0.000000	0.000000

Grade 6 Original to Grade 6 Retest #1 (n = 25,945 records)

n_i		1	2	3	4
3700	1	0.382703	0.579730	0.037027	0.000541
22209	2	0.088207	0.558647	0.351839	0.001306
32	3	0.437500	0.187500	0.375000	0.000000
4	4	0.500000	0.250000	0.000000	0.250000

Grade 7 Original to Grade 7 Retest #1 (n = 24,477 records)

n_i		1	2	3	4
5182	1	0.445388	0.498842	0.055770	0.000000
19273	2	0.117729	0.508016	0.373528	0.000726
18	3	0.555556	0.166667	0.222222	0.055556
4	4	1.000000	0.000000	0.000000	0.000000

Figure 2-3: Intermediary probabilities representing movement from the original administration of the EOG mathematics test to the first retest of the exam within time period *t* (Note the small sample sized associated with the movements from states 3 and 4 shaded in gray. 'Proficient' students are typically not administered a retest, thus we assume that these are recording errors and disregard them from further analyses).

3.2.2 Cost Estimates

There are two types of costs considered in this analysis. The first is a fixed cost of administering a retest to a student. The second is the cost of retaining a student in a grade as a function of their level of achievement on the EOG test. We discuss each in turn in the following sections.

Fixed Cost of a Retest

The total fixed cost of a retest is a combination of administrative costs and the cost of teacher labor. The administrative costs include the test materials (i.e., the test booklet and multiple choice scan sheet) and the cost of processing or scoring the exam. We estimate these costs to be approximately $2 per student. With regard to teacher labor, the US Bureau of Labor Statistics reported that the 2010 mean annual salary of elementary school teachers in North Carolina was $43,200 and $42,670 for middle school teachers (Bureau of Labor Statistics 2011). For simplicity, we assume the average teacher salary is $43,000 per year and we approximate fringe benefits at roughly 30% of the annual teacher salary (NC State 2012). North Carolina teachers are employed ten months out of the year, equating to approximately 200 days. We assume the administration of a retest requires one day (i.e., 8 hours) of teacher labor. This accounts for the four hour time period students have to complete the exam plus time for the teacher to read the instructions, take scheduled breaks, and perform other administrative tasks (Public Schools of North Carolina 2008). Class size restrictions in North Carolina vary by grade level (Public Schools of North Carolina 2011i). For simplicity, we assume that 25 students are administered a retest per classroom. Given the preceding assumptions, we calculate that the administration of a retest costs $11.18 per student in

teacher labor. The total fixed cost of a retest per student is the summation of the administrative cost plus teacher labor, $13.18 per student.

Cost of Retaining a Student in a Grade

In the latest edition of the North Carolina Public Schools Statistical Profile, DPI reports that the total expenditure per pupil excluding child nutrition programs was $8,045.00 per student in 2007-08, the most recent year of data available (Public Schools of North Carolina 2009a, Table 23). We assume that this value can represent the total cost of educating a student for one year and would thus translate into the cost of retaining a student in a grade level for one year.

3.3 Model Validation

To validate the rates of non-promotion we calculated, we compared them to the actual aggregate rates of non-promotion available from DPI. The rates of non-promotion for the 2007-08 academic year, the most recent year of data available at the time of this study, are shown in Table 2-1, column B. We observe that more students were retained in grades 3, 5, and 8 than in the other grades. This was a function of the pre-2009 policy that identified grades 3, 5, and 8 as "gateways" for promotion decisions (Public Schools of North Carolina 2006d). Retests were the primary "safeguard" implemented in the "gateway" grade levels; very few retests were administered in grades 4, 6, and 7 (fewer than 1% of students were retested). As a result of the policy change in 2009, all students that score within achievement level II in all grade levels must be administered a retest (Public Schools of North Carolina 2009b). To more accurately represent the impact retesting has on promotion decisions, we used the NCERDC datasets to calculate the percentages of non-promoted students in 2009.

Using the student's master identification number, we tracked students that appeared in the same grade level in 2009 and 2010, indicating that they were not promoted to the next grade in the subsequent year. These results are shown in Table 2-1, column C. These rates are lower than those of 2008, which we assume in part is a byproduct of the policy change enforced in 2009. In general, the non-promotion rates presented in Table 2-1 represent the percentages of students retained in their current grade after intervention, such as retesting on the EOG exam. In Section 4, we use the values in Table 2-1, column C to validate the total calculated rates of non-promotion for the 'retest' alternative derived numerically. For the sake of the numerical example, we define:

$$P(\rho|a = 1) = [(P(\rho|o_t = I) \cdot P(o_t = I)] + [(P(\rho|o_t = II) \cdot P(o_t = II)] \; \forall \; t$$

as the total calculated probability of non-promotion for grade level *t* for the 'do nothing' action. In the numerical example, we found this value to be comparable across all grade levels, thus, we suppress *t* in the notation for this parameter. For the 'retest' alternative, define:

$$P(\rho|a = 2) = \big[\big(P(r_t = I|o_t = I) \cdot P(o_t = I) + \; P(r_t = I|o_t = II) \cdot P(o_t = II)\big) \cdot P(\rho|r_t = I)\big] + \; \big[\big(P(r_t = II|o_t = II) \cdot P(o_t = II) + \; P(r_t = II|o_t = I) \cdot P(o_t = I)\big) \cdot P(\rho|r_t = II)\big] \; \forall \; t$$

as the total calculated rate of non-promotion for grade level *t* for the 'retest' action and, as discussed previously, we suppress *t* from the notation.

Table 2-1: Actual rates of non-promotion by grade level for the 2007-08 academic year (Public Schools of North Carolina 2009a, Table 7) and calculated rates of non-promotion by grade level for the 2008-09 academic year

(A) Grade Level	(B) Actual Rate of Non-Promotion (2007-08)	(C) Calculated Current Rate of Non-Promotion (2008-09)
3	3.3%	2.9%
4	1.5%	1.3%
5	3.2%	1.0%
6	2.1%	1.1%
7	2.4%	1.1%
8	3.3%	1.1%

4. Results and Discussion

4.1 Characteristics of the Population & Associated Assumptions

Figure 2-4 shows the distribution of 'non-proficient' students after the original administration of the EOG mathematics test for all grade levels in 2009 (i.e., $P(o_t = I)$ and $P(o_t = II)\ \forall\ t$).

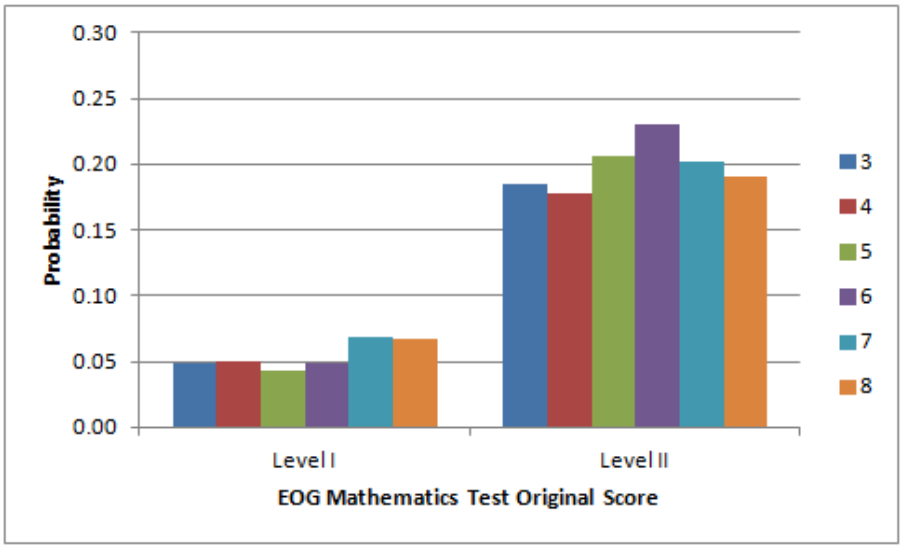


Figure 2-4: Distribution of non-proficient students in the 2008-09 academic year by grade level

Given that $P(o_t = II) > P(o_t = I)$ for all grade levels and that the aggregate rates of non-promotion provided in Table 2-1 are small, we hypothesize that large percentages of students that score within achievement level II on the EOG mathematics test are promoted each year. In an effort to validate the model, we originally assume that the respective percentages of level II students that are not promoted (i.e., $P(\rho|o_t = II)$ or $P(\rho|r_t = II)$) are also small. This ensures that the total calculated probability of non-promotion for a given grade level is comparable to the actual value for that grade.

In general, we assume that a larger proportion of level I students are not promoted to the next grade level, regardless of the action. We consider the relative estimates of non-promoted level I and level II students in two ways:

Scenario 1: We assume that more 'non-proficient' students are retained in their current grade level under the 'do nothing' action since an intervention to their learning does not take place and promotion decisions are judged on their original test score (i.e., $P(\rho|a = 1) > P(\rho|a = 2)$); or

Scenario 2: We assume that more 'non-proficient' students are withheld after the retest since they failed to progress to 'proficiency' after the intervention (i.e., $P(\rho|a = 2) > P(\rho|a = 1)$).

Given the restrictions on the actions available in each state discussed in Section 2.1, our primary emphasis is on procuring optimal decision policies for 'non-proficient' students.

4.2 Optimal Decision Policies

We used Howard's (1960) value iteration method and Matlab to solve the MDP. Preliminary numerical results indicated that the model is highly sensitive to estimates of the

percentages of students retained, given the large cost associated with retaining a student in a grade level. As a result, in Figure 2-5 we present a subset of optimal decision policies as a function of five different configurations of the percentages of 'non-proficient' students that are not promoted. For example, in Case 1, we assumed that $P(\rho|o_t = I) = 0.80$ and $P(\rho|o_t = II) = 0.10$, meaning that 80% of those students that scored within level I on the original administration of the EOG test and 10% of those students that scored within level II were not promoted to the next grade level. For the 'do nothing' action ($a = 1$), this combination of retention assumptions by achievement level yielded a total calculated rate of non-promotion, $P(\rho|a = 1)$, of approximately 6% for each grade. Similar logic can be followed for the 'retest' action ($a = 2$). Relationships between the various rates of non-promotion are indicated in Figure 2-5 under the respective case numbers. Note the total calculated rates of non-promotion for the 'retest' action in some cases are slightly greater than those presented in Table 2-1. Given that the actual 2009 rates were not available at the time of this study, we contend that the calculated rates are reasonable but require further validation.

Case 1

$P(\rho \mid a=1) > P(\rho \mid a=2), P(\rho \mid o_t=1) = P(\rho \mid r_t=1), P(\rho \mid o_t=2) = P(\rho \mid r_t=2)$

Non-Promotion Estimates			Optimal Decision by Grade Level				
State\Action	$a=1$	$a=2$	G3	G4	G5	G6	G7
1	0.80	0.80	2	2	2	2	2
2	0.10	0.10	1	1	1	1	2
3			1	1	1	1	1
4			1	1	1	1	1
$P(\rho \mid a)$*	~6%	~4-5%					

Case 2

$P(\rho \mid a=1) > P(\rho \mid a=2), P(\rho \mid o_t=1) > P(\rho \mid r_t=1), P(\rho \mid o_t=2) > P(\rho \mid r_t=2)$

Non-Promotion Estimates			Optimal Decision by Grade Level				
State\Action	$a=1$	$a=2$	G3	G4	G5	G6	G7
1	0.80	0.75	2	2	2	2	2
2	0.10	0.05	2	2	2	2	2
3			1	1	1	1	1
4			1	1	1	1	1
$P(\rho \mid a)$	~6%	~3%					

Case 3

$P(\rho \mid a=1) \approx P(\rho \mid a=2), P(\rho \mid o_t=1) < P(\rho \mid r_t=1), P(\rho \mid o_t=2) < P(\rho \mid r_t=2)$

Non-Promotion Estimates			Optimal Decision by Grade Level				
State\Action	$a=1$	$a=2$	G3	G4	G5	G6	G7
1	0.85	0.90	1	2	2	2	2
2	0.10	0.15	1	1	1	1	1
3			1	1	1	1	1
4			1	1	1	1	1
$P(\rho \mid a)$	~6%	~5-6%					

Case 4

$P(\rho \mid a=1) < P(\rho \mid a=2), P(\rho \mid o_t=1) < P(\rho \mid r_t=1), P(\rho \mid o_t=2) < P(\rho \mid r_t=2)$

Non-Promotion Estimates			Optimal Decision by Grade Level				
State\Action	$a=1$	$a=2$	G3	G4	G5	G6	G7
1	0.70	0.90	1	1	1	1	1
2	0.05	0.25	1	1	1	1	1
3			1	1	1	1	1
4			1	1	1	1	1
$P(\rho \mid a)$	~4%	~7%					

Case 5

$P(\rho \mid a=1) > P(\rho \mid a=2), P(\rho \mid o_t=1) = P(\rho \mid r_t=1), P(\rho \mid o_t=2) = P(\rho \mid r_t=2)$

Non-Promotion Estimates			Optimal Decision by Grade Level				
State\Action	$a=1$	$a=2$	G3	G4	G5	G6	G7
1	0.90	0.90	2	2	2	2	2
2	0.15	0.15	1	1	2	2	2
3			1	1	1	1	1
4			1	1	1	1	1
$P(\rho \mid a)$	~7%	~5-6%					

Figure 2-5: Optimal decision policy for five possible cases (*The calculated rate represents the total rate of non-promotion given the percentages of level I and level II students assumed to have been retained. These values vary by grade level, but were typically within 1% of each other; thus, we present an approximate rate for each action here)

It is clear that the model is sensitive to small changes in the retention level assumptions. Cases 1, 2, and 5 represent the situation described in Scenario #1 where more students, in total, are retained in the absence of an intervention. In these cases, it is optimal to retest all level I students in all grade levels. In Case 2 where 5% fewer level II students are assumed to be retained for the 'retest' action than for the 'do nothing' alternative, it is optimal to retest all level II students as well. Referencing Cases 1 and 5, when the percentages of level II students retained under each action are equivalent, the decision to retest depends on the grade level. For Case 5, upon further examination of the transition

probabilities in Figure 2-2, retested level II students are twice as likely to regress to a level I than non-retested students in the third to fourth and fourth to fifth grade transitions, hence the decision to forgo the retest intervention in the third and fourth grades, respectively. Additionally, as the difference between the respective total calculated rates of non-promotion increases from ~2% (Cases 1 and 5) to 3% (Case 2), the retest action becomes more attractive for level I and level II students in all grade levels.

Case 4 corresponds to the situation described in Scenario #2 where a larger percentage of 'non-proficient' students are retained after the retest intervention. Given the 3% difference between the total calculated rates of non-promotion for this case, it is optimal to forgo a retest intervention for all students in all grade levels. Per the preceding discussion of Cases 1, 2, and 5, if the total calculated rates of non-promotion differ by more than 3%, we may conclude that the action associated with the lower rate of non-promotion will always be optimal. Additional numerical cases can validate this conclusion.

Case 3 represents the situation where the total calculated rates of non-promotion for each action are nearly equivalent and we vary the individual levels of achievement level I and level II students retained. For Case 3, it is optimal to retest level I students in all grade levels except the third grade, while all level II students should forgo an intervention. Even though 5% more level I students are assumed to be retained after a retest, given that the total calculated probability of non-promotion for this action is slightly lower (with a difference of less than 1%), it is still optimal to retest level I students in grades 4-7. This case shows that the model may be more sensitive to the relationship between the total calculated rates of non-promotion amongst the actions as opposed to the relationship between the assumed

individual levels of 'non-proficient' students retained. From the preceding discussion, it is clear that there is a complex interaction between the assumed rates of non-promotion, and this interaction warrants further study. Ideally, actual values obtained from DPI would serve to validate the model.

As discussed previously, if we assume that large percentages of students that score within achievement level II on the EOG mathematics retest are promoted each year, many of the policies in Figure 2-5 suggest administering a retest to all level I students with the expectation that they would advance to level II on the retest, and would therefore be promoted to the next grade level. Further, given that level I students are at a higher risk of being retained and that there is a lower likelihood of a student scoring within level I in the first place, the decision to retest all level I students in some cases seems reasonable. These policies differ from North Carolina's decision in 2009 to allocate retesting resources to all level II students and to *optionally* retest level I students.

There are a few interesting results to note regarding the one-step transitions probabilities in relation to the intermediary transition probabilities (i.e., the set of observations). Referencing Figure 2-3, although it is unlikely for students that score within achievement level I on the original administration of the exam to score within level III on the retest (fewer than ~5% for all grade levels), level II students advance to 'proficiency' on the retest at least 35% of the time. This value is consistent across all grade levels. More importantly, of those students that score within achievement level III on the retest, more than 50% will remain 'proficient' in the next grade level. Thus, there is utility in administering retests to a subset of level II students that show promise in advancing to 'proficiency.'

Approximately half of the students that score within level I on the original administration of the exam will advance to level II on the retest and about half of level II students will stay within level II after being retested. Of these students that score within level II on the retest, at least 24% will advance to level III, and thus 'proficiency,' in the next grade level.

4.3 Preliminary Investigation of Structural Properties

In this study, the numerical example offers a reality based exploration into the structural properties of the model. The numerical insights suggest that an optimal control-limit type policy may exist given the appropriate conditions (Barlow & Proschan 1965, Chhatwal et al. 2010). As demonstrated by Case 2, if $P(\rho|a = 1)$ is sufficiently larger than $P(\rho|a = 2)$, then a control-limit policy may exist where it is optimal to administer a retest to any student that performs below level III, and to forgo an intervention otherwise. Conversely, if $P(\rho|a = 2)$ is sufficiently larger than $P(\rho|a = 1)$, such as in Case 4, a threshold is not reached and it is optimal to forgo an intervention in all states. Cases 1 and 3 show some evidence of a control-limit policy where students scoring lower than level II should be administered a retest. Here we are speaking generally across all grade levels, but obviously thresholds can be defined for each grade level. A thorough investigation of the structural properties and sufficient conditions required to prove them is an area of future work.

4.4 Sensitivity Analysis on Cost

The model is not sensitive to small changes in the fixed cost of administering the retest given the large disparity between it and the total cost of retaining a student in one grade level. With regard to the retention cost, the model does not appear to be sensitive to small changes within ±10% of the original estimate. We did notice various changes to the optimal

policies as a function of large increases or decreases in the retention cost; however, the practicality of such drastic differences may not make sense in the context of this problem.

5. Conclusions

5.1 Summary

In this study we develop an MDP model of student performance on the North Carolina EOG mathematics test to determine an optimal control policy for the administration of retests. The model offers a basic structure for making sequential decisions about a student's education. A model, by definition, is a representation of reality and we acknowledge the challenges in applying MDP models to the complex social issues found in education systems, but suggest that this study serve as a starting point for discussion. This work is the first step in developing a Markovian framework for optimizing the vertical alignment of mathematics education. Supplementary to the educator's judgment, the model provides a prescription for systematically reviewing a student's performance and determining an appropriate course of action. In the context of EOG mathematics test performance, a key result of this study is that given sufficient conditions, it is optimal to retest students that score within achievement level I on the original administration of the exam. This result contradicts the current policy to optionally retest level I students.

5.2 Model Limitations and Future Work

The reward structure of this MDP model considers costs in isolation of any benefit earned by the system for students that advance to a state of 'proficiency.' An area of future work is to incorporate a reward structure into the model to maximize an individual student's total expected reward, and thus their proficiency, by the eighth grade.

It is clear that promotion decisions are not based on EOG mathematics test scores alone. Further investigation may warrant expansion of the state space to include performance on the EOG reading comprehension test.

Similar to medical decision making models where the progression of disease may not behave as expected, human cognition may not evolve as we expect. As a result, future work will examine potential structural properties of the model to offer educators more efficient optimal policy solutions.

The MDP model in this study considers one possible learning intervention that may impact promotion decisions, but can be expanded to include other interventions that affect a student's performance trajectory. We plan to extend this work by creating other models to measure the effectiveness of various curricular and teaching interventions in North Carolina.

Acknowledgements

The authors thank the NCERDC for granting us access to their datasets. This project has been approved by the NC State University Institutional Review Board for the Use of Human Subjects in Research, IRB Approval Number 1845-11-2.

Chapter 3 - Summary of Contributions

In this chapter we present a summary of the contributions of this research study. In general, the Markovian models described in this research demonstrate a progression of complexity in modeling an uncertain environment, an education system. The Markov chain describes the natural progression of a student's performance on a standardized test over time, while the MDP allows the educator to dynamically alter a student's performance trajectory by their selection of a learning intervention.

1. Contributions of the Markov Chain Model

In Chapter 1 we presented a longitudinal analysis of student performance on the standardized North Carolina EOG mathematics exam and developed Markov chain models to probabilistically characterize the movement of students' scores from one grade level to the next. In doing so, this research was the first study to model the NCERDC EOG test data in this way. The Markov chain model effectively described the evolution of North Carolina students' EOG performance in mathematics in the elementary and middle grades. The model was used to forecast a student's level of proficiency in mathematics in the eighth grade given their level of achievement in the third grade. The results of the model can also inform existing STEM pipeline models linking high school course taking to STEM degree attainment.

2. Contributions of the Markov Decision Process Model

In Chapter 2 we examined a method for optimizing the vertical alignment of STEM learning by modeling the sequential process of learning mathematics over time using an MDP. Specifically, we formulated an MDP model of EOG test performance to determine an

optimal control policy for the administration of retests of the exam and conducted a numerical experiment. The MDP model offers a framework for making sequential decisions about an individual student's education. More explicitly, it allows the educator to dynamically alter a student's performance trajectory by their selection of a learning intervention. Although the MDP in this study considered one control action and associated costs, the model structure can be used to assess other learning interventions and can be used to optimally allocate resources for a large-scale education system. Contrary to many MDP application studies, the numerical example in this research offers a reality based exploration into the structural properties of our model.

3. Future Work

The models presented in this research offer a framework for making sequential decisions about an individual student's education in an effort to optimize their level of mathematical proficiency attained by the eighth grade. Future research will focus on extending these models and applying similar techniques to other aspects of the K-12 STEM education system. In the short-term, we plan to investigate the structural properties of the MDP model and conduct further numerical experimentation. Another area for future work will be to model a student's accumulation of mathematical knowledge over time by allowing for state uncertainty.

REFERENCES

Alagoz, O., Maillart, L., Schaefer, A. & Roberts, M. (2004). The optimal timing of living-donor liver transplantation. *Management Science*, 50(10), 1420-1430.

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher*, 37(2), 65-75.

Atkinson, R. (1972). Ingredients for a theory of instruction. *Stanford University Institute for Mathematical Studies in the Social Sciences Technical Report*, 379-389. Retrieved April 30, 2012 <http://www.rca.ucsd.edu/selected_papers/11_Ingredients%20for%20a%20theory%20of%20instruction.pdf>.

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

Barlow, R. & Proschan, F. (1965). *Mathematical Theory of Reliability*. New York: Wiley.

Baum, L. & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6), 1554-1563

Beal, C., Mitra, S. & Cohen, P. (2007). Modeling learning patterns of students with a tutoring system using hidden Markov models. *Proceedings of 13th International Conference on Artificial Intelligence in Education*, 238-245.

Bifulco, R. & Ladd, H. (2004). The impacts of charter schools on student achievement: evidence from North Carolina. *Duke University Terry Sanford Institute of Public Policy Working Paper Series SAN04-01*. Retrieved August 30, 2011 <http://epsl.asu.edu/epru/articles/EPRU-0412-76-OWI.pdf>.

Bureau of Labor Statistics. (2011). May 2011 State Occupational Employment and Wage Estimates - North Carolina. Retrieved March 30, 2012 <http://www.bls.gov/oes/current/oes_nc.htm#25-0000>.

Burkam, D. & Lee, V. (2003). Mathematics, foreign language, and science course taking and the NELS:88 transcript data. *U.S. Department of Education National Center for Education Statistics*. Retrieved September 14, 2011 <http://nces.ed.gov/pubs2003/200301.pdf>.

Cassandra, A., Kaelbling, L. & Kurien, J. (1996). Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2, 963-972.

Cassandra, A. (1998). A Survey of POMDP applications. *AAAI Fall Symposium: Planning with POMDPs*, Orlando, FL.

Cha, H., Kim, Y., Park, S., Yoon, T., Jung, Y. & Lee, J. (2006). Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. *In M. Ikeda, K. Ashley & T. Chan (Eds.) Proceedings of the 8th International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science*, Berlin, Heidelberg, Springer, 4053, 513-524.

Chhatwal, J., Alagoz, O. & Burnside, E. (2010). Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations Research*, 58(6), 1577-1591.

Common Core State Standards Initiative. (2010). Mathematics Standards. Retrieved April 30, 2012 <http://www.corestandards.org/>.

Confrey, J., Maloney, A., Nguyen, K., Mojica, G. & Myers, M. (2009). Equipartitioning/splitting as a foundation of rational number reasoning using learning trajectories. *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education*, Thessaloniki, Greece.

Confrey, J. & Maloney, A. (2010). The construction, refinement, and early validation of the equipartitioning learning trajectory. *Proceedings of the 9th International Conference of the Learning Sciences*, Chicago, IL, 1, 968-975.

Cook, P. J., MacCoun, R., Muschkin, C. & Vigdor, J. (2008). The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management*, 27(1), 104-121.

Economics & Statistics Administration. (2011). STEM: Good Jobs Now and For the Future. Retrieved March 30, 2012 <http://www.esa.doc.gov/Reports/stem-good-jobs-now-and-future>.

Fam, A., Sarkar, I. & Almuhareb, K. (2006). Academic dishonesty: A probabilistic model using Markov chains. *Proceedings of the American Society for Engineering Education Annual Conference*, Chicago, IL.

Furnee, C., Groot, W. & Maassen van den Brink, H. (2008). The health effects of education: A meta-analysis. *The European Journal of Public Health*, 18(4). Retrieved April 30, 2012 <http://eurpub.oxfordjournals.org/cgi/reprint/ckn028v1>.

Goldhaber, D. & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics*, 89(1), 134-150.

Groot, W. & Maassen van den Brink, H. (2006). What does education do to our health? *Measuring the Effects of Education on Health and Civic Engagement: Proceedings of the Copenhagen Symposium*. Retrieved April 30, 2012 <http://www.oecd.org/dataoecd/15/17/37425763.pdf>.

Hauskrecht, M & Fraser, H. (2000). Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artificial Intelligence in Medicine*, 18, 221-244.

Hoey, J., von Bertoldi, A., Poupart, P. & Mihailidis, A. (2007). Assisting persons with dementia during hand washing using a partially observable Markov decision process. *Proceedings of the International Conference on Vision Systems*. Retrieved March 30, 2012 <http://www.cs.uwaterloo.ca/~ppoupart/publications/coach-icvs/HoeyICVS07.pdf>.

Howard, R. (1960). *Dynamic Programming and Markov Processes*. Cambridge, MA: M.I.T. Press.

Hu, C., Lovejoy, W. & Shafer, S. (1996). Comparisons of some suboptimal control policies in medical drug therapy. *Operations Research*, 44(5), 696-709.

Hui, T.K. & Goldsmith, T. (2012, February 19). Board wrestles with Algebra I. *The Cary News*.

Jaakkola, T., Singh, S. & Jordan, M. (1995). Reinforcement learning algorithm for partially observable Markov decision problems. *Advances in Neural Information Processing Systems*, 7, 345-352.

K12 Academics Website. (2011). Standardized Testing. Retrieved September 19, 2011 <http://www.k12academics.com/standardized-testing>.

Karush, W. & Dear, R. (1966). Optimal stimulus presentation strategy for a stimulus sampling model of learning. *Journal of Mathematical Psychology*, 3(1), 19-47.

Karush, W. & Dear, R. (1967). Optimal strategy for item presentation in a learning process. *Management Science*, 13(11), 773-785.

Klingbeil, N., Mercer, R., Rattan, K., Raymer, M. & Reynolds, D. (2005). Work in progress - The WSU model for engineering mathematics education. *Frontiers in Education - Proceedings of the 35th Annual Conference*, Indianapolis, IN.

Kuravsky, L. & Malykh, S. (2004). Application of Markov models for analysis of development of psychological characteristics. *Australian Journal of Educational & Developmental Psychology*, 4, 29-40.

Laubsch, J. (1969). An adaptive teaching system for optimal item allocation. *Stanford University Institute for Mathematical Studies in the Social Sciences Technical Report*. No. 151.

Littman, M, Cassandra, A. & Kaelbling, L. (1995). Learning policies for partially observable environments: Scaling up. *Proceedings of the Twelfth International Conference on Machine Learning*, San Francisco, CA, 362-370.

Lovejoy, W. (1991a). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28(1), 47-65.

Lovejoy, W. (1991b). Computationally feasible bounds for partially observed Markov decision processes. *Operations Research*, 39(1), 162-175.

Loveless, T. (2008). The misplaced math student: Lost in eighth grade Algebra. *The 2008 Brown Center Report on Education*. Retrieved September 14, 2011 <http://www.brookings.edu/~/media/Files/rc/reports/2008/0922_education_loveless/0922_education_loveless.pdf>.

Maillart, L., Ivy, J., Ransom, S. & Diehl, K. (2008). Assessing dynamic breast cancer screening policies. *Operations Research*, 56(6), 1411-1427.

Matheson, J. (1964). Optimum teaching procedures derived from mathematical learning models. *Stanford University Department of Engineering-Economics Systems EES Technical Report*. No. CCS-2.

McCaffrey, D., Lockwood, J.R., Koretz, D., Louis, T. & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101.

McFarland, D. (2006). Curricular flows: Trajectories, turning points, and assignment criteria in high school math careers. *Sociology of Education*, 79, 177-205.

Miranda, M., Kim, D., Galeano, M., Paul, C., Hull, A. & Morgan, S. (2007). The relationship between early childhood blood lead levels and performance on End-of-Grade tests. *Environmental Health Perspectives*, 115(8), 1242-1247.

Monahan, G. (1982). A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1), 1-16.

Muschkin, C., Bonneau, K. & Dodge, K. (2009). North Carolina Education Research Data Center Annual Report to the Spencer Foundation. *Duke University Center for Child and Family Policy*. Retrieved August 30, 2011 <http://www.childandfamilypolicy.duke.edu/pdfs/projects/NCERDC_SpencerReport2009_final.pdf>.

National Academy of Sciences, National Academy of Engineering & Institute of Medicine. (2007). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington, DC: The National Academies Press.

National Center for Education Statistics. (2006). The Nation's Report Card: Mathematics 2005. Retrieved September 14, 2011 <http://nces.ed.gov/nationsreportcard/pdf/main2005/2006453.pdf>.

National Council of Teachers of Mathematics (NCTM). (2006). Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence. Retrieved September 14, 2011 <http://www.nctm.org/standards/default.aspx?id=58>.

National Research Council Center for Education: Mathematics Learning Study Committee. (2001). *Adding it Up: Helping Children Learn Mathematics*. Washington, DC: The National Academies Press.

National Research Council. (2011). *Successful STEM Education: A Workshop Summary*. Washington, DC: The National Academies Press.

National Science Board. (2007). National Action Plan for Addressing the Critical Needs of U.S. Science, Technology, Engineering, and Mathematics Education System. Retrieved September 14, 2011 <http://www.nsf.gov/nsb/documents/2007/stem_action.pdf>.

National Science and Technology Council. (2011). The Federal Science, Technology, Engineering, and Math (STEM) Education Portfolio. Retrieved April 22, 2012 <http://www.whitehouse.gov/sites/default/files/microsites/ostp/costem__federal_stem_education_portfolio_report.pdf>.

Nicholls, G., Wolfe, H., Besterfield-Sacre, M. & Shuman, L. (2010). Predicting STEM degree outcomes based on eighth grade data and standard test scores. *Journal of Engineering Education*, 99(3), 209-223.

North Carolina Education Research Data Center. (2008). Abstracts of Papers, Presentations, and Research Projects Employing Data from the NCERDC. *Duke University Center for Child and Family Policy*. Retrieved August 30, 2011 <http://www.childandfamilypolicy.duke.edu/pdfs/projects/NCERDC_ABSTRACTS_092908.pdf>.

North Carolina Education Research Data Center. (2012). Homepage. *Duke University Center for Child and Family Policy*. Retrieved August 31, 2011 <http://www.childandfamilypolicy.duke.edu/project_detail.php?id=35>.

North Carolina State University. (2012). Sponsored Programs and Regulatory Compliance Fringe Benefits. Retrieved March 30, 2012 <http://www.ncsu.edu/sparcs/budgeting/fringe.php>.

Pearson School Website. (2008). Algebra Readiness - The Key to Success for California Students. Retrieved September 14, 2011 <http://www.pearsonschool.com/index.cfm?locator=PSZ78a>.

Perdikaris, S. (1992). A Markov chain model in teachers' decision making. *International Journal of Mathematical Education in Science & Technology*, 23(3), 473-477.

Pinar-Yildirim, T., Besterfield-Sacre, M., Shuman, L., Wolfe, H. & Clark, R. (2007). An analysis of engineering design using Markov chains. *Proceedings of the Industrial Engineering Research Conference*, Nashville, TN.

President's Council of Advisors on Science and Technology. (2010). Report to the President - Prepare and Inspire: K-12 Education in Science, Technology, Engineering and Math (STEM) for America's Future. Retrieved September 14, 2011 <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-stemed-report.pdf>.

Proximity Website. (2011). Largest 100 U.S. School Districts 2010-2011. Retrieved September 14, 2011 <http://proximityone.com/lgsd.htm>.

Public Schools of North Carolina. (1999). Assessment Brief: Understanding End-of-Grade Testing - Achievement Levels. Retrieved September 1, 2011 <http://www.ncpublicschools.org/docs/accountability/testing/eog/asb_achlev.pdf>.

Public Schools of North Carolina. (2006a). Standard Course of Study Test Information Sheet. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/accountability/testing/eog/2003scstestgr3mathtis.pdf> .

Public Schools of North Carolina. (2006b). Achievement Level Descriptors for the EOG Mathematics Tests. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/accountability/testing/eog/eogmathgr3to5ald.pdf>.

Public Schools of North Carolina. (2006c). News Release: State Board of Education Raises Standards for Student Achievement. Retrieved August 31, 2011 <http://www.ncpublicschools.org/newsroom/news/2006-07/20061012-01>.

Public Schools of North Carolina. (2006d). Overview of the Student Accountability Standards/Gateways. Retrieved March 30, 2012 <http://www.ncpublicschools.org/docs/promotionstandards/policy/overviewstandards.pdf>.

Public Schools of North Carolina. (2008). Assessment Brief: Time Allowed for North Carolina End-of-Grade Tests. Retrieved March 30, 2012 <http://www.dpi.state.nc.us/docs/accountability/testing/eog/eogassessmentbrief08.pdf>.

Public Schools of North Carolina. (2009a). Statistical Profile. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/fbs/resources/data/statisticalprofile/2009profile.pdf>.

Public Schools of North Carolina. (2009b). End-of-Grade Retest 1. Retrieved March 30, 2012 <http://www.ncpublicschools.org/docs/accountability/testing/eog/eogretest1.pdf>.

Public Schools of North Carolina. (2011a). Evolution of the ABCs. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/accountability/reporting/abc/2010-11/abcevolution.pdf>.

Public Schools of North Carolina. (2011b). ABCs/AYP 2011 Accountability Report Background Packet. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/accountability/reporting/abc/2010-11/backgroundpacket.pdf>.

Public Schools of North Carolina. (2011c). The North Carolina State Testing Results 2009-2010. Retrieved September 14, 2011 <http://www.ncpublicschools.org/docs/accountability/reports/green/greenbook0910.pdf>.

Public Schools of North Carolina. (2011d). North Carolina End-of-Grade Tests Grade 3-8. Retrieved August 31, 2011 < http://www.ncpublicschools.org/accountability/testing/eog/>.

Public Schools of North Carolina. (2011e). North Carolina End-of-Grade Mathematics Tests Grades 3-8. Retrieved August 31, 2011 <http://www.ncpublicschools.org/accountability/testing/eog/math/>.

Public Schools of North Carolina. (2011f). Accountability Services. Retrieved August 31, 2011 <http://www.ncpublicschools.org/accountability/>.

Public Schools of North Carolina. (2011g). The North Carolina Testing Program Overview 2011-2012. Retrieved August 31, 2011 <http://www.ncpublicschools.org/docs/accountability/overview1112.pdf>.

Public Schools of North Carolina. (2011h). “The Green Book”: The North Carolina State Testing Results. Retrieved January 29, 2012 <http://www.dpi.state.nc.us/docs/accountability/reports/green/greenbook0910.pdf>.

Public Schools of North Carolina. (2011i). School Attendance and Student Accounting Manual. Retrieved March 30, 2012 <http://www.ncpublicschools.org/docs/fbs/accounting/manuals/sasa.pdf>.

Public Schools of North Carolina (2012a). Education Value-Added Assessment System. Retrieved March 10, 2012 <http://www.dpi.state.nc.us/evaas/>.

Public Schools of North Carolina (2012b). Academic Services and Instructional Support: Academically or Intellectually Gifted. Retrieved January 23, 2012 <http://www.ncpublicschools.org/academicservices/gifted/>.

Puterman, M. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley and Sons.

Raudenbush, S. & Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Rosenfeld, D. (1976). Markovian deterioration with uncertain information. *Operations Research*, 24(1), 141-155.

Sanders, W., Saxton, A. & Horn, S. (1997). The Tennessee value-added accountability system: A quantitative, outcomes-based approach to educational assessment. *In J. Millman (Ed.), Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* Thousand Oaks, CA: Corwin Press, 137-162.

Sanders, W. (2006). Comparisons among various educational assessment value-added models. *The Power of Two - National Value-Added Conference*, Columbus, OH. Retrieved March 10, 2012 <http://www.sas.com/resources/asset/vaconferencepaper.pdf>.

Sanders, W., Wright, P., Rivers, J. & Leandro, J. (2009). A response to criticisms of SAS EVAAS. SAS White Paper. Retrieved March 10, 2012 <http://www.sas.com/resources/asset/Response_to_Criticisms_of_SAS_EVAAS_11-13-09.pdf>.

Sanford, E. (1996). North Carolina End-of-Grade Tests: Technical Report #1. *North Carolina Department of Public Instruction, Office of Instructional and Accountability Services*. Retrieved August 30, 2011 <http://www.dpi.state.nc.us/docs/accountability/NC End of Grade Tests.pdf>.

SAS Website (2012). SAS EVAAS for K-12. Retrieved March 10, 2012 <http://www.sas.com/govedu/edu/k12/evaas/index.html>.

Schmidt, W. (2011). *STEM Reform: Which Way to Go*. Paper prepared for the workshop of the Committee on Highly Successful Schools or Programs for K-12 STEM Education, National Research Council, Washington, DC, May 10-12, 2011.

Shah, C. & Burke, G. (1999). An undergraduate student flow model: Australian higher education. *Higher Education*, 37, 359-375.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26, 114-145.

Simon, M. & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning*, 6(2), 91-104.

Smallwood, R. (1971). The analysis of economic teaching strategies for a simple learning model. *Journal of Mathematical Psychology*, 8, 285-301.

Smallwood, R. & Sondik, E. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5), 1071-1088.

Sondik, E. (1971). *The Optimal Control of Partially Observable Markov Processes*. Doctoral Dissertation. Stanford University, Stanford, CA.

Spaan, M. & Vlassis, N. (2005). Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research*, 24, 195-220.

Stevens, R., Johnson, D. & Soller, A. (2005). Probabilities and predication: Modeling the development of scientific problem-solving skills. *Cell Biology Education*, 4, 42-57.

Stevens, R. & Thadani, V. (2006). A Bayesian network approach for modeling the influence of contextual variables on scientific problem solving. *In M. Ikeda, K. Ashley & T. Chan (Eds.) Proceedings of the 8th International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science*, Berlin, Heidelberg, Springer, 4053, 71-84.

Tekwe, C., Carter, R., Ma, C., Algina, J., Lucas, M., Roth, J., Ariet, M., Fisher, T., & Resnick, M. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11-36.

Theocharous, G., Beckwith, R., Butko, N. & Philipose, M. (2009). Tractable POMDP planning algorithms for optimal teaching in SPAIS. *IJCAI Workshop on Plan, Activity and Intent Recognition*.

Tyson, W., Lee, R., Borman, K. & Hanson, M.A. (2007). Science, Technology, Engineering, and Mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, 12(3), 243-270.

Tyson, W. (2011). Modeling engineering degree attainment using high school and college Physics and Calculus course taking and achievement. *Journal of Engineering Education*, 100(4), 760-777.

United States Department of Education. (2007). Report of the Academic Competitiveness Council. Retrieved September 14, 2011 <http://www2.ed.gov/about/inits/ed/competitiveness/acc-mathscience/report.pdf>.

United States Department of Education. (2008). Foundations for Success: The Final Report of the National Mathematics Advisory Panel. Retrieved September 14, 2011 <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>.

United States Department of Education. (2012). Elementary and Secondary Education Act. Retrieved March 30, 2012 < http://www.ed.gov/esea>.

Vendlinski, T. (2002). A Markov model analysis of problem-solving progress. *Annual Meeting of the American Education Research Association*, New Orleans, LA.

Voskoglou, M. (1996). An application of ergodic Markov chains to analogical problem solving. *The Mathematics Education (India)*, 3(2), 96-108.

Voskoglou, M. (2007). A Stochastic Model for the Modeling Process. *In C. Haines, P. Galbraith, W. Blum and S.Khan (Eds.) Mathematical Modeling (ICTMA12): Education, Engineering and Economics*. Chichester: Horwood Publishing, 149-157.

Voskoglou, M. (2009). A Stochastic Model for the Process of Learning. *In L. Paditz and A. Rogerson (Eds.), Models in Developing Mathematics Education (MEC21: Proceedings of the 10th Int. Conference)*, Dresden, Saxony, Germany, 565-569. Retrieved April 16, 2011 <http://math.unipa.it/~grim/21_project/Voskoglou565-569.pdf>.

Wake County Public School System. (2011). Demographics. Retrieved August 31, 2011 <http://www.wcpss.net/demographics/>.

White, III, C. (1991). A survey of solution techniques for the partially observed Markov decision process. *Annals of Operations Research*, 32, 215-230.

Whitehill, J. & Movellan, J. (2009). Optimal teaching machines. Retrieved March 29, 2012 <http://mplab.ucsd.edu/~jake/otm.pdf>.

Willink, R. (2005). A confidence interval and test for the mean of an asymmetric distribution. *Communications in Statistics - Theory and Methods*, 34, 753-766.

Wilson, J. (2011). Skewness-adjusted confidence interval based on Student's *t*-distribution. White Paper/Personal Communication: August 27, 2011.

Wright, P., White, J., Sanders, W. & Rivers, J. (2010). SAS EVAAS statistical models. SAS White Paper. Retrieved March 10, 2012 <http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf>.

Young, R. & Sweet, A. (1982a). Confidence intervals for transaction probability estimates in Markov chains, Part I. *Communications in Statistics: Part A-Theory*, 11(2), 165-179.

Young, R. & Sweet A. (1982b). Confidence intervals for transaction probability estimates in Markov chains, Part II. *Communications in Statistics: Part A-Theory*, 11(2), 181-198.

APPENDICES

Appendix A: NCERDC Variable Descriptions and Coding

In the longitudinal dataset for each cohort, we included the following variables from the respective eighth grade datasets. Note that the coding of some variables changes from year to year; these are provided as an example.

- MASTID = Student's master identification number (and a count variable to double check for duplication)
- BDATE = Birth date set to the 15th of the student's birth month (ex: 1/15/1992)
- SEX = Student's gender (F = Female, M = Male)
- ETHNIC = Student's ethnicity (A = Asian, B = Black, H = Hispanic, I = American Indian, M = Multi-Racial, W = White)
- LEP = Limited English Proficiency status (a variety of codes identifying LEP status and when a student exited the LEP program)
- EXCEPT = Exceptionality (a variety of codes identifying students with special needs, i.e. hearing impaired, visually impaired, autistic, multi-handicapped, etc.)
- SCHLUNCH = Participation in a school lunch program (2 = Free lunch, 3 = Reduced price lunch, 4 = Full pay, school not participating, student not eligible)
- LDMATH = Learning disabled in mathematics (1 = Yes, 0 = No)
- LDREAD = Learning disabled in reading (1 = Yes, 0 = No)
- LDWRITE = Learning disabled in writing (1 = Yes, 0 = No)
- LDOTH = Learning disabled other (1 = Yes, 0 = No) [Note: 5 new LD variables were introduced in 2009].
- AIGMATH = Academically/Intellectually Gifted in mathematics (1 = Yes, 0 = No)
- AIGREAD = Academically/Intellectually Gifted in reading (1 = Yes, 0 = No)
- ADMINIST = Test administration (0 = Regular administration, 1 = Retest #1, 2 = Retest #2)

After joining the tables from the respective grades 3-8 in the span of years included in a cohort, we included the following variables for each grade:

- TESTNAME = Grade level and year test was administered (ex: G32004)
- MATHSCAL = EOG mathematics test scale score
- MATHACH = EOG mathematics test achievement level (1-4)
- READSCAL = EOG reading comprehension test scale score
- READACH = EOG reading comprehension test achievement level (1-4)

Note on Retest Scores

Starting in 2001, public school students in grades 3, 5, and 8 were required to meet specific statewide standards or gateways for promotion and the EOG mathematics test was used as an indicator of proficiency at each gateway. Students must score at or above achievement level III on the EOG mathematics test to demonstrate grade-level proficiency and if they failed to do so, various safeguards including retesting and/or other formal evaluation procedures were implemented (Public Schools of North Carolina 2006). Beginning in 2001, the NCERDC included a variable to indicate if a student took a retest. Prior to the 2007-08 academic year, the NCERDC received raw data from DPI that included retests in the dataset. Separate records for a student's original attempt at an EOG test and their retest score were not included. It is hypothesized that some original test scores were overwritten by the retest score without changing the variable used to indicate the type of administration of the test. Due to a change in the structure of the data received from DPI in 2008, the NCERDC began receiving retest scores as separate records. Thus, the NCERDC was able to include separate records for the original administration of an EOG test, the first retest score, and the second retest score if applicable. We acknowledge the potential bias and inflation in test scores prior to 2008, due to the recording of the retest scores. A new policy on retesting was instituted by the state in the 2008-09 academic year. All public school students in grades 3-8 who score within achievement level II on the first administration of the EOG mathematics and/or reading test are required take a retest. Those who score within achievement level I may opt-in to retest. The higher of the two scores are considered in DPI summary reports.

Appendix B: Example Data Mining Procedure (2000-2005 Cohort)

Data Mining Summary (Similar analysis performed for each cohort)

Step 1: Remove duplicate Master IDs, missing Scale Scores, missing Achievement Level Scores, and all Retests from each raw dataset

Dataset	Grade Level	Academic Year	No. Records in Raw Dataset	After Step 1	Percentage Excluded
eog3pub00	3	1999-2000	106,433	100,709	5.38
eog4pub01	4	2000-2001	100,418	97,545	2.86
eog5pub02	5	2001-2002	107,037	98,375	8.09
eog6pub03	6	2002-2003	108,473	102,874	5.16
eog7pub04	7	2003-2004	109,886	106,184	3.37
eog8pub05	8	2004-2005	109,992	106,122	3.52

Step 2: Create longitudinal dataset

Total Records in Longitudinal Dataset	68,695
Percentage Tracked from 3rd Grade	68.21

Step 3: Remove records with blank entries for select variables & exclude LD and Exceptional students (other than those identified as AIG)

Variable	No. Blank Entries Removed	Percentage Blank Entries Removed		Exclusions	Percentage Excluded	Codes Excluded
Gender	0	0.00		n/a	n/a	
Ethnicity	1	0.00		n/a	n/a	
LEP	4	0.01		n/a	n/a	
Exceptionality	94	0.14		5286	7.69	Codes 03-16
School Lunch	0	0.00		n/a	n/a	
LD Math	0	0.00		975	1.42	Code 1 (Yes)
LD Reading	0	0.00		2197	3.20	Code 1 (Yes)
LD Writing	0	0.00		2299	3.35	Code 1 (Yes)
LD Other	0	0.00		58	0.08	Code 1 (Yes)
AIG Math	0	0.00		n/a	n/a	
AIG Reading	0	0.00		n/a	n/a	

Total Records in Revised Longitudinal Dataset	63,227
Percentage Excluded from Original Longitudinal Dataset	7.96

Step 4: Sample sizes of populations of longitudinal datasets

	Before Exclusions			After Exclusions	
Variable	**Sample Size**	**Percentage**		**Sample Size**	**Percentage**
Gender					
Female	35623	51.86		33855	53.55
Male	33072	48.14		29372	46.45
Total	68695	100.00		63227	100.00
Ethnicity					
Asian	1114	1.62		1080	1.71
Black	19541	28.45		17874	28.27
Hispanic	2143	3.12		2011	3.18
Am. Indian	1041	1.52		973	1.54
Mult-Racial	1265	1.84		1163	1.84
White	43590	63.45		40126	63.46
Blank	1	0.00		0	0.00
Total	68695	100.00		63227	100.00
Limited English Proficiency (LEP)					
Never LEP	67076	97.64		61712	97.60
LEP at some time	1615	2.35		1515	2.40
Blank	4	0.01		0	0.00
Total	68695	100.00		63227	100.00
Exceptionality					
Not Exceptional	49626	72.24		49572	78.40
Identified as AIG	13689	19.93		13655	21.60
Exceptional	5286	7.69		0	0.00
Blank	94	0.14		0	0.00
Total	68695	100.00		63227	100.00
School Lunch Assistance Program					
Free lunch	20605	29.99		18460	29.20
Reduced price lunch	5608	8.16		5138	8.13
Full pay, no participate	42482	61.84		39629	62.68
Total	68695	100.00		63227	100.00
Academically/Intellectually Gifted (AIG) Math					
Yes	11075	16.12		10945	17.31
No	57620	83.88		52282	82.69
Total	68695	100.00		63227	100.00
Academically/Intellectually Gifted (AIG) Reading					
Yes	11072	16.12		10980	17.37
No	57623	83.88		52247	82.63
Total	68695	100.00		63227	100.00

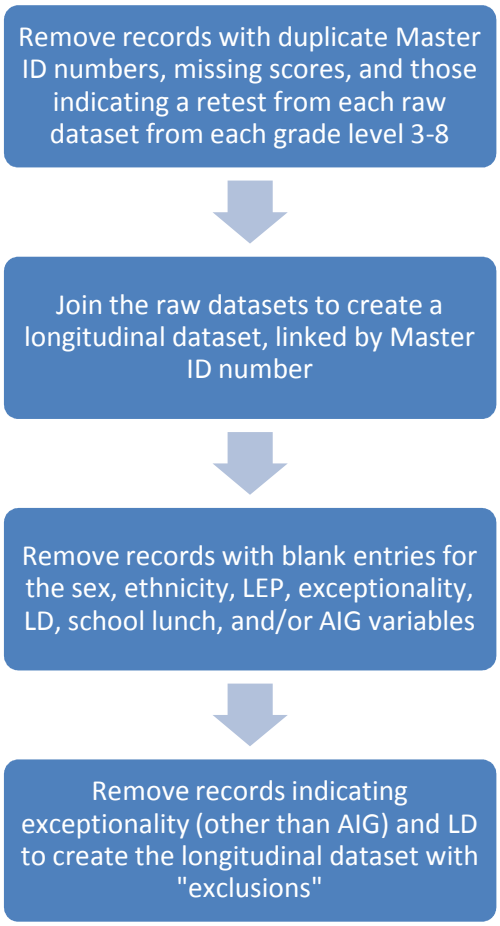


Figure B-1: Flow chart of the procedure followed for analyzing the NCERDC databases and creating the longitudinal datasets

Appendix C: Descriptive Statistical Analysis Methods & Results (2000-2005 Cohort)

The results from one cohort of students (2000-2005) on the gender, ethnicity, school lunch program, and AIG mathematics variables are presented here. We also calculated the correlation between mathematics and reading comprehension test scores for this cohort. All statistical analyses were conducted using SAS Enterprise Guide 4.2 and Microsoft Excel.

Note on Normality and Confidence Interval Calculations

The scaling of the EOG mathematics raw test scores may change from year to year, thus we restrict our review to the achievement level scores. The distribution of the achievement level scores is not normally distributed. This discrete distribution on the integer set {1, 2, 3, 4} is from a truncated normal since it is bounded below by achievement level I and bounded above by achievement level IV. In this case, for comparison of population means, standard *t*-tests are not exact because the sampling population is not normally distributed (Willink 2005). To evaluate the statistical differences in population means, we calculated confidence intervals for the means as described by the following example.

Consider the achievement level scores of a random sample of third grade students of Asian ethnicity from the 2000-2005 cohort and let $\{A_i: i = 1, \ldots, n_A\}$, where n_A = 1,080 students. We aim to construct a $100(1 - \alpha)\%$ confidence interval for the population mean, where the population from which the sample was selected has the following theoretical characteristics: mean μ_A, variance σ_A^2, and skewness, $Sk_A \equiv \mathrm{E}\left[\left(\frac{A-\mu_A}{\sigma_A}\right)^3\right]$ (Wilson 2011).

For asymmetric distributions such as this one, Willink (2005) presented a skewness-adjusted confidence interval based on the Student's *t*-distribution when population skewness $|Sk_A| \leq 4.0$. For sample skewness $\widehat{Sk_A} = 0$, the skewness-adjusted confidence interval reduces to the standard Student's *t*-distribution confidence interval (Wilson 2011). We calculated the sample mean, sample variance, and sample skewness, for each level of each variable under consideration and the results are shown in the next section. Given that our sample sizes are quite large, we are sampling from a bounded distribution, and the sample skewness values are close to 0, we assume normality and calculate the confidence intervals based on the standard Student's *t*-distribution,

$$\left[\bar{A} - t_{n_{A-1},\alpha_{A/2}} \frac{S_A}{\sqrt{n_A}}, \bar{A} + t_{n_{A-1},\alpha_{A/2}} \frac{S_A}{\sqrt{n_A}}\right].$$

For example, for the Asian student population, we calculated the 99.98% ($\alpha = 0.0002$) confidence interval for μ_A and compared it to the 99.98% confidence intervals of the Black, Hispanic, American Indian, Multi-Racial, and White populations to identify true differences in mean performance among third grade students in the 2000-2005 cohort. By the Bonferroni inequality, the probability that all six confidence intervals are valid and simultaneously cover their respective population means is at least $100(1 - 6\alpha)\% = 99.84\%$ (Wilson 2011). For confidence intervals that do not overlap, we conclude that the population means are significantly different. For cases where the confidence intervals of two or more populations do in fact overlap, we examine the degree of overlap to draw conclusions about significant differences with regard to the practicality of such differences. Given our assumption of normality, we also conducted standard *t*-tests to compare population means.

Results and Discussion of Statistical Hypothesis Testing

Only the summaries of the third grade and eighth grade scores are shown for each variable.

1.) Gender

Male and female scores were statistically different at $\alpha = 0.05$ across all grade levels ($p < 0.0001$ in grades 3-7, $p = 0.0131$ in grade 8). Males scored higher than females across all grades; however in eighth grade this difference is not statistically significant at $\alpha = 0.0002$ as evidenced by the overlapping 99.98% confidence intervals highlighted in Table C-1. Although statistical differences in female and male achievement level scores exist, practically the assessment outcomes are the same. In third grade, males' average achievement level scores were only 0.0882 higher than their female counterparts. Given the large width of the achievement level score ranges; it is impractical to assume that male students outperform female students on the EOG mathematics test. The gap between male and female achievement level scores decreased as students progressed to eighth grade (difference of 0.0134).

Table C-1: 2000-2005 cohort – Descriptive statistics by gender

Grade 3 - 2000								
EOG Mathematics	Achievement Level							
Gender	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_i**	
Female	33855	3.1504	0.5823	0.7631	0.0042	-0.4978	3.1350	3.1658
Male	29372	3.2386	0.5797	0.7614	0.0044	-0.6511	3.2221	3.2551
Diff (1-2)		-0.0882						
95% t-test p-value	< 0.0001		Reject null, male/female means are statistically different					

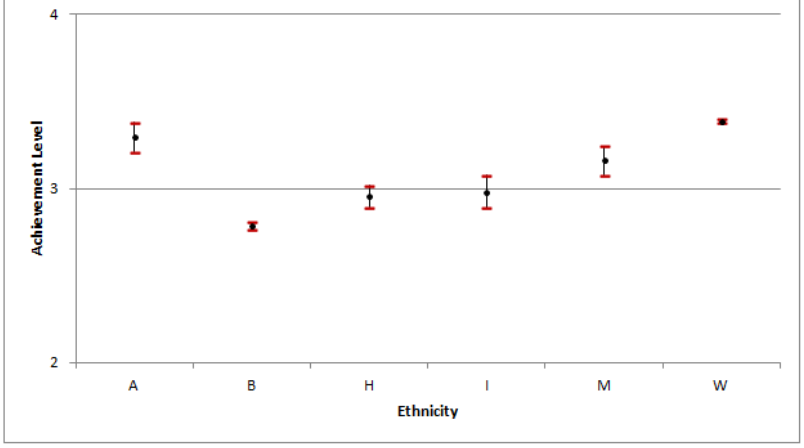
Grade 8 - 2005								
EOG Mathematics	Achievement Level							
Gender	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_i**	
Female	33855	3.5080	0.4473	0.6688	0.0036	-1.2477	3.4945	3.5215
Male	29372	3.5215	0.4735	0.6881	0.0040	-1.3811	3.5066	3.5364
Diff (1-2)		-0.0134						
95% t-test p-value	0.0131		Reject null, male/female means are statistically different					

2.) Ethnicity

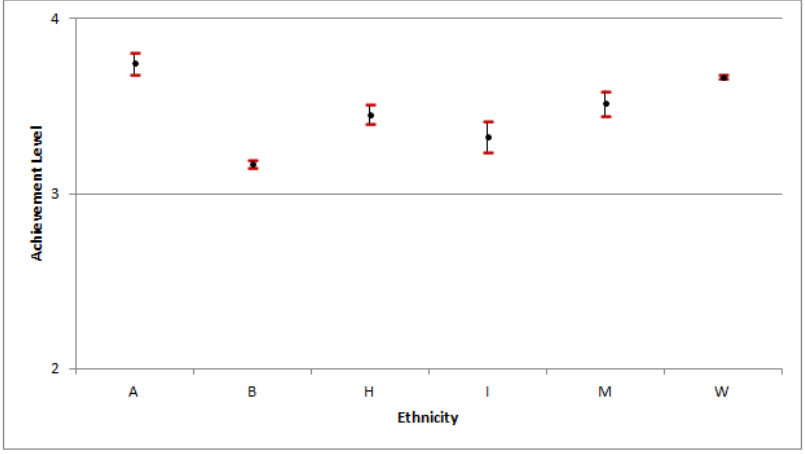
A one-way ANOVA on the EOG mathematics test achievement level scores shows that the means of all levels of the ethnicity variable are not equal at $\alpha = 0.05$ across all grade levels ($p < 0.0001$). However, various pairings of ethnicities were found to be non-significant at $\alpha = 0.05$. As shown in Table C-2, the third grade scores of Hispanic and American Indian students were not statistically different at $\alpha = 0.05$ ($p = 0.3889$). Taking a more narrow view and inspecting the 99.98% confidence intervals in eighth grade, we find that the Hispanic and American Indian confidence intervals overlap as do the Hispanic and Multi-Racial confidence intervals (despite non-significance of these groups at $\alpha = 0.05$ ($p < 0.001$ and $p = 0.0127$, respectively)). Across all grade levels, Asian and White students consistently scored higher than all other ethnicities, while Black students scored lower.

Table C-2: 2000-2005 cohort – Descriptive statistics by ethnicity

Grade 3 - 2000								
EOG Mathematics	Achievement Level							
Ethnicity	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Asian (A)	1080	3.2944	0.5583	0.7472	0.0227	-0.6581	3.2096	3.3793
Black (B)	17874	2.7864	0.5891	0.7675	0.0057	-0.1491	2.7650	2.8077
Hispanic (H)	2011	2.9557	0.5747	0.7581	0.0169	-0.2692	2.8928	3.0187
American Indian (I)	973	2.9815	0.6087	0.7802	0.0250	-0.3715	2.8881	3.0749
Multi-Racial (M)	1163	3.1599	0.5665	0.7527	0.0221	-0.5511	3.0776	3.2423
White (W)	40126	3.3868	0.4659	0.6826	0.0034	-0.7791	3.3741	3.3995
95% p-value	< 0.0001		Reject null, all means not equal					



Grade 8 - 2005								
EOG Mathematics	Achievement Level							
Ethnicity	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Asian (A)	1080	3.7435	0.2780	0.5273	0.0160	-2.1917	3.6836	3.8034
Black (B)	17874	3.1706	0.5922	0.7695	0.0058	-0.6870	3.1492	3.1920
Hispanic (H)	2011	3.4535	0.4579	0.6767	0.0151	-1.0604	3.3973	3.5097
American Indian (I)	973	3.3258	0.5182	0.7199	0.0231	-0.8428	3.2396	3.4120
Multi-Racial (M)	1163	3.5150	0.4324	0.6576	0.0193	-1.2560	3.4431	3.5870
White (W)	40126	3.6688	0.3259	0.5709	0.0028	-1.7214	3.6582	3.6794
95% p-value	< 0.0001		Reject null, all means not equal					



3.) Participation in a School Lunch Program

The school lunch variable is used as an indicator of socioeconomic status and the coding of this variable changed between 2000 and 2005. Upon inspection of the eighth grade results in Table C-3, a one-way ANOVA on EOG mathematics test achievement level scores shows that the three sample distributions are statistically different at $\alpha = 0.05$ ($p < 0.0001$). On average, students that participated in the reduced price lunch program scored lower than those students that did not participate in the program, and students that participated in the free lunch program scored lower than both of these populations. These findings were consistent across all grade levels despite the changes made to the coding of the school lunch variable. In third grade, on average, students that participated in a free lunch program were non-proficient in mathematics, however, by eighth grade these students advanced to proficiency.

Table C-3: 2000-2005 cohort – Descriptive statistics by school lunch program participation

Grade 3 - 2000								
EOG Mathematics	Achievement Level							
School Lunch Participation Status	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Not Eligible (0)	34736	3.3876	0.4789	0.6920	0.0037	-0.8234	3.3737	3.4014
Reduced Price Lunch (1)	5372	3.0352	0.5661	0.7524	0.0103	-0.3333	2.9970	3.0734
Free Lunch (2)	17937	2.8374	0.5900	0.7681	0.0057	-0.1829	2.8160	2.8587
Information Not Available (3)	3638	3.2889	0.5189	0.7204	0.0119	-0.6450	3.2444	3.3334
School Not Participating (4)	429	3.1422	0.6363	0.7977	0.0385	-0.6536	2.9977	3.2867
95% t-test p-value	< 0.0001		Reject null, all means not equal*					

Grade 8 - 2005								
EOG Mathematics	Achievement Level							
School Lunch Participation Status	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Free Lunch	18460	3.1966	0.5896	0.7679	0.0057	-0.7290	3.1756	3.2176
Reduced Price Lunch	5138	3.3968	0.4882	0.6987	0.0097	-0.9636	3.3606	3.4331
Full Pay; School Not Participating	39629	3.6775	0.3198	0.5655	0.0028	-1.7550	3.6669	3.6881
95% t-test p-value	< 0.0001		Reject null, all means not equal					

4.) Academically/Intellectually Gifted (AIG) in Mathematics Students

According to the North Carolina General Statues, "AIG students perform or show the potential to perform at substantially high levels of accomplishment when compared with others of their age, experiences, or environment" (Public Schools of North Carolina 2012b). Initial identification of a student as AIG occurs in third grade upon review of the student's test scores and other informal indicators. Students are classified as AIG in mathematics and/or reading. As shown in Table C-4, the two sample t-tests on EOG mathematics test achievement level scores indicate that the population means of AIG mathematics and non-AIG mathematics students are statistically different at $\alpha = 0.05$ with $p < 0.0001$ in third grade and eighth grade. AIG students scored significantly higher than non-AIG students across all grade levels, although the gap progressively decreased by eighth grade for this cohort of students. This may be an artifact of the nature of the bounded distribution, in that students cannot score higher than an achievement level IV.

Table C-4: 2000-2005 cohort – Descriptive statistics by AIG mathematics status

Grade 3 - 2000								
EOG Mathematics	Achievement Level							
AIG Math Status	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Not AIG Math	59617	3.1478	0.5799	0.7615	0.0031	-0.4991	3.1362	3.1594
AIG Math	3610	3.9102	0.0878	0.2963	0.0049	-3.2820	3.8918	3.9286
Diff (1-2)		-0.7624						
95% t-test p-value	< 0.0001		Reject null, AIG/not AIG means are statistically different					

Grade 8 - 2005								
EOG Mathematics	Achievement Level							
AIG Math Status	**N**	**Mean**	**Variance**	**Std. Dev.**	**Std. Error**	**Skewness**	**99.98% CI for μ_j**	
Not AIG Math	52282	3.4163	0.4963	0.7045	0.0031	-1.0576	3.4048	3.4278
AIG Math	10945	3.9825	0.0185	0.1361	0.0013	-8.3170	3.9777	3.9873
Diff (1-2)		-0.5662						
95% t-test p-value	< 0.0001		Reject null, AIG/not AIG means are statistically different					

Correlation Analysis

There is significant correlation between EOG mathematics test scale scores and EOG reading comprehension test scale scores. As shown in Table C-5, correlation coefficients were consistent across grade levels. This indicates that students with low mathematics test scores are likely to have low reading test scores.

Table C-5: 2000-2005 cohort – Correlation between EOG mathematics and reading test scale scores

Correlation Analysis on Math and Reading Scale Scores (n = 63,227)				
Grade	Pearson Coefficient	Prob > |r|, H_0: Rho = 0	Spearman Coefficient	Prob > |r|, H_0: Rho = 0
3	0.74120	< 0.0001	0.74705	< 0.0001
4	0.76830	< 0.0001	0.77762	< 0.0001
5	0.75101	< 0.0001	0.76228	< 0.0001
6	0.73074	< 0.0001	0.73966	< 0.0001
7	0.73508	< 0.0001	0.74599	< 0.0001
8	0.72616	< 0.0001	0.73227	< 0.0001

Appendix D: One-Step Transition Probablity Matrices for Each Cohort

The one-step transition probability matrices for each cohort studied in Chapter 1 are shown in Figure D-1. Yellow shading indicates the 2006 change to the EOG mathematics test scale scores. Grey shading indicates matrices that were included in the post-2006 analysis.

1995-2000

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.365787	0.526403	0.105061	0.00275
2	0.066536	0.487898	0.430734	0.014832
3	0.002706	0.093296	0.6907	0.213298
4	0.000288	0.002732	0.220042	0.776939

4 -> 5	1	2	3	4
1	0.36523	0.528853	0.101534	0.004383
2	0.103371	0.486832	0.391362	0.018436
3	0.005043	0.113672	0.592488	0.288798
4	0	0.002038	0.109307	0.888655

5 -> 6	1	2	3	4
1	0.239224	0.528017	0.228448	0.00431
2	0.06352	0.367861	0.546766	0.021853
3	0.005131	0.092701	0.678254	0.223913
4	0.000101	0.002425	0.152117	0.845357

6 -> 7	1	2	3	4
1	0.2	0.536047	0.25814	0.005814
2	0.068959	0.37083	0.535801	0.02441
3	0.00642	0.08378	0.630127	0.279673
4	0.000048	0.001059	0.088143	0.91075

7 -> 8	1	2	3	4
1	0.309637	0.508689	0.176935	0.004739
2	0.132819	0.455856	0.399228	0.012098
3	0.0128	0.142124	0.682183	0.162893
4	0.000411	0.002838	0.144972	0.851779

1996-2001

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.324932	0.524523	0.144414	0.006131
2	0.057518	0.454048	0.468711	0.019723
3	0.002654	0.079112	0.665157	0.253077
4	0.000242	0.001878	0.172261	0.825618

4 -> 5	1	2	3	4
1	0.309307	0.52646	0.154197	0.010036
2	0.085174	0.429479	0.460372	0.024975
3	0.0061	0.092585	0.607338	0.293977
4	0.000207	0.001756	0.119233	0.878804

5 -> 6	1	2	3	4
1	0.175094	0.55618	0.263109	0.005618
2	0.054303	0.369654	0.552972	0.02307
3	0.005443	0.086356	0.668341	0.23986
4	0.000211	0.00173	0.146323	0.851736

6 -> 7	1	2	3	4
1	0.232558	0.55814	0.204319	0.004983
2	0.099359	0.446429	0.43544	0.018773
3	0.009015	0.121263	0.629386	0.240336
4	0.000241	0.001887	0.106842	0.89103

7 -> 8	1	2	3	4
1	0.266843	0.486129	0.232497	0.014531
2	0.125796	0.428723	0.425965	0.019516
3	0.020887	0.154365	0.669268	0.15548
4	0.000482	0.00563	0.171951	0.821937

1997-2002

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.211729	0.58949	0.18888	0.009901
2	0.036259	0.398647	0.532471	0.032623
3	0.001804	0.06576	0.630298	0.302137
4	0	0.001719	0.151359	0.846922

4 -> 5	1	2	3	4
1	0.260997	0.552786	0.178886	0.007331
2	0.057831	0.407315	0.497343	0.037512
3	0.003307	0.074172	0.584521	0.338001
4	0.000077	0.001533	0.089767	0.908624

5 -> 6	1	2	3	4
1	0.2187	0.624406	0.153724	0.00317
2	0.082438	0.445041	0.458678	0.013843
3	0.008665	0.128118	0.684664	0.178554
4	0.000186	0.003722	0.173227	0.822865

6 -> 7	1	2	3	4
1	0.185754	0.547486	0.251397	0.015363
2	0.067356	0.458477	0.439533	0.034635
3	0.00649	0.126738	0.59466	0.272112
4	0.000033	0.001757	0.100517	0.897693

7 -> 8	1	2	3	4
1	0.238019	0.496805	0.258786	0.00639
2	0.101883	0.395157	0.473543	0.029417
3	0.013797	0.133119	0.66134	0.191743
4	0.000362	0.004617	0.151448	0.843573

1998-2003

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.149584	0.581717	0.25831	0.010388
2	0.022517	0.294045	0.631982	0.051456
3	0.001186	0.039601	0.599901	0.359313
4	0	0.001481	0.11377	0.884749

4 -> 5	1	2	3	4
1	0.296223	0.55666	0.143141	0.003976
2	0.071061	0.443513	0.45361	0.031816
3	0.004606	0.093134	0.575745	0.326515
4	0.000175	0.001929	0.089672	0.908224

5 -> 6	1	2	3	4
1	0.226357	0.576744	0.192248	0.004651
2	0.063181	0.432957	0.491186	0.012676
3	0.004858	0.112146	0.727126	0.15587
4	0.000145	0.002527	0.185467	0.811861

6 -> 7	1	2	3	4
1	0.20318	0.568905	0.222615	0.0053
2	0.06439	0.460811	0.44929	0.025509
3	0.004367	0.100565	0.585838	0.30923
4	0.000064	0.000932	0.068454	0.930549

7 -> 8	1	2	3	4
1	0.231203	0.473684	0.280075	0.015038
2	0.090486	0.362955	0.515182	0.031377
3	0.013795	0.105115	0.667218	0.213872
4	0.000193	0.002645	0.123391	0.87377

1999-2004

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.107804	0.580048	0.304103	0.008045
2	0.015044	0.273669	0.665526	0.045762
3	0.00066	0.034808	0.60545	0.359082
4	0.000088	0.00101	0.111794	0.887108

4 -> 5	1	2	3	4
1	0.218069	0.566978	0.196262	0.018692
2	0.053266	0.406257	0.504122	0.036356
3	0.003505	0.06954	0.605429	0.321526
4	0.000033	0.00118	0.087608	0.91118

5 -> 6	1	2	3	4
1	0.148681	0.573141	0.273381	0.004796
2	0.056679	0.426717	0.50437	0.012235
3	0.004073	0.090973	0.717296	0.187658
4	0.000055	0.001585	0.144118	0.854243

6 -> 7	1	2	3	4
1	0.150794	0.584656	0.253968	0.010582
2	0.058243	0.441124	0.465688	0.034946
3	0.004312	0.103964	0.590347	0.301377
4	0	0.001387	0.073459	0.925154

7 -> 8	1	2	3	4
1	0.21039	0.475325	0.288312	0.025974
2	0.094857	0.353371	0.514514	0.037257
3	0.013149	0.108257	0.662358	0.216236
4	0.000353	0.003957	0.129354	0.866336

2000-2005

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.072761	0.586754	0.333022	0.007463
2	0.008234	0.260486	0.71578	0.015499
3	0.000403	0.034036	0.785403	0.180158
4	0.000041	0.000366	0.233158	0.766435

4 -> 5	1	2	3	4
1	0.171429	0.571429	0.245714	0.011429
2	0.043713	0.4	0.524794	0.031492
3	0.001748	0.047552	0.496933	0.453766
4	0	0.000251	0.031118	0.968631

5 -> 6	1	2	3	4
1	0.083032	0.537906	0.364621	0.01444
2	0.023363	0.291318	0.650418	0.0349
3	0.00226	0.049369	0.638699	0.309672
4	0.000077	0.000792	0.089199	0.909932

6 -> 7	1	2	3	4
1	0.176471	0.535948	0.248366	0.039216
2	0.072893	0.463326	0.432802	0.030979
3	0.00833	0.147177	0.634497	0.209996
4	0.000167	0.003283	0.12191	0.87464

7 -> 8	1	2	3	4
1	0.25641	0.45584	0.273504	0.014245
2	0.104213	0.364747	0.502368	0.028671
3	0.01638	0.129932	0.663937	0.189752
4	0.000417	0.004805	0.142276	0.852503

2001-2006

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.079853	0.62285	0.291155	0.006143
2	0.007654	0.244332	0.715268	0.032746
3	0.000322	0.022229	0.622579	0.354871
4	0	0.000469	0.090551	0.90898

4 -> 5	1	2	3	4
1	0.117647	0.457516	0.405229	0.019608
2	0.027042	0.266594	0.642447	0.063917
3	0.00089	0.03122	0.47449	0.4934
4	0	0.000304	0.032729	0.966967

5 -> 6	1	2	3	4
1	0.113475	0.588652	0.297872	0
2	0.055351	0.344755	0.568793	0.031102
3	0.005871	0.093443	0.696743	0.203943
4	0.000095	0.002464	0.138264	0.859177

6 -> 7	1	2	3	4
1	0.159091	0.595455	0.209091	0.036364
2	0.067573	0.508712	0.398215	0.025499
3	0.008996	0.160285	0.629347	0.201371
4	0.000101	0.003206	0.129689	0.867005

7 -> 8	1	2	3	4
1	0.679558	0.281768	0.038674	0
2	0.450537	0.460818	0.08796	0.000685
3	0.147437	0.480277	0.369899	0.002387
4	0.007929	0.078791	0.555057	0.358223

2002-2007

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.063063	0.387387	0.524775	0.024775
2	0.003057	0.091712	0.79563	0.109601
3	0.000214	0.006322	0.409151	0.584313
4	0	0.000042	0.030201	0.969757

4 -> 5	1	2	3	4
1	0.131148	0.557377	0.278689	0.032787
2	0.049138	0.348276	0.566379	0.036207
3	0.002367	0.058923	0.61656	0.322149
4	2.48E-05	0.000818	0.080895	0.918263

5 -> 6	1	2	3	4
1	0.160714	0.553571	0.285714	0
2	0.058787	0.381188	0.535891	0.024134
3	0.007726	0.103957	0.703643	0.184673
4	0.000115	0.003205	0.156805	0.839875

6 -> 7	1	2	3	4
1	0.626556	0.340249	0.033195	0
2	0.441343	0.515372	0.043285	0
3	0.127831	0.537619	0.333122	0.001429
4	0.003071	0.072206	0.574326	0.350397

7 -> 8	1	2	3	4
1	0.379465	0.492194	0.128076	0.000265
2	0.149422	0.477023	0.369378	0.004178
3	0.013154	0.133471	0.704408	0.148967
4	0.000072	0.002169	0.194969	0.80279

2003-2008

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.102941	0.485294	0.367647	0.044118
2	0.012809	0.20262	0.741776	0.042795
3	0.000715	0.018018	0.541572	0.439695
4	0	0.000223	0.055671	0.944106

4 -> 5	1	2	3	4
1	0.151899	0.582278	0.253165	0.012658
2	0.060092	0.471495	0.44453	0.023883
3	0.005928	0.104239	0.650874	0.238959
4	0	0.002087	0.124279	0.873634

5 -> 6	1	2	3	4
1	0.538095	0.438095	0.02381	0
2	0.372069	0.565628	0.062303	0
3	0.106916	0.552619	0.33779	0.002674
4	0.004036	0.088179	0.566656	0.341129

6 -> 7	1	2	3	4
1	0.454062	0.470901	0.074742	0.000295
2	0.158073	0.518702	0.32191	0.001316
3	0.010167	0.138528	0.732159	0.119146
4	0	0.001248	0.228102	0.77065

7 -> 8	1	2	3	4
1	0.351716	0.500915	0.146224	0.001144
2	0.12632	0.444704	0.423734	0.005243
3	0.011159	0.120787	0.695662	0.172392
4	0.000068	0.001083	0.17275	0.826099

2004-2009

Transition Probability Matrices

3 -> 4	1	2	3	4
1	0.11215	0.523364	0.345794	0.018692
2	0.014013	0.290996	0.658915	0.036076
3	0.000356	0.035118	0.598612	0.365914
4	6.37E-05	0.00051	0.090434	0.908992

4 -> 5	1	2	3	4
1	0.676056	0.28169	0.042254	0
2	0.412776	0.52973	0.057002	0.000491
3	0.090009	0.520106	0.383997	0.005888
4	0.002209	0.076285	0.591966	0.32954

5 -> 6	1	2	3	4
1	0.352003	0.567889	0.079769	0.000339
2	0.108517	0.538835	0.34838	0.004268
3	0.00646	0.14022	0.691979	0.161341
4	0.000154	0.00216	0.223225	0.77446

6 -> 7	1	2	3	4
1	0.406368	0.510442	0.082506	0.000685
2	0.130281	0.494433	0.372588	0.002697
3	0.007179	0.122666	0.733658	0.136497
4	0	0.001053	0.218567	0.78038

7 -> 8	1	2	3	4
1	0.333536	0.503192	0.161751	0.00152
2	0.109272	0.44177	0.442578	0.00638
3	0.008646	0.111687	0.70698	0.172687
4	0.000125	0.001059	0.17885	0.819966

Figure D-1: One-step transition probability matrices for each cohort considered

Appendix E: Results of Markov Chains Based on Demographic Factors

The one-step transition probability matrices for each of the demographic factors considered are presented in Figures E-1 through E-4. Note, in these figures, the frequencies of each transition are shown for each level of each variable (n_x).

2000-2005 Cohort (n = 63,227 total records): Gender											
FEMALE (n_F = 33,855 records)						MALE (n_M = 29,372 records)					
Grade 3 to Grade 4						Grade 3 to Grade 4					
n_{F0}		1	2	3	4	n_{M0}		1	2	3	4
595	1	0.075630	0.598319	0.321008	0.005042	477	1	0.069182	0.572327	0.348008	0.010482
5909	2	0.008292	0.264343	0.713488	0.013877	4414	2	0.008156	0.255324	0.718849	0.017671
15160	3	0.000396	0.029551	0.796504	0.173549	12105	3	0.000413	0.039653	0.771499	0.188435
12191	4	0.000082	0.000410	0.232795	0.766713	12376	4	0.000000	0.000323	0.233516	0.766160
Grade 4 to Grade 5						Grade 4 to Grade 5					
n_{F1}		1	2	3	4	n_{M1}		1	2	3	4
101	1	0.158416	0.584158	0.237624	0.019802	74	1	0.189189	0.554054	0.256757	0.000000
2371	2	0.040067	0.413749	0.518768	0.027415	1884	2	0.048301	0.382696	0.532378	0.036624
19320	3	0.001449	0.047205	0.507557	0.443789	15568	3	0.002120	0.047983	0.483749	0.466149
12063	4	0.000000	0.000166	0.029843	0.969991	11846	4	0.000000	0.000338	0.032416	0.967246
Grade 5 to Grade 6						Grade 5 to Grade 6					
n_{F2}		1	2	3	4	n_{M2}		1	2	3	4
139	1	0.057554	0.568345	0.352518	0.021583	138	1	0.108696	0.507246	0.376812	0.007246
1954	2	0.021494	0.294268	0.651996	0.032242	1513	2	0.025777	0.287508	0.648381	0.038334
11420	3	0.001839	0.049912	0.642382	0.305867	8937	3	0.002797	0.048674	0.633994	0.314535
20342	4	0.000000	0.000541	0.087012	0.912447	18784	4	0.000160	0.001065	0.091567	0.907208
Grade 6 to Grade 7						Grade 6 to Grade 7					
n_{F3}		1	2	3	4	n_{M3}		1	2	3	4
71	1	0.169014	0.577465	0.183099	0.070423	82	1	0.182927	0.500000	0.304878	0.012195
1235	2	0.074494	0.460729	0.432389	0.032389	960	2	0.070833	0.466667	0.433333	0.029167
10429	3	0.008246	0.144501	0.642535	0.204718	8419	3	0.008433	0.150493	0.624540	0.216534
22120	4	0.000181	0.002893	0.124231	0.872694	19911	4	0.000151	0.003717	0.119331	0.876802
Grade 7 to Grade 8						Grade 7 to Grade 8					
n_{F4}		1	2	3	4	n_{M4}		1	2	3	4
194	1	0.268041	0.407216	0.309278	0.015464	157	1	0.242038	0.515924	0.229299	0.012739
2181	2	0.089409	0.347547	0.534617	0.028427	1830	2	0.121858	0.385246	0.463934	0.028962
9996	3	0.012905	0.125650	0.675670	0.185774	8075	3	0.020681	0.135232	0.649412	0.194675
21484	4	0.000465	0.004189	0.146667	0.848678	19310	4	0.000363	0.005489	0.137390	0.856758

Figure E-1: 2000-2005 cohort – One-step transition probability matrices by gender

2000-2005 Cohort (n = 63,227 total records): Ethnicity

ASIAN [A] (n_A = 1,080 records)

Grade 3 to Grade 4

n_{A0}		1	2	3	4
9	1	0.000000	0.666667	0.333333	0.000000
162	2	0.006173	0.179012	0.783951	0.030864
411	3	0.000000	0.021898	0.783455	0.194647
498	4	0.000000	0.002008	0.144578	0.853414

Grade 4 to Grade 5

n_{A1}		1	2	3	4
1	1	0.000000	1.000000	0.000000	0.000000
45	2	0.022222	0.244444	0.666667	0.066667
524	3	0.001908	0.022901	0.446565	0.528626
510	4	0.000000	0.000000	0.011765	0.988235

Grade 5 to Grade 6

n_{A2}		1	2	3	4
2	1	0.000000	0.000000	1.000000	0.000000
24	2	0.000000	0.125000	0.791667	0.083333
270	3	0.000000	0.033333	0.585185	0.381481
784	4	0.001276	0.000000	0.047194	0.951531

Grade 6 to Grade 7

n_{A3}		1	2	3	4
1	1	0.000000	0.000000	1.000000	0.000000
12	2	0.083333	0.500000	0.333333	0.083333
216	3	0.000000	0.078704	0.648148	0.273148
851	4	0.000000	0.002350	0.077556	0.920094

Grade 7 to Grade 8

n_{A4}		1	2	3	4
1	1	0.000000	1.000000	0.000000	0.000000
25	2	0.080000	0.440000	0.440000	0.040000
211	3	0.018957	0.075829	0.630332	0.274882
843	4	0.000000	0.001186	0.067616	0.931198

BLACK [B] (n_B = 17,874 records)

Grade 3 to Grade 4

n_{B0}		1	2	3	4
718	1	0.097493	0.604457	0.293872	0.004178
5427	2	0.010319	0.313249	0.669983	0.006449
8684	3	0.000691	0.053777	0.836366	0.109166
3045	4	0.000000	0.000985	0.392447	0.606568

Grade 4 to Grade 5

n_{B1}		1	2	3	4
132	1	0.196970	0.583333	0.219697	0.000000
2604	2	0.049539	0.433180	0.496160	0.021121
12305	3	0.003169	0.071272	0.573182	0.352377
2833	4	0.000000	0.000706	0.054712	0.944582

Grade 5 to Grade 6

n_{B2}		1	2	3	4
194	1	0.077320	0.546392	0.365979	0.010309
2084	2	0.029750	0.320537	0.619962	0.029750
8529	3	0.003400	0.068003	0.689647	0.238949
7067	4	0.000142	0.002972	0.177444	0.819442

Grade 6 to Grade 7

n_{B3}		1	2	3	4
107	1	0.205607	0.551402	0.205607	0.037383
1375	2	0.080000	0.481455	0.419636	0.018909
8499	3	0.011178	0.194140	0.638546	0.156136
7893	4	0.000380	0.006842	0.219435	0.773343

Grade 7 to Grade 8

n_{B4}		1	2	3	4
230	1	0.278261	0.460870	0.256522	0.004348
2425	2	0.117938	0.384330	0.478351	0.019381
7758	3	0.021011	0.170663	0.671178	0.137149
7461	4	0.001340	0.012867	0.256400	0.729393

HISPANIC [H] (n_H = 2,011 records)

Grade 3 to Grade 4

n_{H0}		1	2	3	4
50	1	0.020000	0.600000	0.380000	0.000000
474	2	0.006329	0.253165	0.723629	0.016878
1002	3	0.001996	0.026946	0.802395	0.168663
485	4	0.000000	0.002062	0.288660	0.709278

Grade 4 to Grade 5

n_{H1}		1	2	3	4
6	1	0.166667	0.833333	0.000000	0.000000
178	2	0.061798	0.348315	0.578652	0.011236
1306	3	0.002297	0.046708	0.519908	0.431087
521	4	0.000000	0.000000	0.047985	0.952015

Grade 5 to Grade 6

n_{H2}		1	2	3	4
15	1	0.066667	0.533333	0.333333	0.066667
128	2	0.031250	0.242188	0.703125	0.023438
807	3	0.001239	0.043371	0.657993	0.297398
1061	4	0.000943	0.000943	0.113101	0.885014

Grade 6 to Grade 7

n_{H3}		1	2	3	4
7	1	0.142857	0.428571	0.428571	0.000000
75	2	0.093333	0.506667	0.386667	0.013333
746	3	0.014745	0.131367	0.627346	0.226542
1183	4	0.000845	0.001691	0.158073	0.839391

Grade 7 to Grade 8

n_{H4}		1	2	3	4
20	1	0.250000	0.350000	0.350000	0.050000
141	2	0.070922	0.411348	0.453901	0.063830
687	3	0.008734	0.100437	0.671033	0.219796
1163	4	0.000860	0.009458	0.181427	0.808255

2000-2005 Cohort (n = 63,227 total records): Ethnicity (continued)

AMERICAN INDIAN [I] (n_I = 973 records)

Grade 3 to Grade 4

n_{I0}		1	2	3	4
31	1	0.064516	0.645161	0.290323	0.000000
212	2	0.009434	0.301887	0.674528	0.014151
474	3	0.000000	0.067511	0.765823	0.166667
256	4	0.003906	0.000000	0.417969	0.578125

Grade 4 to Grade 5

n_{I1}		1	2	3	4
5	1	0.000000	0.600000	0.400000	0.000000
116	2	0.051724	0.362069	0.543103	0.043103
622	3	0.009646	0.078778	0.562701	0.348875
230	4	0.000000	0.004348	0.126087	0.869565

Grade 5 to Grade 6

n_{I2}		1	2	3	4
12	1	0.083333	0.666667	0.250000	0.000000
95	2	0.010526	0.294737	0.652632	0.042105
444	3	0.002252	0.069820	0.662162	0.265766
422	4	0.000000	0.004739	0.146919	0.848341

Grade 6 to Grade 7

n_{I3}		1	2	3	4
3	1	0.000000	0.333333	0.333333	0.333333
69	2	0.043478	0.492754	0.391304	0.072464
421	3	0.009501	0.152019	0.631829	0.206651
480	4	0.000000	0.010417	0.170833	0.818750

Grade 7 to Grade 8

n_{I4}		1	2	3	4
7	1	0.142857	0.571429	0.285714	0.000000
104	2	0.105769	0.432692	0.432692	0.028846
376	3	0.010638	0.125000	0.667553	0.196809
486	4	0.000000	0.002058	0.238683	0.759259

MULTI-RACIAL [M] (n_M = 1,163 records)

Grade 3 to Grade 4

n_{M0}		1	2	3	4
23	1	0.043478	0.782609	0.173913	0.000000
182	2	0.005495	0.225275	0.736264	0.032967
544	3	0.000000	0.034926	0.773897	0.191176
414	4	0.000000	0.000000	0.270531	0.729469

Grade 4 to Grade 5

n_{M1}		1	2	3	4
2	1	0.000000	1.000000	0.000000	0.000000
78	2	0.051282	0.371795	0.538462	0.038462
671	3	0.001490	0.043219	0.487332	0.467958
412	4	0.000000	0.000000	0.026699	0.973301

Grade 5 to Grade 6

n_{M2}		1	2	3	4
5	1	0.200000	0.600000	0.200000	0.000000
60	2	0.000000	0.316667	0.650000	0.033333
380	3	0.002632	0.042105	0.647368	0.307895
718	4	0.000000	0.000000	0.086351	0.913649

Grade 6 to Grade 7

n_{M3}		1	2	3	4
2	1	0.000000	0.000000	1.000000	0.000000
38	2	0.052632	0.342105	0.605263	0.000000
348	3	0.005747	0.126437	0.689655	0.178161
775	4	0.000000	0.003871	0.153548	0.842581

Grade 7 to Grade 8

n_{M4}		1	2	3	4
4	1	0.250000	0.500000	0.250000	0.000000
60	2	0.133333	0.216667	0.600000	0.050000
384	3	0.010417	0.127604	0.658854	0.203125
715	4	0.000000	0.004196	0.141259	0.854545

WHITE [W] (n_W = 40,126 records)

Grade 3 to Grade 4

n_{W0}		1	2	3	4
241	1	0.016598	0.502075	0.460581	0.020747
3866	2	0.005691	0.190119	0.777548	0.026643
16150	3	0.000186	0.023158	0.757957	0.218700
19869	4	0.000000	0.000201	0.206452	0.793346

Grade 4 to Grade 5

n_{W1}		1	2	3	4
29	1	0.103448	0.413793	0.413793	0.068966
1234	2	0.028363	0.348460	0.569692	0.053485
19460	3	0.000565	0.032425	0.446763	0.520247
19403	4	0.000000	0.000155	0.026697	0.973148

Grade 5 to Grade 6

n_{W2}		1	2	3	4
49	1	0.102041	0.489796	0.387755	0.020408
1076	2	0.013011	0.242565	0.699814	0.044610
9927	3	0.001410	0.033646	0.593432	0.371512
29074	4	0.000000	0.000241	0.067242	0.932517

Grade 6 to Grade 7

n_{W3}		1	2	3	4
33	1	0.121212	0.575758	0.272727	0.030303
626	2	0.059105	0.421725	0.463259	0.055911
8618	3	0.005222	0.104549	0.628684	0.261546
30849	4	0.000097	0.002334	0.095238	0.902331

Grade 7 to Grade 8

n_{W4}		1	2	3	4
89	1	0.213483	0.449438	0.303371	0.033708
1256	2	0.080414	0.321656	0.556529	0.041401
8655	3	0.013287	0.097400	0.657770	0.231542
30126	4	0.000199	0.002788	0.113058	0.883954

Figure E-2: 2000-2005 cohort – One-step transition probability matrices by ethnicity

2000-2005 Cohort (n = 63,227 total records): Participation in School Lunch Program (Grade 8)

FREE LUNCH (n_{FL} = 18,460 records)

Grade 3 to Grade 4					
n_{FL0}		1	2	3	4
674	1	0.084570	0.612760	0.299703	0.002967
5177	2	0.010238	0.300560	0.679544	0.009658
9022	3	0.000665	0.051984	0.832299	0.115052
3587	4	0.000279	0.001115	0.395038	0.603568

Grade 4 to Grade 5					
n_{FL1}		1	2	3	4
117	1	0.188034	0.564103	0.247863	0.000000
2442	2	0.052007	0.433251	0.489353	0.025389
12646	3	0.003084	0.072118	0.577099	0.347699
3255	4	0.000000	0.001229	0.068510	0.930261

Grade 5 to Grade 6					
n_{FL2}		1	2	3	4
188	1	0.090426	0.510638	0.377660	0.021277
2040	2	0.029902	0.318627	0.619118	0.032353
8745	3	0.003316	0.064723	0.684391	0.247570
7487	4	0.000134	0.002938	0.170028	0.826900

Grade 6 to Grade 7					
n_{FL3}		1	2	3	4
108	1	0.194444	0.546296	0.212963	0.046296
1334	2	0.084708	0.478261	0.411544	0.025487
8592	3	0.010475	0.180168	0.645368	0.163990
8426	4	0.000593	0.008308	0.216710	0.774389

Grade 7 to Grade 8					
n_{FL4}		1	2	3	4
229	1	0.305677	0.423581	0.257642	0.013100
2315	2	0.121382	0.380562	0.473434	0.024622
7943	3	0.021277	0.167569	0.662974	0.148181
7973	4	0.000878	0.011790	0.253731	0.733601

REDUCED PRICE LUNCH (n_{RL} = 5,138 records)

Grade 3 to Grade 4					
n_{RL0}		1	2	3	4
96	1	0.041667	0.562500	0.395833	0.000000
1055	2	0.008531	0.271090	0.711848	0.008531
2578	3	0.000776	0.041505	0.823119	0.134600
1409	4	0.000000	0.000000	0.327182	0.672818

Grade 4 to Grade 5					
n_{RL1}		1	2	3	4
15	1	0.000000	0.600000	0.333333	0.066667
447	2	0.035794	0.355705	0.572707	0.035794
3372	3	0.001483	0.051305	0.529063	0.418149
1304	4	0.000000	0.000000	0.050613	0.949387

Grade 5 to Grade 6					
n_{RL2}		1	2	3	4
21	1	0.095238	0.476190	0.428571	0.000000
341	2	0.017595	0.260997	0.692082	0.029326
2111	3	0.002369	0.044055	0.660351	0.293226
2665	4	0.000000	0.001501	0.131332	0.867167

Grade 6 to Grade 7					
n_{RL3}		1	2	3	4
13	1	0.076923	0.692308	0.230769	0.000000
196	2	0.040816	0.464286	0.459184	0.035714
1989	3	0.009050	0.145802	0.637004	0.208145
2940	4	0.000340	0.001701	0.171429	0.826531

Grade 7 to Grade 8					
n_{RL4}		1	2	3	4
28	1	0.071429	0.750000	0.178571	0.000000
395	2	0.070886	0.397468	0.506329	0.025316
1864	3	0.020386	0.120172	0.674356	0.185086
2851	4	0.000702	0.009470	0.199579	0.790249

FULL PAY/SCHOOL NOT PARTICIPATING (n_{FP} = 39,629 records)

Grade 3 to Grade 4					
n_{FP0}		1	2	3	4
302	1	0.056291	0.536424	0.387417	0.019868
4091	2	0.005622	0.207040	0.762650	0.024688
15665	3	0.000192	0.022470	0.752186	0.225152
19571	4	0.000000	0.000255	0.196720	0.803025

Grade 4 to Grade 5					
n_{FP1}		1	2	3	4
43	1	0.186047	0.581395	0.209302	0.023256
1366	2	0.031479	0.355051	0.572474	0.040996
18870	3	0.000901	0.030419	0.437467	0.531214
19350	4	0.000000	0.000103	0.023514	0.976382

Grade 5 to Grade 6					
n_{FP2}		1	2	3	4
68	1	0.058824	0.632353	0.308824	0.000000
1086	2	0.012891	0.249540	0.696133	0.041436
9501	3	0.001263	0.036417	0.591832	0.370487
28974	4	0.000069	0.000173	0.064437	0.935321

Grade 6 to Grade 7					
n_{FP3}		1	2	3	4
32	1	0.156250	0.437500	0.375000	0.031250
665	2	0.058647	0.433083	0.467669	0.040602
8267	3	0.005927	0.113221	0.622596	0.258256
30665	4	0.000033	0.002054	0.091114	0.906799

Grade 7 to Grade 8					
n_{FP4}		1	2	3	4
94	1	0.191489	0.446809	0.340426	0.021277
1301	2	0.083782	0.326672	0.552652	0.036895
8264	3	0.010770	0.095958	0.662512	0.230760
29970	4	0.000267	0.002503	0.107174	0.890057

Figure E-3: 2000-2005 cohort – One-step transition probability matrices by participation in a school lunch program

2000-2005 Cohort (n = 63,227 total records): AIG Mathematics

AIG MATH (n_{AIG} = 10,945 records)

Grade 3 to Grade 4

n_{AIG0}		1	2	3	4
4	1	0.000000	0.250000	0.000000	0.750000
28	2	0.000000	0.035714	0.750000	0.214286
1098	3	0.000000	0.000000	0.398907	0.601093
9815	4	0.000000	0.000000	0.063882	0.936118

Grade 4 to Grade 5

n_{AIG1}		1	2	3	4
0	1	---	---	---	---
2	2	0.000000	0.000000	1.000000	0.000000
1086	3	0.000000	0.000000	0.090239	0.909761
9857	4	0.000000	0.000000	0.002130	0.997870

Grade 5 to Grade 6

n_{AIG2}		1	2	3	4
0	1	---	---	---	---
0	2	---	---	---	---
121	3	0.000000	0.000000	0.256198	0.743802
10824	4	0.000000	0.000000	0.004989	0.995011

Grade 6 to Grade 7

n_{AIG3}		1	2	3	4
0	1	---	---	---	---
0	2	---	---	---	---
85	3	0.000000	0.058824	0.505882	0.435294
10860	4	0.000000	0.000000	0.006814	0.993186

Grade 7 to Grade 8

n_{AIG4}		1	2	3	4
0	1	---	---	---	---
5	2	0.000000	0.200000	0.800000	0.000000
117	3	0.008547	0.025641	0.495726	0.470085
10823	4	0.000000	0.000000	0.010995	0.989005

NOT AIG MATH (n_{NOT} = 52,282 records)

Grade 3 to Grade 4

n_{NOT0}		1	2	3	4
1068	1	0.073034	0.588015	0.334270	0.004682
10295	2	0.008256	0.261098	0.715687	0.014959
26167	3	0.000420	0.035465	0.801620	0.162495
14752	4	0.000068	0.000610	0.345784	0.653539

Grade 4 to Grade 5

n_{NOT1}		1	2	3	4
175	1	0.171429	0.571429	0.245714	0.011429
4253	2	0.043734	0.400188	0.524571	0.031507
33802	3	0.001805	0.049080	0.509999	0.439116
14052	4	0.000000	0.000427	0.051452	0.948121

Grade 5 to Grade 6

n_{NOT2}		1	2	3	4
277	1	0.083032	0.537906	0.364621	0.014440
3467	2	0.023363	0.291318	0.650418	0.034900
20236	3	0.002273	0.049664	0.640986	0.307076
28302	4	0.000106	0.001095	0.121405	0.877394

Grade 6 to Grade 7

n_{NOT3}		1	2	3	4
153	1	0.176471	0.535948	0.248366	0.039216
2195	2	0.072893	0.463326	0.432802	0.030979
18763	3	0.008368	0.147578	0.635080	0.208975
31171	4	0.000225	0.004427	0.162010	0.833339

Grade 7 to Grade 8

n_{NOT4}		1	2	3	4
351	1	0.256410	0.455840	0.273504	0.014245
4006	2	0.104343	0.364953	0.501997	0.028707
17954	3	0.016431	0.130612	0.665033	0.187925
29971	4	0.000567	0.006540	0.189683	0.803210

Figure E-4: 2000-2005 cohort – One-step transition probability matrices by AIG mathematics status