

EFFECTS OF ADAPTIVE LEARNING TECHNOLOGIES ON MATH
ACHIEVEMENT: A QUANTITATIVE STUDY OF
ALEKS MATH SOFTWARE

A DISSERTATION IN
Education

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF EDUCATION

by
BURAK YILMAZ

M.Ed., University of Central Oklahoma, 2008

B.S., Bilkent University, 2005

Kansas City, Missouri
2017

© 2017
BURAK YILMAZ
ALLRIGHTS RESERVED

EFFECTS OF ADAPTIVE LEARNING TECHNOLOGIES ON MATH ACHIEVEMENT:
A QUANTITATIVE STUDY OF ALEKS MATH SOFTWARE

Burak Yilmaz, M.Ed., Candidate for the Doctor of Education Degree
University of Missouri – Kansas City, 2017

ABSTRACT

The intent of this study is to investigate the effects of a particular adaptive math learning program (ALEKS) on math achievement and its impact on closing the achievement gap in math performance of middle school students. The study is conducted in two small urban school districts in a southern city. The study is a quasi-experimental research design with a sample size of 1110 students in grades fifth through ninth forming a control group and a treatment group of equal sizes. The data is compiled from the 2014-2015 school year (archived data) and has been analyzed using analysis of covariance (ANCOVA) to compare mean scores of the two groups from a norm-referenced test and regression analysis to understand various categorical variables affecting math achievement on ALEKS.

APPROVAL PAGE

The undersigned, appointed by the Dean of the School of Education, have examined a dissertation titled “Effects of Adaptive Learning Technologies on Math Achievement: A Quantitative Study of ALEKS Math Software” presented by Burak Yilmaz, candidate for the Doctor of Education degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee:

Loyce Caruthers, Ph.D., Committee Chair
Department of Educational Leadership, Policy and Foundations

Shirley McCarther, Ph.D.
Department of Educational Leadership, Policy and Foundations

Robert Leachman, Ph. D.
Department of Educational Leadership, Policy and Foundations

Yuankun Yao, Ph. D.
Department of Educational Foundations & Literacy
University of Central Missouri

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF ILLUSTRATIONS	viii
LIST OF TABLES	ix
ACKNOWLEDGEMENTS	xi
Chapter	
1. INTRODUCTION	1
Adaptive Online Learning.....	1
Problem Statement	3
Purpose of Study	6
What is ALEKS?.....	6
What is NWEA MAP?	7
Research Questions & Hypotheses	8
Theoretical Framework.....	9
Significance of the Study	10
Organization of Subsequent Chapters.....	12
2. REVIEW OF LITERATURE	14
Current Landscape of Education in American Public Schools.....	14
The Skills Gap and Industry Expectations.....	15
Mathematics Instruction in the U.S.	16
Standardized Testing & Accountability.....	20

Personalized Learning Initiatives.....	20
Blended Learning.....	20
Adaptive Learning Technologies.....	21
Knowledge Space Theory.....	22
Adaptive Control of Thought.....	27
Cognitive Science & Intelligent Tutoring Systems	30
Nexus between K-12 Personalized Learning and Post-Secondary Education	47
3. METHODOLOGY	55
Participants.....	55
Data Considerations	56
Design.....	56
Data Collection.....	56
Measures	57
Procedure	59
Analysis.....	60
Ethical Considerations	64
4. RESULTS OF ANALYSES AND CONCLUSIONS.....	66
Characteristics of Treatment and Control Groups	66
Statistical Analyses & Data Tables.....	71
Hypothesis Testing.....	77

Performance Analyses of Demographic Groups.....	81
Conclusions and Summary	86
5. DISCUSSION.....	88
Discussion of Findings.....	93
Summary.....	94
Conclusions.....	98
Implications.....	100
Limitations	101
Recommendations for Future Research	102
Concluding Thoughts.....	103
REFERENCES	107
VITA.....	121

LIST OF ILLUSTRATIONS

Figure	Page
4.1. Distribution of independent variable fall RIT score by treatment/control group.....	36
4.2. Distribution of dependent variable spring RIT score by treatment/control group.....	61
4.3. Scatter plot of dependent variable spring RIT score and covariate fall RIT score.....	63

LIST OF TABLES

Table	Page
1.1. Definition of Key Terms	12
4.1. Control/treatment group participation by grade level.....	66
4.2. Control/treatment group participation by race	67
4.3. Control/treatment group participation by sex	67
4.4. Control/treatment group participation by special – education status.....	68
4.5. Control/treatment group participation by economic disadvantage.....	68
4.6. Control/treatment group participation by Limited English Proficiency status.....	69
4.7. Control/treatment group participation by Gifted/Talented status.....	69
4.8. Correlations among selected variables used in the study	70
4.9. Descriptive statistics for dependent and independent variables used in the study, by treatment/control group.....	72
4.10. Results of ANCOVA analysis, test of hypothesis	76
4.11. Results of ANCOVA analysis, mean spring mathematics RIT score by group.....	76
4.12. Test for homogeneity of variance.....	77
4.13. Regression model specification.....	78
4.14. Test for homogeneity of variance.....	79
4.15. ANOVA results associated with the regression model	79
4.16. Coefficients resulting from the regression model.....	80
4.17. Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by grade level	81
4.18. Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by special populations	82

4.19. Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by race/ethnicity	84
4.20. Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by gender	85

ACKNOWLEDGEMENTS

I would never have completed this dissertation without the guidance of my advisor and committee members, assistance of friends, and most importantly love, support, and encouragement of my family, especially my beloved wife.

I would first like to thank Dr. Loyce Caruthers for her support and guidance throughout the development of this dissertation. Her timely and critical feedback was much appreciated and pushed me to produce quality work. I am also extremely thankful for her encouragement and individual support especially after I moved away to Texas. I could not have asked for a better advisor. I would like to also thank Dr. Yuankun Yao for his valuable insight in my methodology and analyses.

Next I would like to thank my friends John, Sean, and Ruili for providing reviews, suggestions, and inputs into my work. Their contributions made this study stronger. I am especially thankful to my classmate and former boss Sean, who encouraged and even insisted that I start this doctorate program. Without his constant encouragement and persistence, I may never have started this program.

I would also like to express my deepest gratitude to my parents, who –despite being oceans apart– have always supported my graduate studies and constantly checked on me and motivated me to pursue my goals. I am thankful for their high expectations and vision for my future.

Last, but by all means not least, I would like to dedicate this dissertation to my family. Without my beautiful wife Kubra, and my adorable children Pinar and Tarik, I would never have found the strength and motivation in myself to pursue this degree and finish this dissertation. My entire purpose and motivation to pursue life-long learning and earn my

doctorate lies within my desire to become the best husband and father, and become a model educator and leader to my family, first and foremost. I am forever grateful to my wife and children, for always being there for me and cheering me on.

CHAPTER 1

INTRODUCTION

Math has traditionally been difficult to understand, a dreaded subject area in K-16 for many students. When educators talk about achievement gaps among ethnic groups, math is definitely at the center of those discussions. A significant number of college freshmen must take remedial math courses, despite having passing scores on state standardized tests or making good grades on their high school math coursework (Harper & Reddy, 2013; Robathan & Wilson, 2011). With advancements in computer and network technologies, new trends and methods for teaching and learning mathematics have emerged, especially web-based assessment and tutoring systems built on artificial intelligence (Albert & Hockemeyer, 1997; Doignon & Falmagne, 1999; Falmagne, Cosyn, Doignon, & Thiery, 2006). Despite this, attrition rates in math courses, online or classroom-based, are still higher than courses of other disciplines (Shakerdge, 2016; Varsavsky, 2010). Even with the support of adaptive learning technologies and intelligent tutoring systems, many students still fail, withdraw from or perform poorly in their math courses. Therefore, there is a clear need for evaluation of computer-based adaptive learning software to determine their effects on learning and math achievement.

Adaptive Online Learning

Computer based learning environments in mathematics have evolved rapidly in the past decade. It started with programs that provide drills and tutorials based on behaviorist learning philosophy; then continued with using games and simulations in the learning process advocated by cognitive learning theory; and finally lead to the extensive use of hypertext and

hypermedia that promote constructivist learning approaches (Anderson, 1970; Baker, 1968; McKenney & Dill, 1966; Albert & Hockemeyer, 1997).

Cognitive tutoring and assessment systems built on artificial intelligence opened a new wave of reform in public education. In the last two decades, smart algorithms started helping teachers identify student knowledge gaps and create learning spaces; an especially effective method for teaching mathematics (Aleven & Koedinger, 2001; Doignon & Falmagne, 1999; Falmagne & Doignon, 2011). Such advancements lead to digital learning reforms aimed at personalizing learning for all students. With personalization in mind, local districts began putting individual devices into the hands of students as early as first grade in an attempt to engage learners through digital content providers and adaptive instructional software, thus allowing students to progress at their own pace using a mastery based approach.

Online instructional delivery by means of an interactive device such as a computer or tablet in order to engage students with content tailored to student learning needs forms the basis of adaptive learning. “How well it adapts to the individual is entirely dependent on the sophistication of the software that drives the device” explains Vander Ark (2013). Education expert Tom Vander Ark goes on to explain how smart learning algorithms are starting to keep user experience in mind when delivering just-right content, just in time:

The same sophisticated, predictive, intelligent use of data that has accustomed us to personalization in online shopping and music, and show us content we are mostly likely to appreciate and potentially use, has come to learning and education. Smart instructional content adjusts its path based on response to questions. Like computer games, adaptive systems calibrate the difficulty to maintain an appropriate level of challenge. But there is a level beyond shopping and gaming where an entirely new class of adaptive learning software exists, and it is called Intelligent Adaptive Learning. (Vander Ark, 2013, p. 8)

Intelligent adaptive learning holds promise for robust personalization in education, specifically most beneficial for high need students in public schools. These new learning models can be a cure for closing the persistent achievement disparities in the U.S.

Problem Statement

Public education in America has long suffered from disparities of achievement among ethnic groups. According to the Coleman Report of 1966, cited in Camera (2016), average black student in grade 12 ranked in the 13th percentile of the score distribution in both math and reading, outperformed by 87% of their white counterparts in grade 12. Data from 2013 National Assessment for Educational Progress Education placed average black twelfth grader in the 19th percentile in math, a slight improvement in nearly half a century showing that achievement gaps still persist (Camera, 2016). Agencies both at the federal and state levels have made countless attempts to reform public education with a single goal in mind: closing the achievement gap. Starting with *No Child Left Behind Act* and followed by *Race to the Top* initiatives, education reformers placed heavy emphasis on improving math and reading instruction, and promoted personalized learning models. These legislative initiatives (“Every Student Succeeds Act (ESSA)”, 2017; “No Child Left Behind”, 2017; “U.S. Department of Education”, 2017) put federal and state mandates on schools to make adequate yearly progress on accountability measures, and ensure that all students perform ‘proficient’ on standardized state assessments. Because students in the very same classroom begin at different levels, and hence have very different needs, differentiation of instruction, and even more importantly, personalizing the learning experience for each student has become necessary. With the pressure of performance targets on standardized tests, schools and districts are designing or adopting new programs and tools to boost student performance on

state assessments. In recent years, the use of computer-based adaptive learning programs has been on the rise, and schools are purchasing and using these programs in hopes that they will help increase test performance (Dickard, 2003).

Traditional instructional approaches have not been successful in narrowing and closing the achievement gaps in mathematics between black and white students, as well as economically disadvantaged and affluent students. The demographic makeup of American public schools is closely related to the achievement disparities among racial and socio-economic groups, particularly with growing resegregation of schools in recent decades (Frankenberg, Lee, & Orfield, 2003; Orfield, Kucsera, & Siegel-Hawley, 2012). Based on data from National Assessment of Educational Progress (NAEP) in 2011, “on average, White students attended schools that were 9 percent Black while Black students attended schools that were 48 percent Black, indicating a large difference in average Black student density nationally” (Bohrnstedt, Kitmitto, Ogut, Sherman, & Chan, 2015, p. 1). Student performance analysis showed that math achievement was significantly lower in the highest black density schools for both black and white students than in the lowest black density schools, while the achievement gap in math remained the same between black and white students in the same schools regardless of black density level (Bohrnstedt et al., 2015).

Some researchers identified teacher quality as the root cause of achievement disparities in math (Flores, 2007; Kane & Staiger, 2005), while others argued that poverty is a more significant cause than teacher quality (Haberman, 1991; Marder, 2012). Indeed, Haberman (1991) claimed that recruiting good teachers is the solution for closing achievement gaps caused by poverty. Marder (2012) reported that educational performance

was strongly correlated with poverty. Based on SAT/ACT math scores of Texas high school students in 2010 and college readiness criterion, Marder (2012) concludes:

Among schools where less than 15% of the students are eligible for free and reduced meals, there are virtually none where fewer than 20% of the students graduate college-ready. Conversely, among schools where more than 85% of the students are eligible for free and reduced meals, there are none where more than 20% of the students graduate college-ready. In short, the least successful schools serving the wealthy do better than the most successful schools serving the poor. (p. 15)

Although new edtech products can seem appealing to educators right away, the problem becomes testing the adaptivity and curricular alignment of these programs and determine whether they are able to contribute to math achievement measured by standardized assessments. Teachers and administrators should ask themselves this very question: Is our educational technology moving the needle in the classroom? The U.S. Department of Education is recommending schools and districts to use a new approach to evaluating new programs called rapid-cycle evaluations prior to making large scale decisions on edtech acquisition. Rapid cycle evaluation offers educational leaders a low-cost, quick-turn-around evaluation option to make evidence-based decisions (“Mathematica Policy Research”, 2017). School districts need timely and reliable evidence on the effectiveness of new innovative programs to be able to determine whether it is the right fit for the need. However, anecdotal experience, references, and marketing materials are typically the only data available to administrators for decision-making when purchasing a new edtech product or implementing a new curricular program. Because education technology is a rapidly changing and developing field, traditional research and evaluation approaches do not meet the needs of school districts as they take too long and cannot usually keep up with new developments (“Office of Educational Technology”, 2017). Theoretically, with the personalized learning path adaptive learning programs create to cater to the needs of individual students, they seem promising in

bringing success to all students. The problem being investigated in this study is whether or not adaptive learning programs are able to make significant contributions to learning outcomes in mathematics in reality.

Purpose of Study

The purpose of this research study is to investigate the effects of a particular adaptive math learning software (ALEKS) on math achievement and its impact on closing the achievement gap in math performance of middle school students in two public charter school districts in a large urban southwestern city. This research will help educators assess how similar digital learning tools can help create personalized learning environments for middle school students.

An Intelligent Tutoring System (ITS) is defined as a computer based program that has problem-solving capabilities, which can identify users' current knowledge and skills, and can help users close the achievement gap between themselves and the software itself. ALEKS is a perfect example of an ITS, and is widely used in teaching and learning of math across many K-12 school districts and higher education institutions.

What is ALEKS?

This research is aimed at studying the effectiveness and impact of adaptive learning programs built on artificial intelligence and how they can support personalized learning for students in mathematics. For the purposes of this research, I chose to study ALEKS (Assessment of Learning in Knowledge Spaces), a web-based intelligent tutoring system (ITS) that is widely used at both K-12 and college levels, particularly for math ("ALEKS", 2017; Carpenter & Hanna, 2006; Harper & Reddy, 2013; LaVergne, 2007; Taylor, 2008). The adaptive questioning feature of ALEKS is able to determine quickly and accurately what

a student knows and does not know in a math course. The program then creates an instruction plan to teach students math topics they are most ready to learn based on the precedence relation among topics. As a student works through a course in ALEKS, the program periodically reassesses the student to ensure that topics learned are also retained over time. Another great feature of ALEKS is that it avoids multiple-choice questions, and instead uses flexible answer input tools, which is equivalent to what students would traditionally need to do with paper and pencil to work out a problem.

ALEKS is an artificially intelligent assessment and learning system that is built on Knowledge Space Theory (KST), which is the theoretical framework that enables a computerized algorithm to identify the knowledge state of a learner and create a learning space individualized to the needs of that particular student (Falmagne, Cosyn, Doignon, & Thiery, 2006). KST reveals the knowledge state of the individual and determines the best learning path based on what the learner is capable of doing in math (Doignon & Falmagne, 2011).

What is NWEA MAP?

Measures of Academic Progress (MAP) is a norm-referenced assessment developed by non-profit Northwest Evaluation Association (NWEA). MAP creates a personalized assessment experience by adapting to each student's learning level in order to measure student progress and growth for each individual. NWEA MAP assessments are used by many districts across the nation as they provide valuable data on student percentile ranks, student growth, and instructional gains ("NWEA", 2017). The districts participating in this study administers the MAP tests twice a year, once at the start of school and once again at the end of school year.

Research Questions & Hypotheses

This study intends to address the following two broad questions:

- 1- Does use of ALEKS math software improve student achievement in mathematics?
- 2- Does increased time spent with ALEKS correlate with improved student achievement in mathematics?

For the first broad question, I will compare student performance on NWEA MAP test, which is a norm-referenced adaptive assessment that is used in participating districts as a summative assessment, between ALEKS users and non-ALEKS users.

In order to tackle the second question, I will need to measure time spent within ALEKS and study the relationships between time-spent and measured subsequent achievement levels. Below are more specific questions I will seek answers to, when tackling the second question:

- What is the relationship between time spent in ALEKS and latest test performance?
- What is the relationship between concept mastery and latest test performance?

These questions might help me uncover the best learning indicator for teachers to use for instructional purposes, especially when teachers are forming small groups to reteach certain topics based on common areas students struggle in math.

In this study, I have formed three hypotheses as follows:

H_{01} : There is no statistically significant difference in mean spring NWEA mathematics scores between students who received a regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₂: Among ALEKS users, there is no statistically significant relationship between time spent on ALEKS and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₃: Among ALEKS users, there is no statistically significant relationship between pie mastery percentage and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$. The ALEKS pie shows the amount of topics mastered in the course.

In order to test the first hypothesis, spring RIT scores from the NWEA MAP math assessment will be used as the dependent variable. Previous (fall) RIT scores from the NWEA MAP math assessment will be used as a covariate to control for prior skill level. The independent variable will be ALEKS use (i.e., ALEKS users as treatment group, and non-users as the control group).

To test the second and third hypothesis, a two – block regression model will be used. In the first block, previous (fall) RIT scores from the NWEA MAP math assessment will be used as a control variable and regressed onto spring RIT scores (dependent variable). In the second block, the significance of time spent on ALEKS instruction and PIE mastery percentage will be used to test the hypotheses.

Theoretical Framework

Many adaptive math software programs, including ALEKS, are built on Knowledge Space Theory (KST), which constitutes the theoretical framework of this study. Falmagne and Doignon (1999) developed the knowledge space theory and explained the science behind ALEKS through KST. This theory makes it possible to uncover the knowledge state of a particular student in a particular math topic through an online assessment. Knowledge state is

defined as the complete set of problems in a particular topic that a student is able to solve (Falmagne & Doignon, 1999). Based on this basic principle, Falmagne and Doignon (1999) explains, an artificially-intelligent adaptive assessment in ALEKS creates two shortlists of problems and concepts that guide students and teachers on what each student can do and what he/she is ready to learn. These two lists uncover the complete knowledge state of an individual student being assessed and then their learning space is mapped out accordingly (Doignon & Falmagne, 2011). The connectivity of student responses made possible by KST allows ALEKS to track thinking patterns of the students in learning the math content (Taagepera & Noori, 2000). As a result, knowledge space theory provides new means for assessing the cognitive organization of student knowledge, which enables educators to deliver instruction with greater insight (Falmagne, Cosyn, Doble, Thiery, & Uzun, 2007).

Significance of the Study

Given the contemporary nature of merging software development with cognitive science, there has been very limited previous research in this field. Lead researchers in this field mainly addressed theoretical aspects of advanced computer tutoring in an attempt to introduce the phenomenon to educators and touched on achievement gains through cognitive tutors in a broad sense as case studies (Anderson et al., 1995, & Koedinger and Alevan, 2007). There are numerous adaptive math software programs in the market that are built on artificial intelligence (“ALEKS”, 2017; “DreamBox Learning”, 2017; “MIND Research Institute”, 2017; “Reasoning Mind”, 2017). Many of them can simulate responses of a human tutor and provide tutoring, hints, explanations, and feedback on problem solving to students. These programs are also designed to help students progress at their own pace through a competency based approach, which means students are continually assessed on a particular

objective until they show evidence of mastery based on their performance before moving on to the next learning target.

Many entrepreneurs in the field of education used aforementioned principles and theories to develop adaptive learning software in math and they have been widely in use in K-16 education institutions (Horn & Staker, 2014; Vander Ark, 2012). Advancements in learning technologies and software development have been so rapid and many school districts have been pushed to implement personalized education reforms and adopt such adaptive learning software quickly with federal and state mandates, without a chance to see research based pilot programs first. Shortly, the market of edtech tools is growing rapidly and more and more districts are adopting these programs every year. Unfortunately, research in this area has not been able to keep up with the fast paced advancements and quick changes in implementation. Although there is a clear need for further research to study effectiveness of these learning technologies, much of the existing evaluations come from companies who develop them, which questions the reliability of those studies and brings up ethical concerns (Krimski, 2012; Pigott, Polanin, Valentine, Willians, & Canada, 2013). Furthermore, investors, developers, and entrepreneurs spend more time and money on advertising these products rather than conducting evaluations and research to test and improve them. Much of the existing evaluations conducted has a single goal of presenting the product with a “research proven” label and thus increasing its market value, which ultimately brings in more customers and more profit. Therefore, there is an immediate need for an investigation on these adaptive learning programs to be conducted by an independent researcher with no ties to the developers and/or vendors. Majority of the previous independent research on ALEKS and other adaptive learning technologies have been conducted with college student samples,

and very few research conducted on K-12 student samples (“ALEKS”, 2017). A thorough experimental research with participants from K-12 school districts would allow educators to see the true value of these programs and help them make informed decisions before they spend tax-payer dollars from their already constrained budgets.

Table 1.1 lists key technical terms along with their common abbreviations and basic definitions. These terms will be referenced many times throughout the literature review and they can also be encountered in other scholarly articles on adaptive learning technologies and intelligent tutoring systems.

Table 1.1.

Definition of Key Terms

Term	Abbreviation	Definition
Adaptive Control of Thought	ACT	A network model that takes into account all of our human cognition such as language, learning, decision-making, and so on.
Artificial Intelligence	AI	The human-like intelligence exhibited by machines or software. The science and engineering of making intelligent machines, especially intelligent computer programs.
Assessment of Learning in Knowledge Space	ALEKS	A web-based, artificially intelligent assessment and learning system.
Cognitive Tutor	CT	A particular kind of intelligent tutoring system that utilizes a cognitive model to provide feedback to students as they are working through example problems.
Cognitive Science		The interdisciplinary scientific study of the mind and its processes.
Computer Assisted Instruction	CAI	A program of instructional material presented by means of a computer or computer systems.
Computer Based Education	CBE	The use of electronic media and information and communication technologies (ICT) in education.

Depth of Knowledge	DOK	Represents the comparison of the cognitive demand of learning standards and the cognitive demand of the assessments given to determine their mastery.
Hypermedia Based Instruction	HBI	Instruction based on electronic literature including but not limited to audio, video, plain text, hyperlinks, and other multimedia tools.
Intelligent Tutoring System	ITS	A computer system that aims to provide immediate and customized instruction or feedback to learners, usually without intervention from a human teacher.
Knowledge Space Theory	KST	A set-theoretical framework, which proposes mathematical formalisms to operationalize knowledge structures in a particular domain.
Practical Algebra Tutor	PAT	An intelligent tutoring system for algebra problem solving.

Organization of Subsequent Chapters

Chapter two of this study is a review of the literature that brings together rich discussions on relevant research topics, including an overview of adaptive learning technologies. Chapter 3 presents the research design for this study and describes the statistical methodology used to perform hypothesis testing. Chapter 4 presents the results of the study along with data analyses. Finally, chapter 5 is a discussion of findings, conclusions, and implications of the study, and it concludes with limitations of current study and recommendations for further research.

CHAPTER 2

REVIEW OF LITERATURE

This chapter is organized to present a thorough review of relevant literature in the field being studied, in order to establish the significance of the research and to provide a benchmark for assessing the results of this study with comparable research in the field. In an effort to establish the context for the current literature in this field, the chapter starts with a brief discussion of select topics in the current landscape of American public schools, and continue with a more extensive discussion of trends and reform efforts in public education. The latter broader topic includes the conceptual framework of this study, Knowledge Space Theory (KST), as well as online Intelligent Tutoring Systems (ITS), mainly ALEKS. The literature review includes a wide range of studies on ALEKS and its effects on mathematics instruction in both secondary and post-secondary education institutions.

Current Landscape of Education in American Public Schools

American public education system was designed based on a Prussian factory model of education introduced by Horace Mann in late 1800s. This industry model of education was the right fit for decades in the twentieth century, because graduates would take manufacturing jobs at factories that did not require higher order skills. The world has changed dramatically since then. Advancements in technology, robotics, and automation has caused a shift towards a knowledge economy. The job markets have changed dramatically, so did the expectations of what a high school or college graduate must know and be able to do. Despite all these changes, our education system has remained almost the same for nearly 130 years.

The Skills Gap and Industry Expectations

Math skills have always been vital to employers for successful job performance. Even lower level or mid-level jobs require some kind of math skills in order to operate high-tech equipment (Agondi, Harris, Atkins-Burnett, Heaviside, Novak, & Murphy, 2009). Many states have invested in high school courses under career and technical education to make sure students leave high school with necessary mid-skills to be successful in such careers. An increasing number of business and industry leaders reported dissatisfaction with new generation workers coming out of high schools or two year colleges lacking foundational math skills to perform above average at their jobs (Vincent, 2005).

It is also known that United States has been lagging behind many other European and Asian countries in international student assessments in the areas of math and science. It has been common over the past decade for researchers and the popular press to report the deficits in U.S. student achievement relative to many other nations (Kerr, 2015). For example, Pew Research reports that the Programme for International Student Assessment (PISA), a comparative international test of mathematics science and literacy skills, ranked the United States 38th among 71 countries in mathematics, and only 24th in science. The Organization for Economic Cooperation and Development (OECD), a 35 – nation group that sponsors PISA testing, ranked the US 30th in mathematics and 19th science. (Desilver, 2017). Kerr (2016) reported in the Associated Press (AP) that the Trends in International Mathematics and Science Study, (TIMSS), which administers tests every four years in many countries, that the United States placed 10th in fourth grade science and eighth grade mathematics. Kerr also reported that American students placed 14th in fourth grade mathematics on the TIMMS, behind Portugal and Kazakhstan. Employer dissatisfaction and the increasing

negative distance between American student achievement and students of other countries are the reasons that led educational policy makers to design new and more rigorous learning standards, especially in math, which subsequently led to the adoption of common core standards by 46 states and District of Columbia in the United States (Coburn, Hill & Spillane, 2016; Watt 2015).

Mathematics Instruction in the U.S.

Many researchers also argue that in addition to a viable, coherent curriculum, it is also equally important how such a curriculum is delivered to students (Schmoker, 2011; Marzano, Waters, & McNulty, 2005). That is, we should focus both on what we teach and how we teach. This argument brings up the question of what effective instructional tools and strategies must be implemented to be able to ensure a highly effective delivery of the curriculum. Furthermore, the delivery of instruction is supposed to make sure that all student groups (based on demographics and educational needs) master the content and skills before the teacher can move on. American public education has long been suffering from disparities of achievement among ethnic groups. Education agencies both at the federal and state levels have made countless attempts to reform public education with a single goal in mind: closing the achievement gap. The U.S. Department of Education defines the term achievement gap as follows:

Achievement gap: The difference in the performance between each ESEA subgroup (as defined in this document) within a participating LEA or school and the statewide average performance of the LEA's or State's highest achieving subgroups in reading/language arts and mathematics as measured by the assessments required under the ESEA. (U.S. Department of Education, 2012)

Starting with *No Child Left Behind* Act and followed by *Race to the Top* initiatives, education reformers put heavy emphasis on improving math and reading instruction and

promoted personalized learning models. With personalization in mind, technology and digital learning came into picture. Now more than ever, local districts started putting individual devices into the hands of students as early as first grade in an attempt to engage learners through digital content providers. The role of the teacher shifted from provider of information to facilitator of learning (Juan et al., 2011).

According to National Assessment of Educational Progress (NAEP), also known as the nation's report card, only 39% of fourth-graders, 34% of eighth-graders, and 23% of twelfth-graders score proficient in math. These statistics confirm that the longer students are in school, the wider the achievement gap is in math. This impacts higher education negatively as well. The demand for remedial math courses is very high in colleges. Since many students enter into higher education with critical deficiencies in mathematics, Robathan and Wilson (2011) state that universities are obligated to rectify those deficiencies by means of remedial course offerings. According to a report by Jobs for the Future, 60% of community college students end up taking at least one remedial math course before they are allowed to enroll in college-level courses. This is simply because many high school graduates arrive at college lacking not just basic algebra but also basic arithmetic. Varsavsky (2010) confirms that one of the main reasons why students drop out of college is their lack of mathematical skills.

Minimizing drop out and failure rates have become a major priority for university administrators across the U.S. (Gury, 2011). There is a strong need for evidence-based innovative programs and interventions, especially in math as Kezar (2011) points out, to improve student skills for college level requirements and college persistence rates. Padilla-Oviedo, Mundy, and Kupczynski (2016) suggest a blended learning model for improving

mathematics instruction that effectively complements the learning environment with a computer-based component. There has been slow but steady shifts in the educational system from traditional methods to technology-assisted instruction (Gano, 2011). The advancements in technology, computer science, and internet connectivity enabled effective utilization of technology tools in education, which has transformed the outlook of instruction in American classrooms (Juan, Steegmann, Huertas, Martinez, & Simosa, 2011). These trends are further discussed in the blended learning section of this literature review.

All of these facts point to a significant problem in our math education. Mathematics and policy experts at the National Mathematics Advisory Panel acknowledge that math education is broken and must be fixed in the United States, and they list the following six issues to be addressed (Hechinger Report, 2010):

1. Standards: There are too many content standards that are scattered and repetitive. Teachers need clear guidance on what is important. Many higher achieving countries focus on a few set of standards each year and expect students to study those in greater depth.
2. Curriculum: Elementary and middle school curricula must be redesigned to build a strong foundation in whole numbers, fractions, measurement, and basic geometry. The goal should be to prepare students to take algebra by eighth grade.
3. Graduation Requirements: Increased requirements must align with student achievement and preparedness as measured by SAT and end-of-course exams.
4. Textbooks: American textbooks are too big and cover a wide range of topics. Teachers are unable to get through the textbooks and thus sample through the

material without a coherent plan. Moreover, the textbooks address merely definitions and formulas and do not go beyond the mechanics of mathematics.

5. Teaching: There is no consensus among researchers about what good math teaching is. Teachers must have a deep knowledge of the math content they are teaching. Teacher preparation programs must incorporate more math content into their design. Schools and districts must support less experienced math teachers with strong mentorship and professional development.
6. Culture: American society has accepted a culture of inadequate expertise and fluency in math. It is ok for many students to say “I hate math.” Many parents do not seem to mind that their children lack basic math literacy.

There are obvious roadblocks for students to understand math. Many students experience math anxiety and teachers feel pressured to cover material too quickly. Korbey (2013) stated that math teachers are strapped for time and resources to be able to explore the beauties of math in meaningful ways. The way textbooks are designed and written as well as pressures to meet curriculum pacing requirements are among main reasons for this problem (Korbey, 2013). More students can understand and appreciate math if teachers are granted more time and freedom.

Personalized learning approaches advocate for connecting learning to student interests, which helps student retain information for longer periods of time. According to Schwartz (2013), this approach might work specifically well with math since many students dislike the subject for not seeing its relevance to their lives. In many classrooms today math is not being applied to the student’s world in a meaningful way. One strategy to change that

is tailoring questions to individual student interests when teaching difficult and abstract math concepts in particular (Schwartz, 2013).

Standardized Testing & Accountability

In recent decades, American public schools have focused heavily on standardized testing, which sometimes even supersedes instruction in importance. This prioritization has led some schools and teachers to feel pressured to “teach to the test” rather than implement authentic lesson plans to make learning more meaningful and enjoyable for students. The achievement gap among different student groups (based on demographics and educational needs) urges educators to find creative ways to use data in decision making to inform their instruction and make it more targeted. Teachers are expected to use formative and interim assessments more than ever to constantly assess student skills and knowledge. Critics who oppose testing argue that the more focus on assessments the less time teachers have for instruction. But why should we lose valuable instructional time at the expense of extracting formative data every so often? Cognitive tutoring allows integrating assessment into classroom instruction, which enables students to learn during a test (Anderson et al., 1995).

Personalized Learning Initiatives

Blended Learning. One of the emerging models that support personalized learning environments is blended learning. Simply put, this model blends individual and group instruction with facilitated technology through digital curricula. Blended learning encompasses a wide spectrum of tools and practice such as extensive use of technology in the classroom, use of online and formative assessment, adaptive software for students, learning management systems and many other technological advancements. A formal and commonly

accepted definition of blended learning is given below by research fellows at Clayton Christensen Institute for Disruptive Innovation:

Blended learning is a formal education program in which a student learns, at least in part, through online learning. Blended learning includes some element of student control over time, place, path, and/or pace and includes learning, at least in part, in a supervised brick-and-mortar location away from home. The modalities within a course or subject along each student's learning path are connected to provide an integrated learning experience. (Staker & Horn, 2012, p. 3)

Wolf and Schneiderman (2014) stated that the ultimate goal of blended learning is personalized learning for all students.

No Child Left Behind Act (2001) was ambitiously passed by the United States Congress with an intent to to close the achievement gap in reading and math. It required schools to make adequate yearly progress with the ultimate goal of all students meeting or exceeding proficiency by 2014. Unfortunately, many schools across the nation have fallen far behind this ambitious goal. Blended learning was suggested by experts as a powerful model to help close the achievement gap in reading and math. As NCLB unfolded over the years, we have started seeing early adopters of blended learning by innovative districts and schools in different pockets of the country. Although many schools implement different curricular initiatives in math classrooms, there is little empirical research to demonstrate their effectiveness (Slavin & Lake, 2007). My research topic is the impact of adaptive learning software on math achievement and how such digital learning tools can help create personalized learning environments for middle school students.

Adaptive Learning Technologies

In *the* area of adaptive learning technology research, there are two groundbreaking theories that form the foundation of adaptive learning technologies. The first is called the *Knowledge Space Theory* (KST), developed by Jean-Claude Falmagne, a Belgium-born

American mathematical psychologist, along with his colleague Jean-Paul Doignon. Falmagne and Doignon's research made significant contributions to educational technology, specifically in mathematics. The second predominant theory in adaptive learning technologies is Adaptive Control of Thought (ACT).

Knowledge Space Theory. Knowledge Space Theory (KST) is a key principle used in creating artificially intelligent adaptive learning software in math. KST is basically a knowledge representation and it is based on precedence relation. Due to strong prerequisite requirements among many mathematical subjects, Falmagne et al. (2004) argued that precedence relation can be used to design effective and rigorous assessment tools. This concept revolutionized the digital learning market for K-12 math education, which is now a billion dollar industry thanks to entrepreneurial spirit of elite-class American businessmen.

Adaptive online learning programs are now being largely used in order to leverage advances in artificial intelligence and cognitive science. Falmagne and his research team explained the science behind adaptive math software through knowledge space theory. This theory makes it possible to uncover the knowledge state of a particular student in a particular math topic through an online assessment (Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990). Knowledge state is defined as the complete set of problems in a particular topic that a student is able to solve (Falmagne & Doignon, 1999). Based on this basic principle, Falmagne and Doignon (1999) explains, an artificially intelligent adaptive assessment creates two shortlists of problems and concepts that guide students and teachers on what each student can do and what he/she is ready to learn. These two lists uncover the complete knowledge state of an individual student being assessed (Falmagne et al., 2007).

Effective teachers first assess the knowledge state of their students, find out their strengths and weaknesses, and then attempt to tailor their lessons to meet the varying needs of their individual students. This is how, in the most basic idea, teachers can differentiate instruction to personalize learning for each individual student. Knowledge space theory simulates the skills of an expert teacher to assess the knowledge state of a student in a computerized fashion (Falmagne & Doignon, 1985). As opposed to a theory of human cognition, knowledge space theory automatically informs an online assessment of student knowledge and keeps it accurate and continuously updated. A knowledge space for a particular math topic consists of all possible knowledge states related to that topic (Falmagne et al., 1990).

Adaptive learning software designed with knowledge space theory maps out the details of each student's knowledge and determines whether they mastered a particular topic with that continuous assessment cycle, and it knows what they are ready to learn next based on their updated knowledge state. Such programs use this knowledge to provide feedback to learners, make learning more efficient and provide continuous growth path by offering students a selection of only the topics they are ready to learn at that specific moment in time. This, in turn, provides a personalized learning environment for students, where each learner can progress at their own pace as if they are studying with a private tutor.

Knowledge Space Theory provides a framework to be able to formally describe knowledge domain structures in particular knowledge spaces. Albert and Hockemeyer (1997) argues that "the set of possible knowledge states is restricted by prerequisite relationships between the items" (p. 553). They went on to apply this phenomenon to describe hypertext structures, where prerequisite relationships among hypertext components are specified by

prerequisite links (Albert, Hockemeyer, and Held, 1997). The structures of knowledge spaces and these prerequisite links in hypertexts turn out to be quite similar, which made it possible to design intelligent hypertext tutoring systems for individualized instruction through combination of a hypertext model and knowledge space theory (Albert & Hockemeyer, 1997). The authors attribute the efficiency of such an adaptive hypertext-based ITS to two main things. First one is the effective procedures presented by knowledge space theory in identifying structures of a given knowledge domain. The second one is the description of a hypertext model that uses mathematical relations to connect prerequisite links to specific components of the knowledge domain within the student's knowledge state. Applying methods of relational database theory to knowledge domain structures "concerns not only individual access to document appropriate for student's actual knowledge but also the construction of sub hypertexts due to educational objectives or to the student's prior knowledge" (Albert & Hockemeyer, 1997, p. 555).

Combining knowledge space theory with relational hypertext model Hockemeyer, Held, and Albert (1998) created an ITS prototype called RATH, which stands for Relational Adaptive Tutoring Hypertext. Elementary probability theory was chosen as the initial course to be incorporated into RATH as a knowledge domain due to the fact that many students had difficulties with this course, hence further instructional support was needed (Hockemeyer et al., 1998). Any other math courses could be easily added into RATH, provided that a tutoring hypertext is available containing lessons, examples, assessments, and any other teaching materials for the course. Although the initial prototype of RATH was instrumental in providing instructional support to students, Hockemeyer et al. (1998) acknowledge several areas of improvement for RATH, such as the addition of an initial assessment to determine a

student's current knowledge state rather than starting everyone off with the assumption of zero knowledge in the course, and improving analysis of student answers to exercise questions in order to provide meaningful feedback rather than just giving a true/false response. The authors conclude that RATH was an important stepping stone for incorporating psychological theories into intelligent tutoring systems.

Villano and Bloom (1992) also used knowledge space theory as the foundational architecture of a probabilistic student model to be embedded into an ITS. The researchers were particularly interested in a probabilistic student model in order to be able to represent the uncertainty of estimating student knowledge. "Several factors contribute to uncertainty in student modeling such as careless errors and lucky guesses in the student's responses, changes in the student knowledge due to learning and forgetting, and patterns of student responses simply unanticipated by the designer of the student model" (Villano & Bloom, 1992, p. 1). In knowledge space theory, Falmagne and Doignon, (1985) define the basic unit of knowledge as an item. Villano and Bloom (1992) considered two basic steps in the construction of their probabilistic student model: (i) building the structural relationships among the items in domains of knowledge, and (ii) determining the initial probability values in the models. Following these steps required expert judgments and/or empirical student data. The authors explored both of these approaches with the addition of a novel application of neural networks for constructing knowledge structures. This research was able to leverage knowledge space theory and its utilization to develop adaptive, computerized student assessment and tutoring systems for building a probabilistic student model in KST. The probabilistic student model demonstrated many effective applications such as adaptive assessment item selection, adaptive assessment updating routine, knowledge type

representations, curriculum representation, hint level selection, advancement criterion, and student feedback. Villano and Bloom (1992) concluded that applying probabilistic student models to an intelligent tutoring system could lead to “developing a dynamic, non-deterministic student model capable of robust, individualized assessment” (p. 9).

Conlan, Hampson, O’Keeffe, and Heller (2006) presented a series of case studies where KST principles were applied to the analysis and determination of a learner’s knowledge for creating highly effective adaptive systems to support learner modeling and personalization in instruction. Over the course of six years, the researchers designed and refined an adaptive engine which powers the personalized learning services for adaptive instruction. The first case study examined a personalized course on mechanics, which tested the initial prototype of the adaptive engine in 2001. The adaptive nature of this personalized physics course was four-fold. It started with a *pre-test* to build a basic model of the learner’s initial knowledge. Next, the *dynamic personalization* mechanism presented a series of modules that the learner is capable of mastering. As the learner interacts with the eLearning modules and consumes the course materials, the *dynamic modeling* tool mapped out the learner’s evolving competencies aligned to created modules and objectives of the course. And the final piece of its adaptivity was *learner choice to expand the course*, which meant that additional content would be added to the eLearning space when learner decided he/she is ready to learn more, provided that all prerequisite content is already mastered.

The second case study presented by Conlan et al. (2006) was called Personalization in iClass, and it examined two particular personalization technologies. One is the personalization of content and activities, while the other is monitoring and profiling the learner. The authors report dramatic evolutions in both the use of adaptive engine and the

implementation of knowledge space theory during the iClass project. One innovative approach in this case study was to incorporate confidence degrees in knowledge assessment by soliciting the confidence degree of the respondent after each question item. Another addition to the initial assessment was to assess also skills along with concepts. The confidence degree attached to each response enabled the adaptive engine to create more robust personalization experience for learners as it hinted the ITS whether a wrong answer was merely a guess or a serious misconception. The iClass program offered a more powerful personalization experience with a two-fold approach: (i) the Selector service that adapts concepts and activities recommended to a learner based on their knowledge and personal preferences, and (ii) the LO generator that selects new learning objects from the content pool. Conlan et al. (2006) reports this separation of knowledge assessment from the personalization of eLearning as the primary advancement made through this case study. In the first case study presented above, the knowledge assessment approach was intricately tied to the adaptation mechanism. “This separation has enabled the evolution of the knowledge assessment to be carried out independently to the evolution of the personalized eLearning, thus enabling different pedagogical approaches to be adopted for different learners while still using the same knowledge assessment facilities” (Conlan et al., 2006, p. 1916).

Adaptive Control of Thought. The second theory supporting human-computer interaction as a teaching/learning tool is also widely used in cognitive learning systems, and is called adaptive control of thought (ACT), developed by Kenneth Koedinger, a cognitive psychologist at Carnegie Mellon University. Koedinger extensively studied human-computer interaction and his research significantly contributed to the development of intelligent tutoring systems. Cognitive tutoring software is a particular intelligent tutoring system that is

designed to provide personalized instruction and immediate feedback to students while assessing their knowledge and skills (Koedinger & Alevan, 2007). Cognitive tutors are designed based on a set of eight principles of the advanced computer tutoring theory, which are listed as follows:

1. Use production system models of the student
2. Communicate the goal structure of the problem space
3. Provide instruction on the problem-solving context
4. Promote an abstract understanding of the problem-solving knowledge
5. Minimize working memory load
6. Provide immediate feedback in errors
7. Adjust the grain size of instruction according to learning principles
8. Enable the student to approach the target skills by successive approximation

(Anderson, Boyle, Farrell, & Reiser, 1987)

Similar to knowledge space theory, adaptive control of thought also provides means to assess current knowledge and skills of a student, thereby providing personalized instructional methods via intelligent tutoring systems (Anderson et al., 1987; Koedinger & Alevan, 2007).

Although there is evidence that computer-aided instruction and intelligent tutoring systems increased student engagement and motivation in math classes (Schofield, 1995), it would be naïve to assume that such tech-enabled classrooms are exempt from motivation and engagement issues that are chronic in traditional lecture-based classrooms. Alevan and Koedinger (2001) stated that some students misuse intelligent tutoring systems designed with principles of adaptive control of thought. Baker, Corbett, Koedinger, and Wagner (2004) conducted a study on off-task student behaviors while working with a cognitive tutor. The

researchers found that students who misuse the cognitive tutor and cheat the system learned only two-thirds as much as their counterparts who used the system properly. The misuse of the tutoring system was referred to as “gaming the system” which meant systematically abusing the feedback and help features of the program for the sole purpose of finding correct answers and advancing through the material as quickly as possible (Baker et al., 2004). This study provoked the idea of designing enhanced ITSs which can adapt to not only varying levels of student cognition but also differences in student motivation. This requires a robust assessment of motivational state of students. Baker, Corbett, and Koedinger (2004) further investigated motivational issues when students interact with ITSs and ways to remedy such issues. During their study, Baker et al. (2004b) observed students gaming the system in two different ways, one by abusing the help feature and the other by engaging in systemic trial-and-error. The authors reported a strong negative correlation between a student’s frequency of gaming and their learning gains. Interestingly, they also reported no significant correlation between learning and other off-task behavior such as talking, sleeping, web-surfing, etc. This finding suggests that various types of low motivation do not impact learning outcomes the same way. When choosing a remediation approach to prevent or reduce gaming, it is critical to be able to detect which students are gaming the system and which ones are not, because the remediation approaches will likely frustrate learners who do not engage in gaming the system. Therefore, Baker et al. (2004b) “present and discuss a machine-learned Latent Response Model (LRM) that is highly successful at discerning which students frequently game the system in a way that is correlated with low learning” (p. 535). Creating an algorithm to accurately detect which students were gaming the system was the purpose of this study. In order to accomplish this, Baker et al. (2004b) included a sample of 70 students

using the same cognitive tutor lesson during their regular math instructional setting. Three different types of data were collected and combined as participating students interacted with the ITS. The first data set was recording a log of student activities for each student, the second data was observations of student behavior, while the third data came from learning outcomes of the participants. By following the patterns and frequency of errors in student responses and cross-validating this data with learning outcomes, the algorithm calculates probability rates for estimating whether a student is gaming or not and whether their learning outcomes were hurt due to gaming. Out of the 70 students participating in the study, the algorithm reported 53 students never engaged in gaming, 9 students gaming the system but not impacted negatively by their gaming behavior, and the remaining 8 students engaged in gaming with low learning outcomes. This machine-learned Latent Response Model (LRM) was proven to be able to detect gaming students who have low learning outcomes with high accuracy. Baker et al. (2004b) indicates that this research will have significant contributions to the further development of intelligent tutors that can adapt to behavioral characteristics of students as well as their knowledge and cognition levels.

Cognitive Science & Intelligent Tutoring Systems

Technological advancements across the world in the past few decades caused tremendous shifts in the way we live, work, and organize our lives. Industries and jobs have also gone through major transformations that changed the outset of the global economy. For example, automation and coding almost completely wiped out manual labor needs in manufacturing, networking, and infrastructures. Advancements in computer science and software design enabled major breakthroughs in the field of artificial intelligence, which has been quite instrumental in education. Online and digital learning mediums powered by

artificial intelligence and smart algorithms continue to create huge opportunities for cognitive science to expand learning methodologies and their efficiency in education. Online adaptive instructional programs have been designed to interact with learners by mimicking human responses and attitudes when providing performance feedback and learning recommendations to users, very much like the recommendation engine we see at Amazon or Netflix that customizes experience to user needs, interests, and personal preferences. Such educational software programs have the potential to function as an Intelligent Tutoring System (ITS) or as a cognitive tutor through machine learning.

Anderson et al. (1995) report significant achievement gains through use of cognitive tutors, where students in some cases reach the same proficiency levels as in traditional instructional setting in one-third of the time. The rich problem-solving environment combined with instructional guidance through step by step feedback and on-demand content hints are some of the features that make these theory-based tutoring systems highly interactive and appealing to learners (Koedinger & Alevan, 2007). As intelligent tutoring systems have matured and more widely used in K-12 education, more researchers have been attracted to the adaptive learning technology field. One intriguing question posed by Koedinger and Alevan (2007) was “How should learning environments balance information or assistance *giving* and *withholding* to achieve optimal student learning?” (p. 239).

How best to achieve this balance remains a fundamental open problem in instructional science. We call this problem the “assistance dilemma” and emphasize the need for further science to yield specific conditions and parameters that indicate when and to what extent to use information giving versus information withholding forms of interaction. (Koedinger & Alevan, 2007, p.239)

Intelligent tutoring systems provide a powerful tool for educators to leverage advances in artificial intelligence and cognitive science. When these scientific breakthroughs

are merged with the evolving power of the Internet, effective learning in math becomes scalable and personalized for every single student. Before intelligent tutoring systems, more traditional computer-based instruction was in use, which was not as robust as ITS in terms of tracking student performance and adjusting the teaching approach based on the needs and strengths of students (Woolf, 2009).

A number of researchers reported positive learning outcomes in math as a result of using computer-aided instruction, including ITS models (Murphy, Penuel, Means, Korbak, Whaley, & Allen, 2001; Beal, Arroyo, Cohen & Woolf, 2010). On the other hand, some researchers claimed that many computer-based instruction studies had design flaws, which made their findings invalid (Waxman, Lin, & Michko, 2003). According to Waxman et al. (2003), many studies were only descriptive and lacked relevant data and specificity, and only a handful created a randomized experimental design.

One of the evaluation studies examined use of a computer-aided algebra tutoring program by PLATO Learning in a high school remedial math setting. The overall program objective was to boost student performance on state mandated testing. In this comparative experimental study, a treatment group of 87 students spent 80% of the instructional time on PLATO's computer-based algebra program, 39 students in the control group received regular instruction without any use of technology. Hannafin and Foshay (2006) reported significant gains made by both treatment and control group on the state exams. While the mean score for the control group was a lot higher than the mean score for treatment group, the achievement gains of the treatment group were significantly higher than the gains made by the control group (Hannafin & Foshay, 2006).

In a similar study, Koedinger, Anderson, Hadley, and Mark (1997) evaluated an algebra curriculum called PUMP, which also had a supplemental intelligent tutoring system called PAT. In this rather larger experimental study, the treatment group included 470 students enrolled in algebra classes who were taught with PUMP curriculum and used the ITS as well. The control group had 120 students also enrolled in algebra class who did not use the ITS at all and were taught with a traditional curriculum. Students were given two standardized tests and two custom-created tests by the end of the class, and the students in the treatment group outperformed their counterparts in control group on all tests (Koedinger et al., 1997).

A rather contemporary study attempted to evaluate and compare the performance of students from two Algebra I classes; one that uses a computer-based instructional software and another that uses an intelligent tutoring system. Campuzano, Dynarski, Agodini and Rall (2009) set up a control group where students accessed Larson computer-based algebra program as part of their traditional curriculum, while the treatment group used Cognitive Tutor as the core algebra curriculum. Control group students accessed the program 313 minutes per year over 6-week periods. Treatment group students used the Cognitive Tutor software an average of 2,149 minutes per year over a 24-week period. This large scale study reported no significant difference between the performances of treatment and control groups using computer-based and intelligent tutoring softwares for algebra instruction (Campuzano et al., 2009).

Barrus, Sabo, Joseph, Atkinson, & Perez (2012) conducted their research by means of a summative evaluation of two off-the-shelf intelligent tutoring systems: Carnegie Learning's Cognitive Tutor and ALEKS. The goal of this study is to measure the effectiveness of these

two ITS models when they are used as an exclusive method of instructional delivery. 30 remedial high school algebra students were selected and they were randomly assigned to either ALEKS or Cognitive Tutor and worked on their respective adaptive software every day during the 14-day summer school. Students' gains in math were measured through the Accuplacer algebra and arithmetic reasoning subtests and both groups made significant gains from Day 1 to Day 13 using either ITS model (Barrus et al., 2012).

In another recent study, Nwaogu (2012) conducted research on ALEKS by means of a quasi-experimental study using one-group, non-randomized, pretest-posttest design to measure the effect of ALEKS on student math achievement. 80 students participated in this study as part of an online 5-week long summer College Mathematics course. Nwaogu (2012) reports a strong correlation between the concept mastery and achievement scores on the quizzes and posttest. There was not a significant relationship between time spent learning in ALEKS and achievement scores. The results also did not suggest any correlation between the two independent variables: concept mastery and time spent learning in ALEKS (Nwaogu, 2012).

Foster, et al (2016) studied the effects of using Building Blocks Software, specifically with kindergarten – level low income and minority children in an urban school district. Student – subjects for this study were 243 monolingual and predominantly minority (63% black, 30% Hispanic, 4% mixed/other and 2% White) who attended a Title I school. Students' pre – and post – test scores (beginning and end of kindergarten year) on the "Research Based Early Math Assessment," were recorded as a measure of numeracy. Subjects were randomly selected to receive Building Blocks Software (mathematics) or "Earobics Step 1", a literacy oriented computer - assisted tutoring program. The authors

found a significant positive affect of using Building Blocks Software to supplement instruction ($F(1, 178) = 8.08, p < 0.01$), controlling for subjects' initial numeracy score.

Computer – assisted tutoring has been shown to be effective with at – risk students, as well. Salerno (1995) examined the effect of computer – assisted tutoring on students defined as ”at – risk“ (i.e. dropping out of or failing in school for reasons of SES, drug use, lowered English skills, or previous failure in academic subjects). Salerno’s study randomly selected 150 students deemed at – risk from a large urban school district. Of the selected students, 50 were assigned to each of two experimental groups and the remaining 50 were assigned to a control group. The first experimental group received computer assisted instruction (CAI) in mathematics for the usual amount of time used in the school district. The second experimental group received CAI for an 60 minutes each week beyond the time used for the first experimental group. The control group spent an equivalent amount of time performing math – related tasks without CAI. The experimental and control groups were further divided by student – sex. Alternate forms of a criterion – referenced test were used as a pretest/posttest measure.

Salerno (1995) found in subsequent analysis that additional time spent with CAI had a positive affect for both boys and girls; but, the effect was statistically significant only for boys. Both boys and girls in the treatment groups showed a statistically significant positive affect relative to the control group. Salerno demonstrated a positive effect of CAI; but, the author also identified a differential effect by student sex.

Ravenel, Lambeth & Spires (2014) studied the effects of CAI on student attitude towards mathematics and the amount of engagement among students. The researchers divided 31 fourth grade subjects into two groups, hands – on and CAI. The students were

approximately representative of their district with respect to SES and race (66% White, 33% non – white; 70% economically disadvantaged). Teachers monitored engagement using a checklist while students were receiving instruction, either in the hands-on activity mode or through CAI. Student attitudes toward mathematics was measured by a survey administered to the study subjects. The study was conducted across a seven week period. The findings of the study showed that both CAI and hands – on groups had a positive attitude towards mathematics. The authors also found that the CAI group had a statistically significantly higher level of engagement with their mathematics instruction than did the hands – on group.

Adaptive learning programs are designed to enhance cognitive retention of facts, concepts, and principles (“MIND Research Institute”, 2017). That is why many blended learning solutions (digital curricula, adaptive learning software, dynamic assessment systems, etc.) incorporate game-based simulations into their instructional design. Before advancements in hardware and software technologies, education practitioners and researchers tested out simulation-gaming techniques to improve cognitive achievement and retention (Anderson, 1970; Baker, 1968; Lucas, Postma, & Jay, 1974; McKenney & Dill, 1966). Lucas et al. (1974) conducted an experimental study involving 294 high school students enrolled in United States History course in five different public high schools in a Midwestern state over a five-week instructional period. The subjects were chosen from a wide range of geographical and social regions to ensure diversity in sampling. The purpose of the study was to compare cognitive achievement and cognitive retention of participants in traditional lecture-based instruction versus game-based simulation techniques. In order to control for teacher variable, each history teacher from participating schools had one control class and one experimental class. The researchers found that both control and treatment groups made

equivalent gains in terms of cognitive achievement based on the post-test given at the end of the five-week treatment period. However, based on the post-test given to both groups after a delay of ten weeks revealed that simulation-gaming group performed significantly better than traditional lecture group. With these results, Lucas et al. (1974) concluded that simulation-gaming techniques are effective in cognitive retention of facts, concepts, and principles.

Algebra is one of the most critical subjects students take in high school. Every single state education agency across the nation selects Algebra as one of the handful of high school courses to administer a summative end-of-course test for accountability purposes. Some states even make it a policy to pass Algebra state test before high school students can apply for a driver's license. Student performance in Algebra is often one of the key indicators of success in college, and therefore is used as a metric for college readiness by K-12 school districts. For this reason, it is common practice among districts to find alternative interventions for Algebra and pilot them to evaluate the results before scaling it across an entire school or district. Below report presents findings from an action research of a pilot of ALEKS in high school algebra and how intervention correlates with student outcomes in state Algebra assessment.

Lavergne (2007) attempted to answer the question "How does the ALEKS online math tutorial impact the learning and retention of math concepts and skills for Algebra 1 students as measured by the MAP Test (Measures of Academic Progress)?" This action research study was implemented in a mid-sized community high school in the Midwest. The algebra I progress of 98 students was followed for one semester. Participants were pretested at the start of the semester and tested again at the end of the semester using the MAP test. During the semester, participants used ALEKS, a computer – administered mathematics

teaching program. Participants' achievement results across the semester were subsequently compared with the school district as a whole and with national means for ALEKS users.

The author found that ALEKS users had a 2.7 RIT score increase during the semester, compared with a one point increase for non-ALEKS users in the district and a 1.6 point increase nationally. Although no hypothesis test was associated with this report, ALEKS users in the study group scored much higher than the national average and more than 2.5 times the increase of non-ALEKS users within the same school district.

As previously mentioned, ALEKS was originally created to address remediation problems in college mathematics courses. Naturally, it was adopted by colleges and universities first, and eventually the program was enhanced to be scaled in K-12 schools to address learning challenges in mathematics at earlier grades. Due to its widespread use at higher education institutes especially at its inception, the amount of research investigating ALEKS and its effects on math at universities is a lot more than similar research conducted in K-12 schools. Several examples of such studies at various universities are reported next.

Taylor (2008) investigated the utility of using ALEKS, which emphasizes mastery learning, to remediate the mathematics skills of college freshmen enrolled in an intermediate college algebra course. Decreasing mathematics anxiety and improving students' attitude towards mathematics were also a focus of the study. Taylor focused on the following five research questions (p 38):

1. Does a mastery learning perspective of remediation, where students are expected to learn all the objectives in an intermediate algebra class, make a difference in mathematics achievement?
2. What differences exist between students using Assessments and Learning in Knowledge Spaces (ALEKS) compared to students who are taught Intermediate Algebra using a traditional lecture style?

3. Are there differential mathematics effects for either group based on demographic factors such as gender, age, ethnicity, number of mathematics courses taken in the past, and degree plans?
4. Do differences emerge between the two groups of students in their perceived level of mathematics anxiety?
5. Are the students' attitudes toward mathematics a factor in students' inability to be successful in Intermediate Algebra?

Participants in the study were 54 students assigned to ALEKS-based instruction and 39 control students who receive traditional lecture-based instruction. Ethnically, students participating in the study had a similar distribution to the state and university in which the study was conducted, with moderate deviation in the percentage of Hispanic students. The National Achievement Test First Year Algebra Test (NATFYAT) was used as a pretest and posttest measure of achievement. The study was conducted in a four month period between September and December. Paired – sample t-tests and correlational analysis showed that both groups had improved achievement across the study, but the control group had a greater gain and achievement than the ALEKS group. The authors hypothesized that this difference may be attributable to the fact that ALEKS is differentially effective based on individual student characteristics. The authors found no differences between groups based on ethnicity, gender, or age.

Further analysis showed that students in both the experimental and control groups had lowered anxiety levels across the period of the study; but, the experimental ALEKS group showed a statistically significantly greater reduction in anxiety than did the control group. The F-S scales, a Likert-type questionnaire, was used to determine changes in student attitude during the study. Results showed that there was no statistically significant change in student attitude among the experimental students, but that there was a net negative change in student attitudes among the control group (Taylor, 2008).

Although the study was limited both in size and complexity, the results indicate that ALEKS can potentially increase student achievement in algebra while decreasing anxiety. Students using ALEKS maintain their attitude during the course while students in the lecture – type course exhibited a negative change in attitude (Taylor, 2008).

Another study involving ALEKS at a university level examined a curriculum redesign to leverage blended learning models by combining theory and web-based learning modules. Hagerty, Smith, and Goodwin (2010) presented a case – study of redesigning the algebra curriculum at a four-year university, including the addition of ALEKS, a web – based learning tool for mathematics. At the outset of the curriculum redesign, the University was faced with declining enrollment in upper-level mathematics courses, such as trigonometry and calculus, because students were not well – prepared in their algebra coursework to engage in the upper-level mathematics courses. The mathematics faculty, in conjunction with psychology and sociology faculty, explore the attributes of mathematics students in order to determine the most effective method or redesigning the curriculum. One finding was that students had poor self – efficacy when beginning their studies at the college algebra level (Hagerty et al., 2010). To counteract this, they determined that mastery experiences should be emphasized in students’ mathematics study. ALEKS was evaluated for both content and underlying principles, and found to be adequate for algebra instruction. The incorporation of ALEKS, along with other substantive changes in the curriculum, was credited with several beneficial effects on the algebra and mathematics achievement of students (Hagerty et al., 2010).

- Algebra course passing rates (“C” or above) increased by 21% (54% to 75%) within a four-year period. The authors attribute this to improvements made by

ALEKS publishers, improved understanding by the faculty of ALEKS, and improved incorporation of ALEKS by the faculty.

- The Collegiate Assessment of Academic Proficiency (CAAP), a nationally – normed algebra assessment was used to measure the difference between students taking the traditional curriculum versus those taking the redesigned curriculum which incorporated ALEKS. Students taking the redesigned curriculum scored moderately higher on the ACT mathematics portion than did students taking the traditional curriculum. The authors did not statistically test this measure as a hypothesis.
- Class attendance improved by 26% across the four-year redesign period of the course.
- Improved self-efficacy was another benefit of the algebra redesign project. This was determined by the fact that three times as many students enrolled in trigonometry immediately following the college algebra course.

The authors conclude that ALEKS, although it was not the only part of the course redesign, was a significant contributor to the success of the course redesign.

ALEKS is also used to enhance learning and instruction in statistical courses. Another study at a university used mixed methods to measure the impact of ALEKS on students' learning and attitude towards statistics. Xu, Meyer, and Morgan (2009) studied the effectiveness of Assessment and LEarning in Knowledge Spaces (ALEKS) in addressing the needs of students based on their individual characteristics. The two primary research questions of the study were (p 5):

- Does the integration of ALEKS improve student performance in Stats I?

- How does the hybrid class with an online commercial tutoring system impact students' learning and attitude about statistics?

The authors collected data from the fall semester of one year (traditional instruction, n=45), which was considered the control group and the fall semester of the following year (traditional instruction + ALEKS, n=41) which was considered the experimental group. Additionally, qualitative data were collected via a survey and three focus group interviews from the experimental (ALEKS) group. ANCOVA, using incoming GRE quantitative scores as a covariate to control for initial statistics levels, showed no statistically significant difference between the control and experimental groups (ALEKS/non-ALEKS) in achievement gains during the course. Neither age nor race were statistically significant in the ANCOVA results (Xu et al., 2009).

Subsequent questionnaires and focus groups with students in the experimental (ALEKS) group found two primary concerns with using ALEKS to augment instruction: 1) They found ALEKS to be time-consuming; and 2) the ALEKS coursework was not well matched with the associated classroom instruction (Xu et al., 2009). A third area, concerning ALEKS assessments, as Xu et al., (2009) reports, showed higher satisfaction among students that were high performing at the start of the course than with other students. The authors conclude that the relationship between student attributes, such as initial skills, knowledge and attitudes, may be significantly related to the students' perceptions of ALEKS and the efficacy of ALEKS in improving their learning. The authors also hypothesize that these student characteristics may be an important determinant of whether or not ALEKS should be used with selected students.

Many college freshmen end up dropping out of college after having academic struggles, mainly reported in mathematics courses (Shakerdge, 2016). ALEKS is a

commonly used intervention to address this problem in many universities (“ALEKS”, 2017). A state university located in northwestern U.S. particularly had low freshman retention rate of 64% among its engineering students, compared to national average of 69% (Pyke, Gardner, Hampikian, Belcheir, & Schrader, 2007). This statistic prompted further investigation to find ways for improving freshman retention rates among engineering students. The university turned to ALEKS as an instructional support system for freshman engineering students.

Hampikian, Guarino, Chyung, Gardner, Moll, Pyke, and Schrader (2007) conducted a study on ALEKS in a Precalculus class for freshman engineering students. In this study, 84 freshman engineering students were enrolled in Precalculus classes. The university added a non-compulsory freshman engineering course (ENGR 110) in an effort to increase retention rates. The main focus of ENGR 110 was supporting the Precalculus students via ALEKS. Of the 84 freshman engineering students enrolled in Precalculus, 37 opted in to enroll in ENGR 110, which had a requirement of making a weekly progress of 4%-6% on ALEKS, as well as a goal of completing 65%-75% of their knowledge space by the end of the semester. These students spent most of their time working on ALEKS. Engineering students who did not choose to take ENGR 110 still remained in the Precalculus course without using ALEKS.

Hampikian et al. (2007) created an assessment rubric and used it on all engineering Precalculus students comparing ALEKS and non-ALEKS users on both interim performance over the course of the semester and also on their final grades. They found that mean math scores in Precalculus for ALEKS users were higher than non-ALEKS users across the board, however the results were not statistically significant. The assessment rubric also had an exit survey component for ALEKS users only, which provided qualitative data on student

experiences and attitudes in Precalculus class. The survey data revealed positive regard for ALEKS, as 63% of students reported that ALEKS helped them succeed in Precalculus as well as increase their confidence in math (Hampikian et al., 2007). The authors attributed the statistically insignificant gains made by ALEKS users to small sample size.

In another study involving ALEKS use in college algebra courses, Padilla-Oviedo et al. (2016) examined how instructional strategies, college division, and gender may have an impact on student performance and learning outcomes as measured by final course grades and drop/withdrawal data. The motivation for this study was due to high drop/failure rates in college algebra courses and their direct influence on 4- to 6-year graduation rates. The fact that college algebra courses have become gatekeepers towards graduation makes it all the more worthwhile to invest in evidence-based instructional practices and research on their effectiveness. Padilla-Oviedo et al. (2016) studied 253 students enrolled in college algebra across seven different college divisions in a southern university serving predominantly Hispanic students (90%). The participants were receiving instruction via three different instructional strategies including a traditional lecture-based instruction, technology-based instruction with adaptive learning program ALEKS, and a targeted intervention program called *College Completion America Fundamentals of Conceptual Understanding & Success* (CCA-FOCUS). CCA-FOCUS started out as a summer bridge program and expanded into a semester-based implementation where developmental math students are placed in college algebra while receiving targeted support and just-in-time content-specific remediation (Loredo, 2012). Padilla-Oviedo et al. (2016) found that final grades of study participants in college algebra course was significantly different across the three different instructional strategy groups. Both ALEKS group and CCA-FOCUS group performed significantly better

in college algebra than traditional lecture group, while the CCA-FOCUS group had the highest mean score. The authors also looked at gender and college division to see if those independent variables would impact student outcomes in college algebra. Padilla-Oviedo et al. (2016) reported no significant differences in performance between male and female students as well as across college divisions. The intervention approach in CCA-FOCUS group was similar to what ALEKS offered in terms of personalizing the learning experience of students, but it also included a cooperative learning element where students also leveraged peer-tutoring through discussion of material with their peers. The authors attributed the higher scores achieved by the CCA-FOCUS group to the cooperative learning strategy.

Achievement disparities among racial groups has been a hot topic in education in the accountability and testing era of the standards movement. It is also a common issue in regards to high school graduation and college persistence. Researchers often reported that performance of African American and Hispanic students often lag behind their White counterparts in both SAT/ACT scores and course grades in mathematics (Harris & Herrington, 2006; Orr, 2003). Hu, Luellen, Okwumabua, Xu, and Mo (2008) conducted a study at a large urban university to examine the effectiveness of ALEKS as an intelligent tutoring system on closing the racial performance gaps in an undergraduate behavioral statistics course. The researchers stated that ALEKS was adopted as an ITS to provide online instructional support for students who take the behavioral statistics as an online course via distance learning. The adaptive, mastery-based, and self-paced features of ALEKS tailoring the instructional material to the needs of individual students were most appealing to the university in their decision to select ALEKS as an ITS for the online version of the course.

Hu et al. (2008) included 548 undergraduate students in their study from a 10-year span of academic years from spring 1995 to fall 2005 who took both the online and on-site version of the behavioral statistics course from the same professor. The study followed a non-equivalent control group design to compare student performance in ALEKS-using sections of the course to a retrospective comparison group following a traditional lecture style instruction. 137 students were in the control group and 411 students were in the comparison group. After “a full factorial analysis of variance (ANOVA) examining the relationships between cumulative grade point average (GPA) prior to enrolling in behavioral statistics and three factors, passing status (passed vs. failed), race (black vs. white) and course format (lecture vs. online ITS)” Hu et al. (2008) reported no significant two-way or three-way interactions (p. 6). The only significance was observed in the main effects for passing status and race. Since the two-way interaction between race and course format was not significant, Hu et al. (2008) concluded that the trends for failing students were similar in online and on-site course formats. However, the relationship between the standardized course grade and race varied significantly by course format. The average standardized grade in the lecture-based course format was significantly lower for African American students compared to their white counterparts. This is a typical observation of performance disparity among racial groups. Interestingly, the same racial disparity between African American and white students was not observed in the online ITS class. The average standardized grade for African American students in the ALEKS-using online class was equivalent to that of their white counterparts, which is an indicator of elimination of racial disparities (Hu et al., 2008). Moreover, the average standardized scores of African American students enrolled in online ITS class were significantly higher than that of their African American counterparts in

lecture-based classes. White students, on the other hand, did not perform significantly different in online versus traditional lecture format. These findings show promising potential of online intelligent tutoring systems such as ALEKS in closing the achievement gaps between racial groups (Hu et al., 2008).

Nexus between K-12 Personalized Learning and Post-Secondary Education

The issue of knowledge and skills gap of students in the nation's public schools has been the root cause of high school graduates being persistently underprepared for college and career readiness. Innovative districts made bold attempts to design personalized learning initiatives to achieve specific post-secondary outcomes. One approach has been designing K-16 pathways, which offers post-secondary exploratory experiences for K-12 students. One of the key considerations in turning such K-16 initiatives to success is how to provide credentialing to students for completing post-secondary experiences while in high school. To address these design challenges, districts such as Race to the Top grantees partnered with higher education institutes to establish connections and identify relationships between profiles of what secondary and post-secondary graduates should know and be able to do better. Identifying the best measures of competency including and beyond academic skills is critical to fulfill the promise of competency based education, which requires assessing readiness based on mastery of learning targets rather than seat-time requirements. Adaptive learning programs such as ALEKS make it possible for schools to move towards a competency based education model.

Badaracco and Martinez (2011) proposed a new architecture for an intelligent tutoring system, which incorporated competency-based education as its pedagogical model. General architecture of an ITS consists of four components: 1) domain model, 2) student

model, 3) pedagogical model, and 4) interface model (Sleeman & Brown, 1982; Polson & Richardson, 1988). Domain model is the knowledge domain that is being taught. Student model refers to the characteristics of the learner, mainly the initial knowledge state. In other words, student model is the student's knowledge of the domain model. Pedagogical model, also known as the instructional model, is the set of strategies how the material is delivered and how the learner is allowed to progress throughout the material. The interface component designs and supports how the learner interacts with the ITS. These components work together to deliver a customized learning experience for the user.

Competency-based education is an emerging instructional model that is gaining momentum across the globe. Many innovative school districts in the U.S. moved towards a competency-based education model by embedding digital playlists of modular content into their curricula, aligning formative and summative assessments to allow mastery based progression and fluid movement of students as they work through their customized playlists (Vander Ark, 2013; 2014; Horn & Staker, 2014). Competency education movement created a framework for designing personalized, competency-based education in 2011 when 100 innovators came together to establish a consensus on what constitutes a high-quality competency education model. Sturgis (2017) summarizes their working definition as follows:

- Students advance upon mastery.
- Competencies include explicit, measurable, transferable learning objectives that empower students.
- Assessment is meaningful and a positive learning experience for students.
- Students receive timely, differentiated support based on their individual learning needs.

- Learning outcomes emphasize competencies that include application and creation of knowledge, along with the development of important skills and dispositions.

Competency-based education movement also spread to Europe and Latin America with *Tuning Educational Structures in Europe* and *Tuning Latin America* projects. More than 135 universities in Europe joined the movement since 2001 to fine-tune the educational structures and improve cooperation among higher education institutes for the purpose of developing excellence and effectiveness (Tuning Project, 2017). Badaracco and Martinez (2011) explains tuning “as a platform for developing reference points at subject area level. These are relevant for making programmes of studies (bachelor, master, etc.) comparable, compatible, and transparent” (p. 127). The ability to perform effectively in a given situation is a common definition of competence. While knowledge is an essential requirement of a competency, it is also a representation of know-how that integrates conceptual, procedural, and attitudinal knowledge (Badaracco & Martinez 2011). Effective competency-based models identify competencies that correspond to learning targets and clearly outline performance descriptors for evaluating progress within each competency (Sturgis, 2017). Competency-based education models challenge the current time-based requirements of American public schools and demand new policies to move away from seat-time towards personalized, competency-based systems. Following are the main reasons listed for this radical change effort:

- To ensure that all students succeed in building college and career readiness, consistent with the Common Core of world class knowledge and skills;
- To build the capacity of districts, schools and educators to respond more rapidly to the needs of students and engage in continuous improvement;

- To take advantage of the extraordinary technological advances in online learning for personalization, allowing students to learn at their own pace, any time and everywhere;
- To provide greater flexibility for students that would otherwise not graduate from high school because they have to work or care for their families. (Sturgis, 2017)

Although there is a strong, coordinated effort to scale competency-based education across the U.S., current federal, state, and district policies and accountability structures are limiting the abilities of local education agencies to make the shift to competency education. Hence, the movement is still at its infancy.

In the absence of a standards-based competency approach to credentialing students before they are allowed to graduate from secondary schools, post-secondary institutions are presented with the challenge of effectively assessing student readiness, which is a critical task that helps universities place students into courses they are most likely to succeed. A study conducted at the University of Illinois (Harper & Reddy, 2013) compares ALEKS as a placement tool into the university's five entry-level freshman math courses to previously used ACT math scores for placement. As previously mentioned, ALEKS was built upon the principles of Knowledge Space Theory (KST). The initial adaptive assessment in ALEKS uncovers the current knowledge state of a student based on the correct answers of the student. Then ALEKS creates a customized learning environment for that student once it determines a set of items the student already demonstrated mastery over through the mapping of his/her knowledge state. Carpenter and Hanna (2006) presented empirical evidence that ALEKS can be used as a placement measure, which informed the University of Illinois' decision to switch to ALEKS for placement program, Harper and Reddy (2013) reports, due to its ability

to accurately measure knowledge, given that initial knowledge at the start of a course is a strong indicator of student performance in that course (Carpenter & Hanna, 2006; Doignon, 1994).

Harper and Reddy (2013) explain the critical need for accurate placement since students come to university from different backgrounds, locations, and school experiences with varying math skills and backgrounds. The curriculum, assessments, and exit criteria for the same course at different secondary schools could be vastly different, not to mention varying grading practices of these institutions and their educators. Consequently, the math coursework students complete during their secondary education does not constitute readiness for the next college level math course, and certainly not presenting reliable data for success indication in postsecondary math courses. Therefore, there is a dire need to precisely assess readiness before placing students.

Many universities use standardized test data for placement purposes such as ACT, SAT, and AP Calculus exam scores (Baron & Norman, 1992; Carpenter & Hanna, 2006; Harper & Reddy, 2013). This was also the case for the University of Illinois prior to switching to ALEKS as a placement mechanism. It is debatable how well standardized assessments can measure specific skills and/or predict student success later in the course. According to Baron and Norman (1992), there is very weak (in some cases even negative) correlations between SAT scores and student performance in a math course. Harper and Reddy (2013) also argue that use of standardized test scores in placement decisions correlate poorly with student performance due to obvious reasons. One such reason is, many students take their ACTs or SATs in junior year and/or in early senior year, and they do not take any math courses in their senior year since it is not a requirement in many states. One

ramification of this common practice is that data from students' latest math course grade or exit exam as well as ACT/SAT type standardized tests are not reliable measures because students have completed these assessments long time ago prior to starting a college math course, and their attained knowledge in most recent math course have already been lost after new knowledge acquisition. To overcome these challenges, the university decided to require students to take a placement assessment in the last 4 months prior to starting a math course and achieve a cut score for readiness (Harper & Reddy, 2013). When students fall short of the cut score for readiness, they are either removed from the course to be placed into another course, or ALEKS would give students the option to remediate until they become ready using the learning mechanisms built into ALEKS.

Although many colleges use ALEKS as a placement program, the initial assessment of ALEKS was not designed as merely a placement exam ("ALEKS", 2017). It is a part of a learning system where the initial knowledge state of a student determines a customized design of that student's learning experience in that course. The reason behind ALEKS being used as a placement mechanism is that the accurate mapping of a student's initial knowledge state is a strong predictor of eventual student performance (Harper & Reddy, 2013). The adaptive nature of ALEKS adjusts the level and difficulty of each question based on the students' answers to allow exact pinpointing of a student's knowledge state, thereby providing a customized learning path for each student. Falmagne and Doignon (2011) explains that due to the fact that ALEKS is a completely free-response assessment and ITS, it minimizes the risk of inaccurate inferencing when constructing learning spaces, as it takes out the possibility of careless errors and lucky guesses we often see with multiple choice assessments.

The goal of effective and accurate placement is to reduce course failure and withdrawal rates. Evaluating placement mechanisms depends not only on student performance in the course, but also on course drop/withdrawal rates. Harper and Reddy (2013) used three years of placement program data to compare ACT math score and ALEKS assessment. They found that ALEKS assessments are strongly correlated to student performance in all five of the university's entry level math courses. Furthermore, they concluded that the ability of ALEKS to create the initial knowledge state of students was a strong foundation for placement decisions. Compared to ACT math scores, ALEKS assessment provided a significant improvement on the correlation of placement data and course grades in all five courses. The authors attributed this strong correlation to the proximity of the assessment to enrollment in the course.

The use of knowledge state in providing a personalized learning environment customized to the needs of each student is a powerful idea; one that is not possible with use of static item assessments. Adaptive assessments, however, are able to map out a learner's knowledge state with high precision. Two students may make the same score on ALEKS assessment but end up having two completely different knowledge states. Likewise, two students may score equally on ACT by answering almost completely different sets of questions on the exam. In static assessments like ACT though, we are not able to determine the knowledge state of each student and end up placing them into the same course just because they scored the same. In ALEKS, two students may have a completely different learning experience with the program even within the same course despite scoring equally on the initial exam. In other words, the learner experience is customized and personalized with

powerful and smart use of adaptive technologies and principles of knowledge space theory,
all thanks to the advancements in the field of cognitive science.

CHAPTER 3

METHODOLOGY

This quantitative study investigated the effects of ALEKS, a web-based artificially intelligent assessment and learning system, on math achievement of students in fifth through ninth grades in an urban school setting. This chapter describes the methodology for the research study, and includes an overview of the research sites and participants, instruments used to collect data, procedures used to carry out the design and how the data was analyzed.

Participants

Participants in this study come from two different public charter school districts. Both school districts are located in the same city, a southwestern urban metropolitan area. Being located in the same city and serving families from similar underserved neighborhoods and communities, both districts have very similar demographics and characteristics as follows:

- 60-65% economically disadvantaged students
- Predominantly Hispanic/Latino students: 60-65%
- Heavy curricular focus on literacy and numeracy
- Heavy extra-curricular focus on STEM (Science, Technology, Engineering, and Math) events and activities
- Young/inexperienced teaching workforce: Average years of teaching experience ~1.8-2.2

Both districts are managed by the same Charter Management Organization (CMO) and therefore implement the same curriculum, and teachers receive the same professional development services. Despite being separate independent charter districts, it is safe to say that both districts implement the same educational model. Due to a grant available to only

one of these districts, treatment and control populations were naturally formed within these schools.

Data Considerations

The accessible population consists of all students enrolled in grades 5 through 9 in five different schools within the two selected districts. All students enrolled in these targeted grade levels were eligible to participate in the research. Therefore, the accessible population in this sampling frame includes all participants. Hence, population validity was established. The sample size was 1110 students. Missing values treatment and data matching efforts were applied to the collected data to finalize the sample. The sample was divided into two groups forming our experiment group and control group. A propensity score matching technique was used to eliminate selection bias based on unobserved characteristics between control and treatment groups. This precise matching on student demographic characteristics alleviates the need to control for these characteristics during analysis.

Design. The quantitative study has a pretest-posttest quasi-experimental design with control and experiment groups. Because there was no random assignment of participants into control/experiment groups, this is not considered a true experimental design, hence it is deemed quasi-experimental.

Data Collection. The data was collected from the 2014-2015 school year. All participants had already completed fall and spring administrations of NWEA MAP assessments. This data is stored in NWEA servers, and is accessible to school/district leaders at any time. Participants in the experimental group used the ALEKS program throughout the school year. These students spent roughly 45 minutes a day working with ALEKS to make progress on their personal learning plans. The complete data set (concept mastery, time spent,

and latest test performance) for each student is archived on ALEKS servers. This data set is accessible to school/district leaders any time on-demand.

Measures

In this study the researcher has collected demographic information such as gender, race, socio-economic status, grade level, and inclusion in any special programs (ESL, Special Education, and Gifted & Talented) to be able to make correlations and generalize findings. In addition to the demographic data, student achievement data were collected using NWEA MAP math scores and diagnostic assessments in ALEKS as measures of math achievement.

MAP is a norm-referenced adaptive test developed by Northwest Evaluation Association, a non-profit educational organization providing assessment and growth tools to many schools nation-wide. ALEKS is an adaptive educational mathematics software, which is only implemented in one of the participating districts. The ALEKS program includes its own diagnostic tests, which are used as a pretest and posttest to measure concept mastery in math. Both ALEKS diagnostic assessments and NWEA MAP already have reliability and validity evidences provided by test developers. Summary of reliability and validity evidence is presented as follows:

Reliability is an essential measure of consistency of an assessment. There are two types of consistency checks we are looking at; performance across time (test-retest reliability) and performance across forms (parallel forms reliability). Essentially, we are looking to answer these two questions:

- To what extent does the test administered to the same students twice yield the same results from one administration to the next?
- To what extent do two equivalent forms of the test yield the same results?

Both of these questions are answered using the Pearson r correlation coefficient. NWEA conducted several reliability estimate studies. In 2002, reliability estimates for test-retest reliability and parallel forms reliability of NWEA MAP mathematics test from fall to spring yielded Pearson correlation coefficients of .91, .93, .94, .93, and .90 for grades 5 through 9 respectively. Given that minimum expected coefficients (r) for test-retest reliability and parallel forms reliability are .80 and .85 respectively, the results propose strong reliability of the NWEA MAP assessment. NWEA also estimated reliability across test items, which is often referred to as internal consistency. For internal consistency, NWEA developed its own test called the marginal reliability coefficient. In 1999 fall and spring administrations of NWEA MAP mathematics test yielded coefficients ranging from .94 to .96 in grades 5 through 9. These results are nearly identical to the Cronbach's Alpha method for calculating internal consistency.

Concurrent validity is an important measure to determine if an assessment measures what it is supposed to measure. Concurrent validity is important to be able to make inferences based on the results of a test or to make generalizations about a population. NWEA assured content validity by mapping its content standards with a dozen different state assessments. The essential question for validity is, "How well do the scores from NWEA MAP correspond to the scores from another test?" A Pearson r correlation coefficient can be used again to measure this concurrent validity. When this coefficient is in the mid-.80's we can say that content validity is established. The concurrent validity measure for Texas yielded a Pearson correlation coefficient of .82 for 7th grade mathematics in 2003.

ALEKS is a much more complicated assessment system built on artificial intelligence and does not have static tests. That makes it very hard to measure its reliability and validity

in traditional ways. For this reason, a comprehensive study focusing on the validity and reliability of ALEKS has been conducted by the developers and founders of knowledge space theory. This study investigated the extent of how predictive of an ALEKS user's responses to the problems that are not in the assessment, thereby measuring the validity and reliability of the artificial intelligence algorithm that runs the program (Falmagne, Cosyn, Doble, Thiery, & Uzun, 2007). The study reports a mean correlation of .67 between predicted and observed responses and the mean log odds ratio of 2.75 (Falmagne et al., 2007). The authors also point out to an indirect evidence of the validity of ALEKS as follows:

We also presented an analysis of the learning efficiency achieved by students because it is directly related to the validity of the assessment. Indeed, the problem types that are proposed to the student for learning are those located in the outer fringe of that student's state, as revealed by the assessment. The argument is that a valid assessment leads to a correct gauging of the outer fringe, which should entail efficient learning. The distribution of the conditional probabilities of learning successes are displayed and the median of that distribution is .92. (Falmagne et al., 2007, p. 19)

Procedure

Organizational approvals to access and use student data were obtained prior to starting the research. FERPA regulations were followed to maintain student privacy. The researcher did not interact with any of the research participants throughout this study and all necessary data were already archived in district student information systems. Upon receiving IRB approval from UMKC, the researcher contacted the district administration to obtain written permission for collecting and/or accessing archived data in order to conduct the study. There was no need to inform the participants regarding their participation in the study either, because they were already in naturally formed treatment and control group environments and the treatment had already been introduced to the students and completed during the 2014-2015 school year, prior to research taking place. The researcher has simply

accessed student demographic and achievement data from districts' student information systems and conducted the experiment thereafter.

The research design is a quasi-experimental, pretest-posttest design, with control and treatment groups. All of the treatment group participants came from one district, and likewise, all of the control group participants came from the other district. Because the treatment program had already been established in one of the selected districts and it did not exist in the other selected district, the researcher did not have to ask any of the participants to do anything. The treatment group already knew about the treatment program. The control group had no knowledge or access to the treatment program, since they have no connection to the participants in the treatment group. Therefore, this research study naturally avoids any possible threats to validity that are typically seen in most experimental designs where there is a control versus treatment group.

There was no need to compensate participants for their involvement in this research. However, each school who was implementing the treatment program had similar strategies to provide incentives to participating students to make sure that students used the program with fidelity and benefitted from it. These incentive strategies included giving out achievement certificates to students when they reached certain milestones within the program, recognizing them in public, small gifts, and class parties (pizza, ice-cream, etc.) when the entire class met their performance goals.

Analysis

This quantitative study examined the effectiveness of a cognitive tutoring and assessment software (ALEKS) on math achievement as measured by norm-referenced (NWEA MAP) assessments. The study follows a pretest-posttest quasi-experimental design

with a control and a treatment group of middle school students from two urban school districts in the south.

The following three hypotheses were tested in this study:

H₀₁: There is no statistically significant difference in mean spring NWEA mathematics scores between students who received a regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₂: Among ALEKS users, there is no statistically significant relationship between time spent on ALEKS and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₃: Among ALEKS users, there is no statistically significant relationship between PIE mastery percentage and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

In order to test the first hypothesis, spring RIT scores from the NWEA MAP math assessment are used as the dependent variable. Previous (fall) RIT scores from the NWEA MAP math assessment are used as a covariate to control for prior skill level. The independent variable is ALEKS use (i.e., ALEKS users as treatment group, and non-users as the control group).

To test the 2nd and 3rd hypotheses, a two – block regression model is used. In the first block, previous (fall) RIT scores from the NWEA MAP math assessment are used as a control variable and regressed onto spring RIT scores (dependent variable). In the second block, the significance of the beta weights for time spent on ALEKS instruction and PIE mastery percentage are used to test the hypotheses.

The following research questions are addressed in this study:

1. Does use of ALEKS math software improve student achievement in mathematics?
2. How effective is ALEKS in identifying and closing knowledge gaps of students in math?
 - a. What is the relationship between time spent in ALEKS and latest test performance?
 - b. What is the relationship between concept mastery and latest test performance?

Here, time spent and concept mastery are independent variables and the latest test performance is our dependent variable. ALEKS maps out a personal learning plan illustrated as a pie. This pie includes all the topics and objectives a student is expected to master. As students work through ALEKS, they periodically take an assessment to prove mastery. When they prove mastery on a set of learning objectives pertaining to a topic, the student's knowledge gaps in regards to that topic are closed. The latest assessment includes not only problems students are practicing lately, but also a comprehensive set of problems that they have been studying since the beginning of the course. That makes the latest assessment a comprehensive one where we can draw conclusions on how effectively a student's knowledge gaps are identified (based on their pie) and closed (based on mastery and pie completion). The relationship between concept mastery and latest test performance will help us determine how much of an impact concept mastery has on student performance on the latest assessment, thereby the program's effectiveness in closing knowledge gaps.

The following analyses were used in this research:

- Analysis of Covariance (ANCOVA)
- Descriptive Analysis: for mean/median comparisons.

- Correlational Analysis: to investigate relationships between dependent and independent variables and their effect sizes.
- Multiple Regression Analysis: to uncover joint effect of independent variables on math achievement.

Before performing any data analysis, a missing values analysis were conducted. It is very common in pretest-posttest designs that a number of participants miss either part of the test, which results in missing data. The researcher followed the most traditional treatment for missing data and used listwise deletion where necessary.

IBM Statistical Package for the Social Sciences (SPSS) version 23 was used for the analysis, with the alpha level set to 0.05. For the first hypothesis, ANCOVA analysis will allow the researcher to compare mean student performance on the posttest for each participant group (control vs. treatment), with a pretest score used as a controlling covariate. In a recent study conducted by Hyer and Waller (2014), various analytic techniques for two-group, pre-post repeated measures designs were compared and the researchers concluded that ANCOVA was one of the effective statistical analysis methods to test differences on post-test controlling for pre-test. Using ANCOVA for comparing results of two groups controlling for pretest is a classical way of statistical analysis in this type of situation. Reducing the error variance and eliminating systematic bias are main reasons for using pretest scores as a covariate in ANCOVA with a pretest-posttest design (Dimitrov & Rumrill, 2003).

Linear regression analysis is used in this research as well. Regression analysis will help the researcher predict which independent variable in ALEKS has the greatest effect on math achievement. A two – block regression model is used. In the first block, previous (fall) RIT scores from the NWEA MAP mathematics assessment is used as a control variable and

regressed onto spring RIT scores (dependent variable). In the second block, the significance of time spent on ALEKS instruction and PIE mastery percentage is used to test the 2nd and 3rd hypotheses.

Propensity matching the treatment and control groups based on demographic factors such as gender, ethnicity, socio-economic status, grade level, and participation in special programs (i.e., ESL, Special Education, and Gifted/Talented) alleviates the need to control for these variables when testing hypotheses based on group membership. Propensity matching assigns individuals a propensity score based on their demographic characteristics. For each subject in the treatment group, control subjects were randomly selected from a larger control group pool if their propensity score was nearest to or matched a treatment group subject. For this study, there was a randomly-selected control group member that matched each treatment group subject. This resulted in subjects from both groups being demographically almost identical, thus alleviating potential bias attributable to demographic dissimilarity.

In summary, to tackle the first research question, we engage in ANCOVA analysis to determine the statistical significance of the difference in means scores between students who use ALEKS and those who do not. We also do regression analysis to determine the relationship between concept mastery and time spent in ALEKS instruction and posttest mathematics achievement scores.

Ethical Considerations

The researcher used student data from the school districts he works for as a district administrator. This might bring up some ethical considerations to be addressed in this study. First and foremost, the researcher ensures that all necessary measures are taken to maintain

student privacy per FERPA regulations and takes all necessary precautions to eliminate the possibility of matching or tracing any achievement and/or demographic data to individual students. The researcher has carefully selected the most appropriate statistical analysis and methods to maintain integrity of the research and meet ethical expectations for quantitative research.

CHAPTER 4

RESULTS OF ANALYSES AND CONCLUSIONS

As stated in Chapter 3, three hypotheses were tested in this study:

H₀₁: There is no statistically significant difference in mean spring NWEA mathematics scores between students who received a regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₂: Among ALEKS users, there is no statistically significant relationship between time spent on ALEKS and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H₀₃: Among ALEKS users, there is no statistically significant relationship between PIE mastery percentage and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

Correlations between variables and descriptive statistics for the variables used will be provided at the start of this chapter, followed by hypothesis tests and the conclusions derived therefrom.

Characteristics of Treatment and Control Groups

Using propensity score matching allows the researcher to create a randomly sampled control group that is demographically matched to the treatment group. Although random sampling gives the appearance that a truly experimental design was implemented, all selection of treatment and control subjects was performed post hoc with existing data.

Despite the random selection of demographically-matched control subjects' post-hoc to avoid

demographically related bias, subjects could not be randomly assigned to control & experimental groups before the experimental learning conditions were applied: hence, this study is deemed quasi-experimental. Given a control – pool large enough, a control group that is very similar to the treatment group can be created. For this study, demographic variables such as race, sex, special education, economic disadvantage (free or reduced lunch status), gifted/talented status, Limited English proficiency status (LEP) and student grade level were chosen as the matching factors. An almost – exact match was obtained, as shown in the following tables. The only discrepancy in matching was that an American Indian/Alaska native could not be found from the control pool.

Table 4.1.

Control/treatment group participation by grade level

Grade	Percent		Frequency	
	Control	Treatment	Control	Treatment
5 th	15.3	15.3	85	85
6 th	28.3	28.3	157	157
7 th	18.9	18.9	105	105
8 th	22.7	22.7	126	126
9 th	14.8	14.8	82	82
Total	100.0	100.0	555	555

Table 4.1 shows 100% correspondence between treatment and control groups based on grade level.

Table 4.2.

Control/treatment group participation by race

	Percent		Frequency	
	Control	Treatment	Control	Treatment
AI/AN	0.0	.2	0	1
Asian	1.1	1.1	6	6
Black	8.8	8.8	49	49
Hispanic	77.5	77.5	430	430
Multi-ethnic	1.8	1.6	10	9
White	10.8	10.8	60	60
Total	100.0	100.0	555	555

Table 4.2 shows 100% correspondence between treatment and control groups based on student race.

Table 4.3.

Control/treatment group participation by sex

	Percent		Frequency	
	Control	Treatment	Control	Treatment
Male	53.2	53.2	295	295
Female	46.8	46.8	260	260
Total	100.0	100.0	555	555

Table 4.3 shows 100% correspondence between treatment and control groups based on student sex.

Table 4.4.

Control/treatment group participation by special – education status

	Percent		Frequency	
	Control	Treatment	Control	Treatment
SPED_NO	96.6	96.6	536	536
SPED_YES	3.4	3.4	19	19
Total	100.0	100.0	555	555

Table 4.4 shows 100% correspondence between treatment and control groups based on student special – education status.

Table 4.5.

Control/treatment group participation by economic disadvantage (F/R lunch status)

	Percent		Frequency	
	Control	Treatment	Control	Treatment
ED_NO	34.8	34.8	193	193
ED_YES	65.2	65.2	362	362
Total	100.0	100.0	555	555

Table 4.5 shows 100% correspondence between treatment and control groups based on student economic disadvantage (F/R lunch status).

Table 4.6.

Control/treatment group participation by Limited English Proficiency status

	Percent		Frequency	
	Control	Treatment	Control	Treatment
LEP_NO	94.1	94.1	522	522
LEP_YES	5.9	5.9	33	33
Total	100.0	100.0	555	555

Table 4.6 shows 100% correspondence between treatment and control groups based on student Limited English Proficiency status.

Table 4.7.

Control/treatment group participation by Gifted/Talented status

	Percent		Frequency	
	Control	Treatment	Control	Treatment
G/T_NO	93.7	93.7	520	520
G/T_YES	6.3	6.3	35	35
Total	100.0	100.0	555	555

Table 4.7 shows 100% correspondence between treatment and control groups based on student Gifted/Talented status.

Tables 4.1 – 4.7 show an extremely high correspondence of treatment/control group participation based on demographic variables that are historically related to educational achievement. This high correspondence between the treatment/control groups obviates the need to control for these variables during analysis.

Statistical Analyses & Data Tables

Table 4.8 shows the Pearson r correlations among selected variables used in this study.

Table 4.8.

Correlations among selected variables used in the study

Variable	Statistic	RIT Score Spring	RIT Score Fall	ED (Y/N)	LEP (Y/N)
RIT Score Spring	Pearson r	1	.849**	-.158**	-.168**
	Sig. (2-tail)		0.000	0.000	0.000
	N	1110	1110	1110	1110
RIT Score Fall	Pearson r	.849**	1	-.154**	-.173**
	Sig. (2-tail)	0.000		0.000	0.000
	N	1110	1110	1110	1110
ED (Y/N)	Pearson r	-.158**	-.154**	1	.088**
	Sig. (2-tail)	0.000	0.000		0.003
	N	1110	1110	1110	1110
LEP (Y/N)	Pearson r	-.168**	-.173**	.088**	1
	Sig. (2-tail)	0.000	0.000	0.003	
	N	1110	1110	1110	1110
Gifted/Talented (Y/N)	Pearson r	.239**	.229**	-.153**	-.065*
	Sig. (2-tail)	0.000	0.000	0.000	0.030
	N	1110	1110	1110	1110
Total Time ALEKS	Pearson r	0	-0.0116	0.00415	0.03791
	Sig. (2-tail)	0.740	0.786	0.922	0.373
	N	555	555	555	555
Best Pie Mastery pct	Pearson r	.230**	.155**	-0.0593	0.01386
	Sig. (2-tail)	0.000	0.000	0.163	0.745

	N	555	555	555	555
TestPercentile	Pearson r	.913**	.757**	-.133**	-.170**
	Sig. (2-tail)	0.000	0.000	0.002	0.000
	N	555	555	555	555

Table 4.8 Correlations among selected variables used in the study (continued)

Variable	Statistic	Gifted/Talented (Y/N)	Total Time ALEKS	Best Pie Mastery pct	Test Percentile
RIT Score Spring	Pearson r	.239**	.014	.230**	.913**
	Sig. (2-tail)	0.000	0.740	0.000	0.000
	N	1110	555	555	555
RIT Score Fall	Pearson r	.229**	-.012	.155**	.757**
	Sig. (2-tail)	0.000	0.786	0.000	0.000
	N	1110	555	555	555
ED (Y/N)	Pearson r	-.153**	.004	-0.0593	-.133**
	Sig. (2-tail)	0.000	0.922	0.163	0.002
	N	1110	555	555	555
LEP (Y/N)	Pearson r	-.065*	.038	0.014	-.170**
	Sig. (2-tail)	0.030	0.373	0.745	0.000
	N	1110	555	555	555
Gifted/Talented (Y/N)	Pearson r	1	-.150**	.147**	.313**
	Sig. (2-tail)		0.000	0.001	0.000
	N	1110	555	555	555
Total Time ALEKS	Pearson r	-.150**	1.000	.382**	0.00148
	Sig. (2-tail)	0.000		0.000	0.972
	N	555	555	555	555
Best Pie Mastery pct	Pearson r	.147**	.382**	1	.407**
	Sig. (2-tail)	0.001	0.000		0.000
	N	555	555	555	555

Table X8: Correlations among selected variables used in the study (continued)

Variable	Statistic	Gifted/Talented (Y/N)	Total Time ALEKS	Best Pie Mastery pct	Test Percentile
Test Percentile	Pearson r	.313**	.001	.407**	1
	Sig. (2-tail)	0.000	0.972	0.000	
	N	555	555	555	555

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

* . Correlation is significant at the 0.05 level (2-tailed).

Table 4.9.

Descriptive statistics for dependent and independent variables used in the study, by treatment/control group

Statistic	RIT Score Fall		RIT Score Spring		Total Time ALEKS	Best Pie Mastery pct
	Control	Treatment	Control	Treatment	Treatment	Treatment
N	555	555	555	555	555	555
Mean	222.61	223.00	224.61	229.13	2620.58	59.796
Std. Dev.	15.723	15.338	17.670	15.573	1068.655	18.1195
Variance	247.210	235.256	312.231	242.523	1142024.261	328.315
Skewness	-.432	-.439	-.492	-.473	-.391	.182
Kurtosis	.281	.798	.748	.370	-.274	-.359

Figure 4.1 shows the distribution of independent variable fall RIT score by treatment/control group. The distribution for both groups is relatively normal and symmetrical.

Figure 4.1.

Distribution of independent variable fall RIT score by treatment/control group

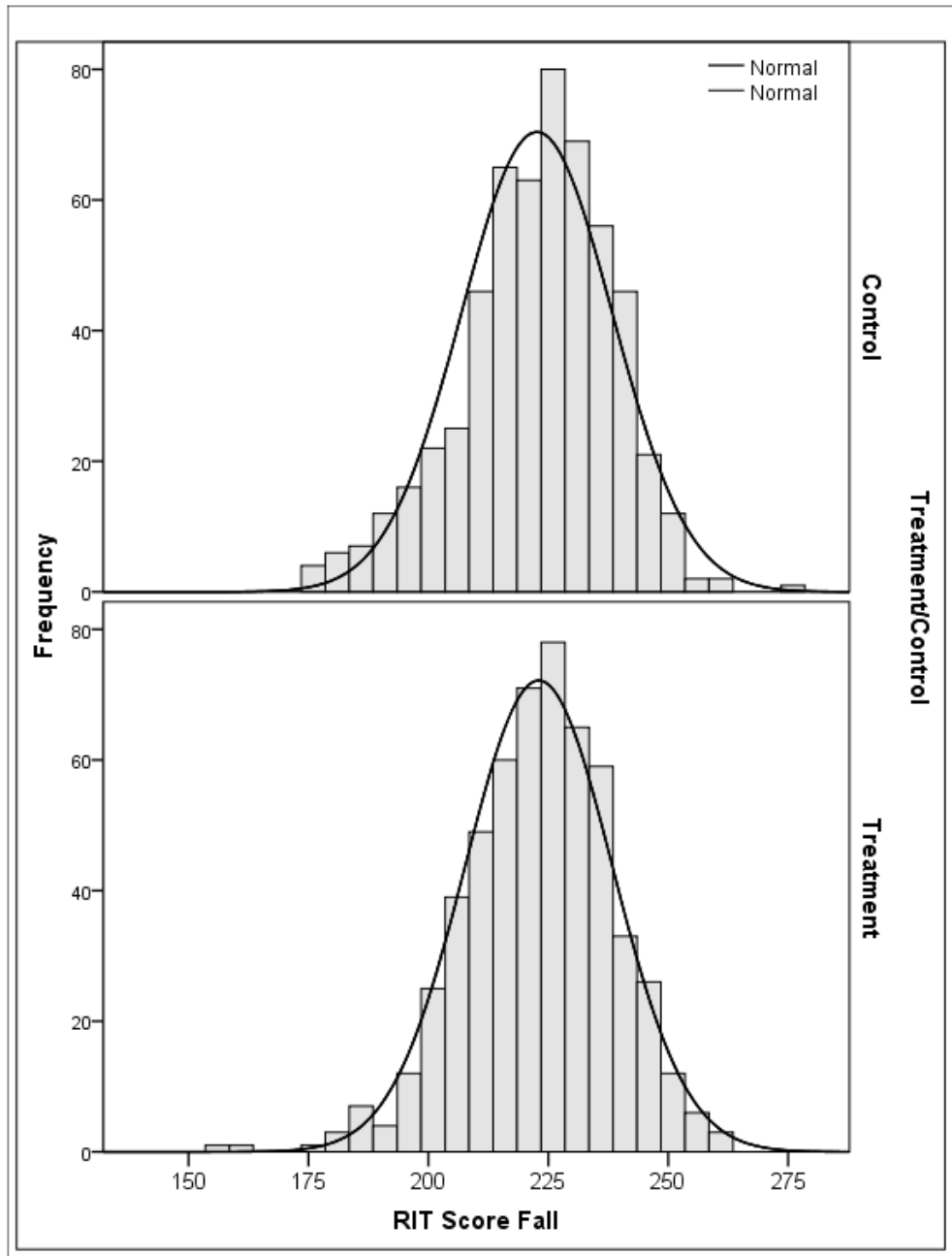


Figure 4.2 shows the distribution of independent variable spring RIT score by treatment/control group. The distribution for both groups is relatively normal and symmetrical.

Figure 4.2.

Distribution of dependent variable spring RIT score by treatment/control group

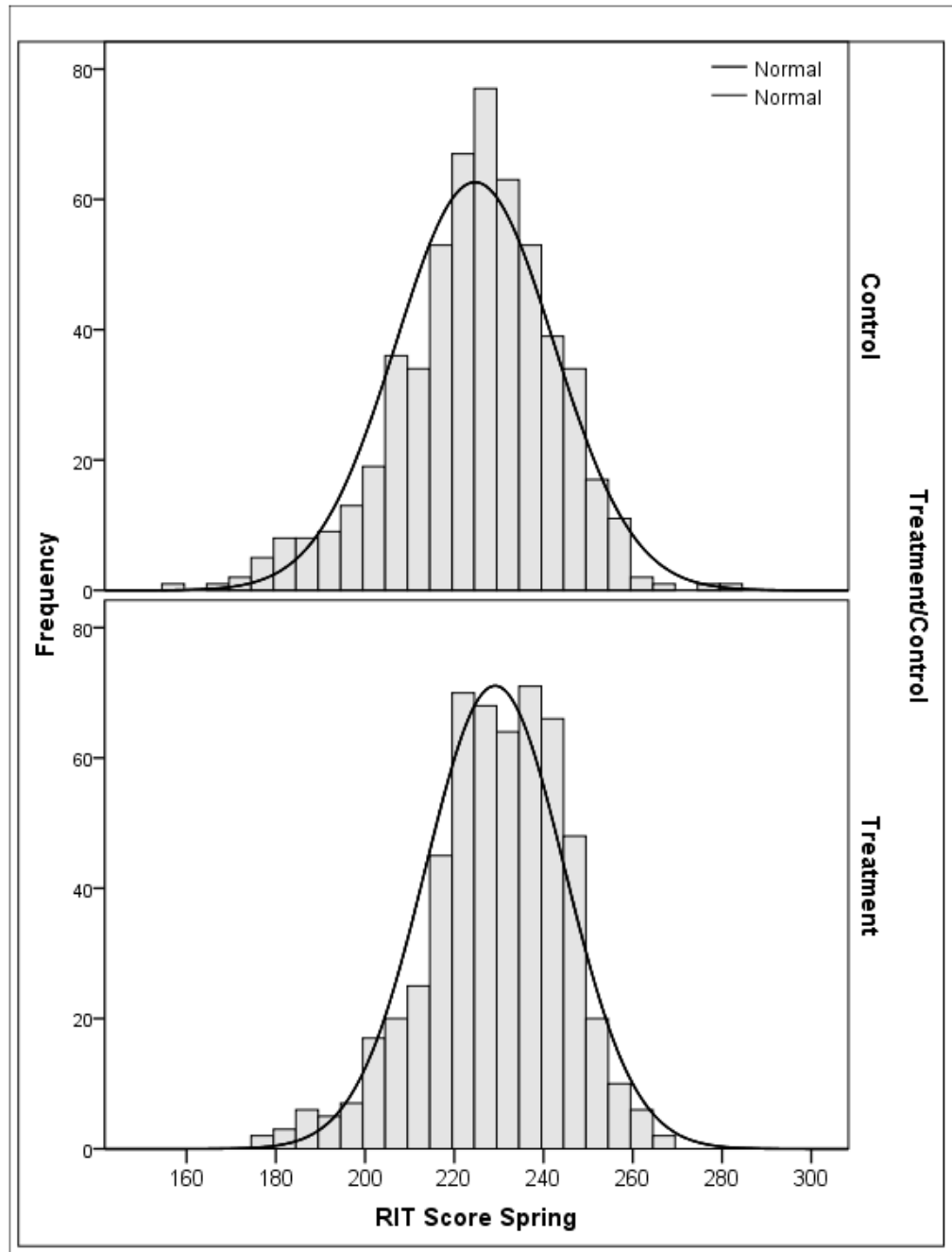
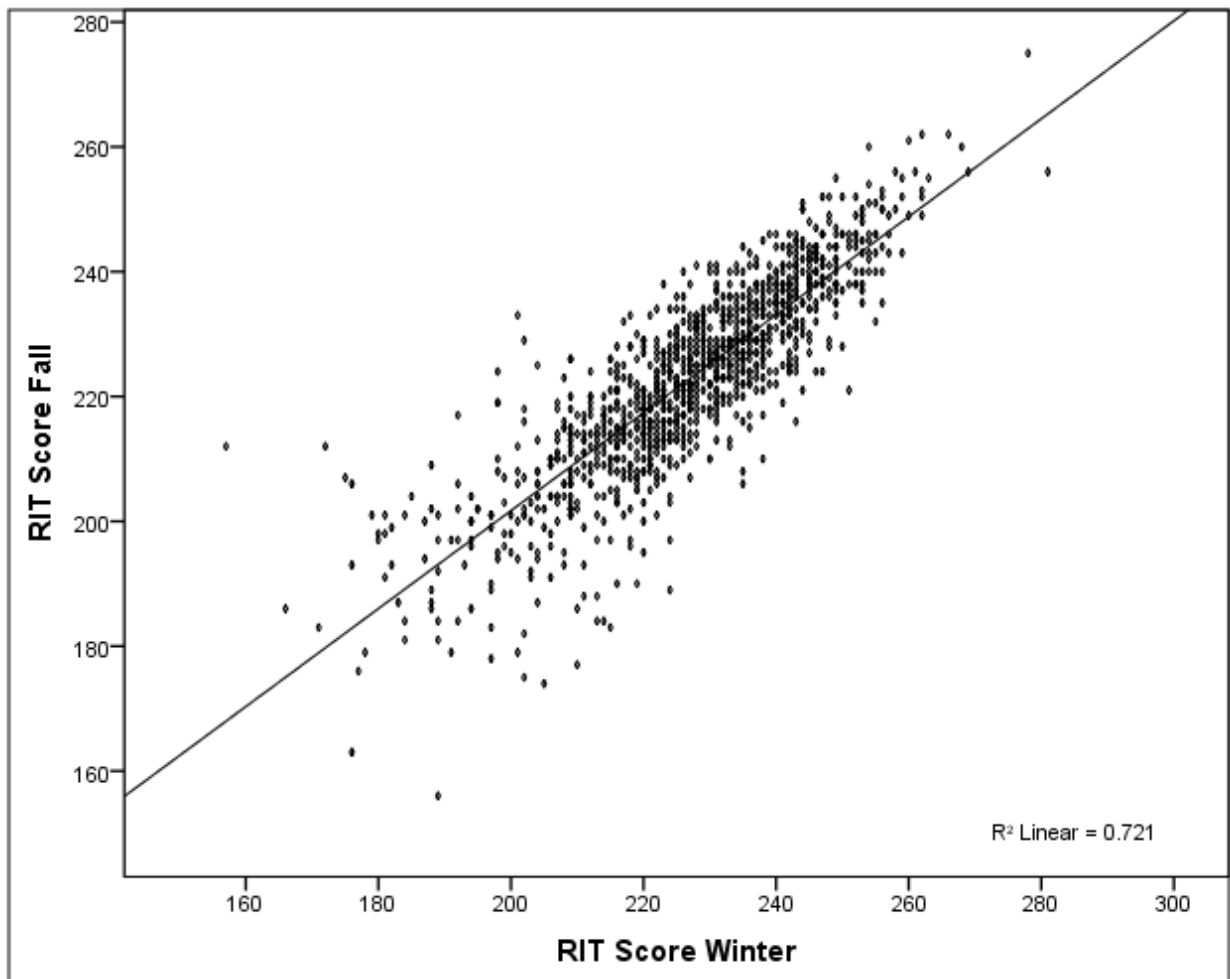


Figure 4.3 shows the relationship between covariate fall RIT score and dependent variable spring RIT score. The relationship between these two variables is quite high ($r=0.849$), making the fall RIT score a good covariate for testing hypothesis number one. Using the fall RIT score as a covariate will hold students' achievement level at the beginning of the study constant, allowing relative gains to be comparable.

Figure 4.3.

Scatter plot of dependent variable spring RIT score and covariate fall RIT score



Hypothesis Testing

The first hypothesis to be tested is:

H_{01} : There is no statistically significant difference in mean spring NWEA mathematics scores between students who received regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

Table 4.10.

Results of ANCOVA analysis, test of hypothesis

Source	Type III SS	df	MS	F	Sig.	Partial Eta Squared	Observed Power ^b
Corrected Model	230559.860 ^a	2	115279.930	1548.044	0.000	.737	1.000
Intercept	2711.233	1	2711.233	36.408	.000	.032	1.000
RIT_F	224897.653	1	224897.653	3020.053	0.000	.732	1.000
Aleks_user	4798.004	1	4798.004	64.430	.000	.055	1.000
Error	82436.199	1107	74.468				
Total	57444375.000	1110					
Corrected Total	312996.059	1109					

a. R Squared = .737 (Adjusted R Squared = .736)

b. Computed using alpha = .05

Table 4.11.

Results of ANCOVA analysis, mean spring mathematics RIT score by group

Group	Mean	SD	N
Control	224.61	17.670	555
Treatment	229.13	15.573	555
Total	226.87	16.800	1110

Table 4.12.

Test for homogeneity of variance

Levene Statistic	df1	df2	Sig.
3.840	1	1108	.051

ANCOVA results show that there is a statistically significant difference in the mean spring mathematics RIT scores (highlighted row Table 4.10), and that the score for the treatment group was higher (highlighted row Table 4.11). The homogeneity of variance between the treatment and control groups is not different (SEE Table 4.12). These results show that there is a statistically significant difference in mean spring NWEA mathematics scores between students who received regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$, and therefore, the null hypothesis may be rejected. Although the results show that the null hypothesis should be rejected, the test has a high level of power due to the very large sample size, and the Eta Squared shows that ALEKS instruction explains 5.5% of the variance in spring NWEA mathematics scores after controlling for fall NWEA mathematics scores. Although some researchers may consider this a small amount of variance explained in a study, educators having the ability to influence 5.5% of achievement in a subject area is not a negligible finding.

The second and third hypotheses will be tested using a single multi – block multiple regression model using the SPSS Linear Regression procedure. These tests involve only the treatment group:

H_{o2} : Among ALEKS users, there is no statistically significant relationship between time spent on ALEKS and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

H_{o3} : Among ALEKS users, there is no statistically significant relationship between PIE mastery percentage and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores, $\alpha \leq .05$.

Table 4.13 specifies the multi-- block regression model. In block 1, fall NWEA mathematics scores are entered into the model to hold them constant. In block 2, time spent in ALEKS and PIE Mastery are entered simultaneously so that their beta weights may be evaluated to test hypotheses 2 and 3.

Table 4.13.

Regression model specification

Variables Entered/Removed^a			
Model	Variables Entered	Variables Removed	Method
1	RIT Score Fall ^b	--	Enter
2	Total Time ALEKS & Best Pie Mastery pct ^b	--	Enter

a. Dependent Variable: RIT Score Spring

b. All requested variables entered.

Table 4.14.

Regression model summary

Model	R	R Square	Adj. R Square	Std. Error of the Estimate
1	.876 ^a	.768	.768	7.506
2	.882 ^b	.777	.776	7.368

a. Predictors: (Constant), RIT Score Fall

b. Predictors: (Constant), RIT Score Fall, Total Time ALEKS, Best Pie Mastery pct

In Table 4.14 the adjusted R square values show that after pretest scores are controlled for, Total Time in ALEKS and Best Pie Mastery pct. Explain only an additional 0.8% of the variance in the posttest scores (see highlighted column).

Table 4.15.

ANOVA results associated with the regression model

Model		SS	df	MS	F	Sig.
1	Regression	103201.306	1	103201.306	1831.724	.000 ^b
	Residual	31156.611	553	56.341		
	Total	134357.917	554			
2	Regression	104448.369	3	34816.123	641.390	.000 ^c
	Residual	29909.548	551	54.282		
	Total	134357.917	554			

a. Dependent Variable: RIT Score Spring

b. Predictors: (Constant), RIT Score Fall

c. Predictors: (Constant), RIT Score Fall, Total Time ALEKS, Best Pie Mastery pct

Table 4.16.

Coefficients resulting from the regression model.

Model	Variable	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	30.691	4.647		6.604	.000
	RIT Score Fall	.890	.021	.876	42.799	.000
2	(Constant)	29.641	4.635		6.395	.000
	RIT Score Fall	.874	.021	.860	42.157	.000
	Total Time ALEKS	.000	.000	-.015	-.689	.491
	Best Pie Mastery pct	.088	.019	.102	4.639	.000

a. Dependent Variable: RIT Score Spring

Table 4.16 (highlighted rows) shows that the Beta for time spent in ALEKS was not significant, and therefore time spent in ALEKS had no statistically significant effect on mean posttest scores controlling for pretest scores. Null hypothesis 2 is retained. Table 4.16 does show, however, that Pie mastery percent does have a statistically significant effect on mean posttest scores controlling for pretest scores. Null hypothesis three is rejected.

Performance Analyses of Demographic Groups

Because ANCOVA analysis shows statistical evidence of significance of the treatment program on mathematics achievement of students, the researcher has decided to look at analysis of student performance data broken down by various demographic groups such as race, gender, grade level, and educational needs. Tables 4.17 – 4.20 below demonstrate performance analysis of student groups based on demographics and educational needs through ANCOVA results. These further analyses will provide more in depth insight

into whether the significance of ALEKS as an adaptive learning software can be generalized for different student groups as typically reported in state and federal accountability measures.

Table 4.17.

Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by grade level

Grade Level	Group	Mean	SD	N
5 th Grade	Control	212.71	16.90	85
	Treatment	214.45	16.47	85
	Total	213.58	16.90	170
6 th Grade	Control	217.46	16.76	157
	Treatment	224.30	15.55	157
	Total	220.88	16.75	314
7 th Grade	Control	227.28	16.57	105
	Treatment	231.23	15.54	105
	Total	229.25	16.57	210
8 th Grade	Control	234.31	16.62	126
	Treatment	236.64	15.53	126
	Total	235.48	16.62	252
9 th Grade	Control	232.33	16.81	82
	Treatment	239.35	15.57	82
	Total	235.84	16.80	164

Highlighted rows in Table 4.17 show that treatment group has outperformed control group in each grade level. The impact of the treatment has been largest in 6th grade and 9th

grade as we see roughly a 7-point performance difference between ALEKS users and non-ALEKS users. The smallest mean difference in NWEA MAP mathematics spring scores between control and treatment group was observed in 5th grade (1.74 RIT score points) and in 8th grade (2.33 RIT score points) respectively. From this analysis, we can conclude that each participating grade level in the treatment was able to make positive gains in mathematics achievement via ALEKS instruction and outperform their peers in control group.

Table 4.18.

Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by special populations

Category	Group	Mean	SD	N
Special Education (SPED)	Control	202.16	16.96	19
	Treatment	211.89	15.05	19
	Total	207.03	16.61	38
Limited English Proficiency (LEP)	Control	213.45	16.65	33
	Treatment	217.88	15.74	33
	Total	215.67	16.65	66
Gifted/Talented (GT)	Control	239.51	16.69	35
	Treatment	245.11	15.65	35
	Total	242.31	16.69	70
Economically Disadvantaged (ED)	Control	222.16	16.80	362
	Treatment	227.70	15.56	362
	Total	224.93	16.80	724

Highlighted rows in Table 4.18 show that treatment group has outperformed control group in each special programs category. The impact of the treatment has been largest among special education students as we see a 9.73 RIT-score difference between ALEKS users and non-ALEKS users. This is considered a very large performance difference based on normative RIT-score guidelines of NWEA MAP assessment and can be interpreted as the treatment group making instructional gains about 1.5 to 2 grade levels in mathematics in a single school year. That is a strong indicator of closing the achievement gap effectively. The mean differences in NWEA MAP mathematics spring scores between control and treatment group among economically disadvantaged and gifted/talented students was very similar, with both treatment groups outperforming their corresponding control groups by more than 5.5 RIT score points. This significant performance difference observed nearly as identical in the historically highest performing student group (gifted/talented) and also in the historically lowest performing student group (economically disadvantaged) confirms the findings of this study that ALEKS is a promising adaptive learning software which can be used both as an intervention tool for low performing students and also as an enrichment program for high performing students. It is also worth noting that students with limited English proficiency made significant gains in math achievement via ALEKS as the treatment group in this category outperformed the control group with a nearly 4.4 RIT-score difference in spring NWEA MAP mathematics test. From this analysis, we can conclude that each participating special-needs groups in the treatment was able to make positive gains in mathematics achievement via ALEKS instruction and outperform their counterparts in control group.

Table 4.19.

Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by race/ethnicity

Race	Group	Mean	SD	N
African American	Control	221.63	16.67	49
	Treatment	222.12	15.75	49
	Total	221.88	16.67	98
Hispanic	Control	223.40	16.80	430
	Treatment	228.82	15.57	430
	Total	226.11	16.80	860
White	Control	232.32	16.68	60
	Treatment	234.30	15.90	60
	Total	233.31	16.68	120

Table 4.19 shows performance analysis across three racial/ethnic student groups. In this table several racial/ethnic groups such as American Indian / Alaskan Native, Asian, and multi-ethnic groups are excluded due to their small size within the sample. This data reveals that among Hispanic students the treatment effect was largest as we see a difference of 5.42 RIT-score points between the treatment and the control group. Among white students, the treatment group outperformed the control group by roughly 2 points, while the performance difference within the African American students are observed to be minimum with a less than 1-point RIT-score between the treatment group and the control group. Nonetheless, it is worth noting that treatment group has outperformed control group across all participating ethnic populations. The significant gains made by the Hispanic students is of particular

interest to the participating school district due to the fact that the district serves predominantly Hispanic students in its community.

Table 4.20.

Results of ANCOVA analysis, mean spring mathematics RIT score by group, broken down by gender

Gender	Group	Mean	SD	N
Male	Control	224.20	16.80	295
	Treatment	229.04	15.57	295
	Total	226.62	16.80	590
Female	Control	225.08	16.68	260
	Treatment	229.23	15.57	260
	Total	227.15	16.68	520

Performance analysis based on gender is shown in Table 4.20. According to this data, both male students and female students made significantly larger gains in mathematics due to treatment effect. Male students using ALEKS outperformed male students in the control group by 4.84 RIT-score points in spring NWEA MAP mathematics assessment, while female students in the treatment outperformed females in the control group by 4.15 RIT-score points.

Conclusions and Summary

The results of analysis show that ALEKS instruction has a statistically significant positive effect on students' end-of-year mathematics scores when beginning-of-year scores are held constant, and that participation in ALEKS instruction explains approximately

5.5% of the variance in end-of-year mathematics scores. Despite the finding that ALEKS use was significantly related to positive mean posttest scores controlling for pretest scores, Table 4.16 (highlighted rows) shows that the Beta for actual time spent in ALEKS was not significant, and therefore time spent in ALEKS had no statistically significant effect on mean posttest scores controlling for pretest scores.

With a deeper dive into demographic group performance analysis, the results reveal that treatment group participating in mathematics instruction via ALEKS outperformed the control group across all grade levels (6 through 9), both gender groups, all racial/ethnic groups (African American, Hispanic, and White), and all special programs (Special Education, Limited English Proficient, Gifted/Talented, and Economically Disadvantaged) in spring NWEA MAP mathematics assessment. The results also show that pie mastery percentages are a statistically significant predictor of subsequent end-of-year mathematics scores, and that they can be a useful benchmark indicator during the year's instruction.

In chapter 5, the educational implications of these results will be discussed, along with implications for future research.

CHAPTER 5

DISCUSSION

The results of this study, presented in detail in chapter 4, are interpreted in this chapter with further discussion in light of the research questions addressed in this study and in conjunction with other relevant literature. Educational implications of the findings of this study for K-12 schools and districts will be in the center of our discussions in this section, as well as implications for further research.

Within the last two decades, public education system in the United States has been subject to various reform efforts both at the federal and at the state levels. While many researchers and practitioners (Horn & Staker, 2014; Vander Ark, 2012; Domenech, Sherman, & Brown, 2016) argue that American public education must be completely redesigned rather than applying changes to its current infrastructure, policymakers are in favor of maintaining the status quo and allowing educational reform in small doses not to disrupt the century old model currently in place. Obama administration has been particularly more bold and innovative in providing funding and flexibility for state education agencies and local education agencies to design and implement new learning models in K-12 education. One of the main goals of these reform efforts has been boosting high school graduation and college readiness rates across the country.

The Department of Education under Secretary Duncan's leadership has proudly announced that high school graduation rates have reached an all-time high record of 82 percent in 2013-2014 school year. "In school year 2013-14, the adjusted cohort graduation rate (ACGR) for public high schools rose to an all-time high of 82 percent" (NCES, 2016). In

fact, the graduation rates have reached another record high of 83 percent in the 2014-2015 school year, marking the fifth straight record-setting year under Obama's presidency. Despite these tremendous improvements in graduation rates and rising emphasis on college-readiness, recent reports on student performance based on SAT and ACT scores suggest that most high school graduates are ill-prepared for the academic rigor of college, especially when it comes to mathematics (Adams, 2015). According to a recent report by Saxe and Braddy (2015), about 50 percent of students are not passing college algebra with a grade of C or above each year. These students are either deemed college-ready in mathematics according to their SAT or ACT scores, or they have taken and successfully passed one or more remedial math courses before they were allowed to take college algebra. These high failure rates spurred universities to take a proactive approach to math classes.

ALEKS was indeed originally created as an intelligent tutoring system for college math classes in the 90s. It was intended to address the very aforementioned problems with college-readiness in math. As the program gained success and popularity, the developers expanded the program to reach K-12 students in an effort to start closing their knowledge and skills gaps in mathematics at earlier ages. The advancements in technology made personalized learning more possible at scale. Horn and Staker (2014) describe digital curricula and online adaptive learning software as the backbone of blended learning and portray how personalized learning comes to life when the infrastructure that supports selected hardware and software is in place with an anytime anywhere learning culture established throughout the school community. Vander Ark (2012) also writes how digital learning and smart algorithms change the way students are learning and causing dramatic shifts in job markets across the globe, hence necessitating drastic changes how schools are run and

learning environments are designed. Domenech, Sherman, and Brown (2016) agree with these ideas as they further argue how shifts in demographic landscape of our country demands personalizing 21st century education for all students, specifically to close the achievement gap between economically disadvantaged students and their affluent counterparts across all U.S. cities and communities in urban, rural, and suburban settings.

Since we live in a performance-driven world under strict accountability guidelines, funders, policy makers, and those in positions of power demand to see hard evidence that any reform initiative will work in education before they authorize large scale projects. Therefore, new innovative approaches to personalized learning are still evaluated by their impact on student achievement and growth, measured by standardized state assessments under NCLB and ESEA. Although many innovators in K-12 education (Horn & Staker, 2014; Vander Ark, 2012; Domenech, Sherman, & Brown, 2016) argue that standardization of curriculum and assessments totally contradicts the basic tenets and premise of personalized learning, current education system is still bound by accountability guidelines that hinder personalization rather than enable it. Seat-time requirements, pacing guides, bell schedules, the sage-on-the-stage stance of teachers in the classrooms, and many other traditional norms and policies in public education still support the one-size-fits-all approach to education and make it inherently more difficult to personalize learning for all students. In spite of these roadblocks, effective teachers and school leaders are able to mobilize their resources with the help of evolving edtech products to launch blended learning environments within the constraints of current realities of public education system and its bureaucracies.

Horn and Staker (2014) described various blended learning models such as flipped classroom, station rotation, and enriched virtual that are all making it possible to personalize

the instruction to individual students through smart use of technology and making a combination of online and offline digital curricula available to students anytime anywhere. Vander Ark (2012) describes how adaptive online learning programs such as ALEKS can mimic the behavior of an effective human tutor and provide instant feedback to students to help them move forward and master the content that is readily available to each individual learner at the right time and at the appropriate pace. These advancements help students who have large learning gaps catch up with their peers at an optimum pace and enable other students who may have previously mastered the content to move forward and be challenged with a more rigorous content appropriate to their level. This way, students who are behind are not frustrated with the curriculum and those who are ahead are not feeling bored and held back until everybody gets the same content.

The digital revolution in education coupled with breakthroughs in cognitive science and artificial intelligence is not meant to replace the teacher by no means. Blended learning is not about providing personalized instruction through computers and technology alone, but rather it is about empowering teachers with real time actionable data and affording them the time and space to tailor their instructional plans and intervene with their students in smaller groups based on timely feedback flowing from adaptive online programs like ALEKS (Horn & Staker, 2014; Vander Ark 2012). Technology becomes an enabler for personalized learning and helps schools and districts scale blended learning in a way it was not possible before. Leveraging the power of technology brings numerous advantages to teachers, such as real-time actionable data to inform their instructional decision making and grouping their students based on their current and ever changing state of knowledge as they interact with adaptive learning programs throughout the school day. Blended learning requires a dramatic

shift in teacher's role from being the 'sage on the stage' to a 'guide on the side' which puts teachers in more of a learning facilitator role and pushes students to take ownership of their own learning and build student agency, Horn and Staker (2014) explain. The 'command-and-control' type classrooms in traditional schooling leaves students unprepared to successfully navigate in an ever changing, unpredictable world. Students are robbed of opportunities when they are taught like robots to read textbooks and find answers to standardized test questions. In blended classrooms and other personalized learning models, teachers see themselves as an innovator and change maker of education, they act like an instigator of thought and create learning sparks in children through their guidance. All teachers must be exposed to this in order to effectively prepare the future leaders of our society.

Many of the federal, state, and even philanthropic grant programs to empower innovation in public education today have distinct requirements to serve high need students and schools and lift their achievement with significant academic gains. Race to the Top (RTTT) and Investing in Innovation (I3) are prime examples of such competitive federal grant opportunities that are designed to promote blended learning and technology investment in K-12 public education. While the end goal in such programs is to boost college- and career-readiness rates, the progress is measured by student achievement and growth data through standardized state assessments, and the grantees are required to report their data broken down by eight different student groups based on demographics and educational needs, many of which are specifically created to see impact of funding on high-needs student populations. The U.S. Department of Education defines the term 'high-needs students' as follows:

High-needs students: Students at risk of educational failure or otherwise in need of special assistance and support, such as students who are living in poverty, who attend

high-minority schools (as defined in the Race to the Top application), who are far below grade level, who have left school before receiving a regular high school diploma, who are at risk of not graduating with a diploma on time, who are homeless, who are in foster care, who have been incarcerated, who have disabilities, or who are English learners. (U.S. Department of Education, 2012)

From this definition, students who are eligible for free or reduced priced lunch are automatically considered to be high-needs students due to being economically disadvantaged (ED), as well as students who receive special education services (SPED) and those who are limited English proficiency (LEP). Additional students who are not classified in any of these groups may still be considered high-needs based on their academic progress and/or living conditions. The participant sample for this study consists of at least 67 percent high-needs students (742 out of 1110) solely based on their economically disadvantaged, special education, and/or limited English proficiency classifications. This rate might be higher if we consider the possibilities of academic failures, learning gaps, and living conditions of participating students, but because we do not have access to such data on the participants it is not possible to get an exact number and percentage of high-needs students in our sample. Nonetheless, a minimum of 67% of high-needs students is considered to be a large enough percentage for schools and makes those schools eligible for being regarded as a high-needs school overall. Therefore, the sample meets ideal characteristics of testing out educational reforms and measuring their impact and effectiveness.

Discussion of Findings

Does use of ALEKS math software improve student achievement in mathematics?

Yes, the use of ALEKS in general has significantly improved student achievement in mathematics. Although both groups showed improvement on the post-test, the instructional

gains made by the treatment group are significantly higher than that of the control group. The details of the statistical analyses are presented in the summary section next.

Does increased time spent with ALEKS correlate with improved student achievement in mathematics?

No, increased time spent with ALEKS does not necessarily correlate with improved student achievement in mathematics. More detailed answer to this question is provided in the following sub-questions tied to this research question.

What is the relationship between time spent in ALEKS and latest test performance?

Based on the statistical analysis conducted in this research, there has not been found any strong correlation between time spent in ALEKS and students test performance in mathematics. Although theoretically students are expected to perform better on math assessments as they spend more time with ALEKS, but there is no way to assess the quality of the time students spend on the program. They may simply get off task at times or leave the program running and engage in other activities especially when they are not supervised by their teachers.

What is the relationship between concept mastery and latest test performance?

The concept mastery, also referred to as pie mastery, has turned out to be a strong predictor of math achievement. Unlike time spent on ALEKS, we can see a strong relationship between concept mastery and test performance. Specifically, the more concepts a student demonstrates mastery on, the more likely he/she is to make larger gains on his/her math test.

Summary

This study aims to uncover the impact of an online adaptive learning software (ALEKS) on math achievement and growth of middle school students in a southern urban school district. This research measured student math achievement and growth using a norm-referenced assessment known as NWEA MAP in mathematics. Students were assessed once in the beginning of the school year in early fall and those scores were recorded as Fall RIT scores. Students were then divided into two groups for control and treatment. Both groups are taught the same curriculum with their teachers going through the same training program. Treatment group had an additional curricular resource using the adaptive ALEKS intelligent tutoring system integrated into the mainstream math curriculum. Later in the spring semester, both groups were administered the NWEA MAP test in mathematics again by the end of the school year. Those scores were recorded as Spring RIT scores. In this research, we looked at how student math achievement improved from fall to spring measured by NWEA MAP and whether or not these improvements can be attributed to adaptive learning software ALEKS, and specifically what data points in ALEKS predicts student achievement significantly.

Demographic indicators such as grade level, gender, race/ethnicity, learning disabilities, language acquisition levels, and socio-economic factors are known to be historically affecting educational achievement of students. Therefore, many educational research studies control for these variables during analysis. To eliminate demographic indicator bias, such variables are usually matched to a high level of correspondence as much as possible. In our case, control and treatment group participation data was matched with a nearly perfect correspondence based on demographic variables, thereby eliminating demographic bias altogether and removing the need to control for demographic variables during statistical analysis.

In the first hypothesis of this study, we assume that there is no statistically significant difference in mean spring NWEA mathematics scores between students who received regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores. In order to test this hypothesis, we first examined the relationship between the pre-test (Fall RIT) and post-test (Spring RIT) scores. The correlation between the two turned out to be very strong, which makes it an ideal case to use pre-test scores for holding student achievement constant in the beginning of the school year and thus analyzing changes in independent variable (post-test scores) to compare relative instructional gains in math.

The analysis revealed statistically significant gains in math achievement from fall to spring in treatment group. Mean Spring RIT score was 224.61 for control group, while it was 229.13 for the treatment group, which is a 4.52 increase in RIT scores. This increase is considered to be a large instructional gain for middle school students made by the treatment group helping those students close the achievement gap and improve their grade level readiness and increase their nationwide percentile ranking measured by the norm-referenced NWEA MAP assessment. Furthermore, the variance between the treatment and control groups is not different, hence homogeneity of variance assumption is verified. As these results show, there is a statistically significant difference in mean spring NWEA mathematics scores between students who received regular math instruction and students who received adaptive math instruction via ALEKS, controlling for previous fall NWEA mathematics scores. Hence, the first null hypothesis may be rejected. One important thing worth

noting here is that the treatment explains only about 5.5% of the variance in Spring RIT scores. Some researchers may consider this a small amount of variance explained by a treatment factor. However, considering the large sample size and thus the high level of power the statistical test holds here, educators would consider it a pretty strong intervention to be able to influence student achievement levels by 5.5% in mathematics.

The second and third hypotheses are listed here as follows:

- Among ALEKS users, there is no statistically significant relationship between time spent on ALEKS and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores.
- Among ALEKS users, there is no statistically significant relationship between PIE mastery percentage and students' spring NWEA mathematics scores, controlling for previous fall NWEA mathematics scores.

Since the second and third hypotheses involve only the treatment group, they were tested using block designs looking at each factor (time spent and pie mastery percentage) separately and then their joint effect on student achievement (Spring RIT scores). The analysis shows that after pretest scores are controlled for, time spent in ALEKS and pie mastery percentage explain only an additional 0.8% of the variance in the post-test scores. Since the joint effect of these two variables has some influence –but not very large- on math achievement, this finding suggests that one of these independent variables might be a more significant predictor as opposed to the other one for math achievement. Further analysis showed that time spent in ALEKS was not significant, while pie mastery percent does have a statistically significant

effect on mean post-test scores controlling for pre-test scores. Therefore, we decide to retain the second null hypothesis, but reject the third one.

Conclusions

The purpose of this study was to test the effectiveness of an adaptive learning software called ALEKS, on mathematics achievement and growth levels. The results of analysis show that mathematics instruction via ALEKS has a statistically significant positive effect on students' math achievement and growth levels measured by a normative end-of-year mathematics assessment when beginning-of-year scores are held constant. Upon observing statistically significant impact of ALEKS, in the second part of this study the researcher looked into various predictors and independent variables in the software to decide which components and data points of the program are best predictors of mathematical success and instructional growth. The conclusions of this study can be summarized as follows:

- ALEKS made a statistically significant impact on mathematics achievement in the treatment group.
 - Participation in adaptive math instruction via ALEKS explains approximately 5.5% of the variance in end-of-year mathematics scores.
- Students receiving math instruction via ALEKS outperformed students who are not using ALEKS in each of the participating grade levels (6 through 9).
 - Students in 6th and 9th grades made the largest gains in mathematics by using ALEKS in end-of-year test scores.

- Both male and female students participating in ALEKS program outperformed their non-ALEKS using counterparts with large instructional gains observed in end-of-year test scores.
- Students receiving math instruction via ALEKS outperformed students who are not using ALEKS across all racial/ethnic groups (African American, Hispanic, and White).
 - Hispanic students made the largest gains in mathematics by using ALEKS in end-of-year test scores.
- Students receiving math instruction via ALEKS outperformed students who are not using ALEKS across all special-needs programs (Special Education, Limited English Proficient, Gifted/Talented, and Economically Disadvantaged).
 - Special Education students made the largest gains in mathematics by using ALEKS in end-of-year test scores. The numbers suggest that students in this group made improvements equivalent to 1.5 to 2 years of growth in a single school year.
 - Gifted/Talented students and economically disadvantaged students achieved roughly the same amount of growth in mathematics making the second largest gains in mathematics by using ALEKS.
- Among ALEKS users, the results show that pie mastery percentages are a statistically significant predictor of subsequent end – of – year mathematics scores, which can be a useful benchmark indicator for teachers and instructional leaders during the year’s instruction.

- Among ALEKS users, the results show that time spent on ALEKS is not a statistically significant predictor of subsequent end – of – year mathematics scores. Therefore, this data point should not be used as a benchmark indicator for teachers and instructional leaders during the year’s instruction.

Implications

Issues of academic growth and equitable access to high quality educational programs has been critical to PK-12 public school reform movements in the United States. Designing blended learning environments to tackle these growth and access issues have gained momentum both in research and practice in public education. Decision makers in PK-12 education are increasingly considering blended learning and adaptive learning technologies as a lever for personalizing instruction for all students. As blended learning becomes an integral component of their academic programs, research studies like this one will provide insight for these decision makers and practitioners. In many cases, educators are trying to make decisions about blended learning to increase student access and growth for equity. According to Picciano (2006), while such efforts are driven by pedagogical reasons “trying to capture the best of online and traditional face-to-face modalities” (p. 99), individual accounts and case studies are not contributing to the research and literature in the absence of adequate data collection processes.

In a recent blog series about what parents want for their child(ren)’s education, which culminated into a book later on, smart parents demand competency-based, personalized learning that happens anytime, anywhere, in a student-centered setting allowing children to take ownership over their learning (Lathram, Schneider, & Vander Ark, 2015). As schools across the United States continue to focus on designing powerful learning experiences and

plan for equitable access to devices and broadband, blended learning will continue to scale allowing both schools and parents making informed decisions through independent research studies like this one to support their students. Furthermore, this research study addresses the current opportunities and challenges encountered by students and schools and provides insights into evidence-based actions students, parents, and educators can take to cultivate effective learning at home and at school.

Finally, this research also has great implications for practice for educational leadership roles. Adaptive learning technologies are an indispensable component of personalized blended learning designs. These learning models create a shift in teacher roles and requires school and district leaders to approach supporting their teachers in new ways (Domenech et al., 2016; Horn & Staker, 2014). Teachers need new skillsets to be able to use data and form dynamic student groups to provide small group instruction while monitoring online progression of other students who are working individually on their personalized learning paths. With the increasing use of online software and digital curricula, teachers get more rapid and robust data on student progress and mastery of concepts. In order to support all teachers and schools systematically in these blended learning models, district leaders must invest in learning management systems and data dashboards to create seamless procedures and data analytic tools and engines so that teachers can feel supported in their new emerging roles as learning facilitators in these highly personalized educational settings.

Limitations

This research study only included samples from grades 5th through 9th. Although ALEKS is available to grade levels 3 through 12, the results of this study can only be

generalized for the sample population and hence would be limited to 5th through 9th grade students.

Another limitation of this study is the learning software studied, which is called ALEKS. Personalized learning continues to gain traction in K-12 public education and the edtech market is growing at a rapid scale. There are so many other adaptive online software similar to ALEKS that are designed to personalize math instruction for K-12 students. The results of this study can only be generalized for ALEKS software and we cannot make generalizations and draw conclusions accordingly for other adaptive programs similar to ALEKS without specifically studying them and their impact on student achievement.

Furthermore, the findings of this study may not be used to make generalizations about the overall effectiveness and merit of blended learning models as we see different implementations of various types of blended learning in many different school communities. Horn and Staker (2014) present case studies of blended learning implementations across the nation, and while they speak highly of the promise of blended learning for 21st century schooling, they also caution educators to set the tone right for blended learning as they see many failed attempts when personalized learning is not done right.

Recommendations for Future Research

To further this research, I would like to include a different sample from elementary grade levels, specifically 3rd and 4th grades. I would also like to study other adaptive math programs and compare their effectiveness with ALEKS. Based on state assessment scores and accountability reports, the schools included in this study experienced more gains in math in grades 5th through 9th where ALEKS was used. However, they have not seen the same level of success with 3rd and 4th grade students, where a different adaptive online software

was used. This is one of the reasons why research findings on a particular instructional software should not be assumed to be relevant for other instructional software and thus we should avoid making generalizations about effectiveness of blended learning initiatives and/or adaptive learning technologies.

As a practicing educational leader, I had the chance to visit many schools around the nation and talked with teachers and principals about their experiences with personalized learning and curricular components of their programs. Their personal reflections can be summed up as follows: You may find one particular program that works really well for middle school or high school students, but the same program may not work well for elementary level students, or vice versa. Since younger learners are more attracted to animations and games, some digital curriculum developers are doing a great job leveraging this fact and creating game-based content with animated effects. The same kid-friendly approach looks and feels childish and boring for high school level students, and causes them to disengage from the program. Therefore, it is commonly observed in K-12 districts which are going blended and investing in adaptive learning programs to choose multiple softwares for different grade bands. For such reasons, the field of personalized learning and adaptive learning technologies seem very research-friendly and such research studies can benefit schools and districts tremendously and help them make informed decisions based on research.

Concluding Thoughts

The factory model of education introduced by Horace Mann was the right fit for our country at the time of industrial revolution. That was well over a century ago. The world has changed so much since then, but our education system has not. We no longer need high

school graduates produced in batches to work in manufacturing jobs at factories. It used to be enough for graduates to have basic literacy and arithmetic skills back in the industry age. But today the job markets and the economy is much different. The automotive industry for example has manufactured a self-driving car. Today many automobile manufacturers do not employ mid-skills workers to manufacture parts and put together a car. These things are being done by sophisticated robots. So instead of mid-skills workers, manufacturers need highly trained engineers, technicians, scientists, and software developers to make and program those robots who replaced human workers. As Vander Ark (2016) puts it, “smart machines will eat jobs!” The employment opportunities for today’s children are changing and being shaped by the artificial and augmented intelligences (Vander Ark, 2016). History is full of examples with how advancements in technology and automation changed job markets and caused some jobs to vanish. These examples include the printing press, steam engines, and the rise of robotics.

Adaptive technologies and artificial intelligence are increasing efficiency and productivity at the workplace and making life so much easier for humans. Why not benefit from these advancements in education, so that teaching and learning can be elevated and optimized? Blended learning seems to be a great opportunity to provide the means for such innovation. Bringing and sustaining change in public education is very difficult and complex process. Teachers, leaders, and policy makers can naturally resist to change and prefer maintaining the status quo. While some blended learning models appear to be disrupting the current traditional school system, educators can implement a hybrid blended learning model as a sustaining innovation relative to the traditional classroom. According to Christensen, Horn, and Staker (2013), “this hybrid form is an attempt to deliver the best of both worlds—

that is, the advantages of online learning combined with all the benefits of the traditional classroom” (p. 3). Blending effective traditional strategies with new innovations in online adaptive learning can be a great fit for most teachers and school leaders especially if they have limited budgetary or architectural control over their schools (Christensen, Horn, & Staker, 2013).

Education technology is evolving and new blended curricular solutions are hitting the market every year. It is imperative that school and district leaders make informed decisions when it comes to choosing the right learning software for their blended learning programs. Soliciting teacher and student feedback, creating pilot opportunities, reviewing literature and research findings, and communicating with other innovative schools and districts are best practices for informed decision making. ALEKS is an online adaptive learning software in mathematics with a solid theoretical framework known as knowledge/space theory and is built on artificial intelligence. It has promising implications for student learning outcomes and strong results as presented in this research study. The researcher hopes that this study makes significant contributions to the literature in this emerging field and provides a benchmark for evidence-based decision making for educators.

Finally, for schools and districts who cannot afford hardware and software to bring adaptive learning technologies to their students, I would recommend exploring cost-free and/or low cost resources to start their personalized learning journey. Khan Academy is a free resource that many schools leverage as a personalized tutoring system. Facebook recently partnered with a charter school district in San Francisco bay area to design a personalized learning platform and dashboard that makes it possible to house online curriculum content and track student progress against personalized goals and timelines. This

personalized learning platform is now offered to all schools and districts across the United States free of charge with pre-loaded digital curricula developed under creative commons licensing.

REFERENCES

- Adams, C.J. (2015). SAT-ACT performance for 2015 graduates called 'disappointing'.
Education Week, September 9, 2015.
- Agondi, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2009).
Achievement effects of four early elementary school math curricula: Findings from
first graders in 39 schools (NCEE 2009-4052). Washington, DC: National Center for
Education Evaluation and Regional Assistance, USDE.
- Albert, D., & Hockemeyer, C. (1997). Adaptive and dynamic hypertext tutoring systems
based on knowledge space theory. *Frontiers in Artificial Intelligence and
Applications*, 39, 553-555, Amsterdam: IOS Press.
- Albert, D., Hockemeyer, C., & Held, T. (1997). A relational model for hypertext structures.
IEEE Transactions on Systems, Man, and Cybernetics. 13(3), 322-324.
- ALEKS -- Assessment And Learning, K-12, Higher Education, Automated Tutor, Math.
(2017). *Aleks.com*. Retrieved March 24, 2017, from <https://www.aleks.com/>.
- Aleven, V., Koedinger, K.R. (2001). Investigations into help seeking and learning with a
cognitive tutor. *Papers of the AIED-2001 Workshop on Help Provision and Help
Seeking in Interactive Learning Environments*, 47-58.
- Anderson, J., Boyle, F., Farrell, R., & Reiser, B. (1987). *Modelling cognition*. Chichester,
UK: John Wiley & Sons.
- Anderson, J., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons
learned. *The Journal of the Learning Sciences*, 4(2), 167-207.
- Anderson, J. (1992). Automaticity and the ACT theory. *The American Journal of
Psychology*, 105(2), 165-180.

- Anderson, J. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), 355-365.
- Anderson, R. (1970). An experiment on behavioral learning in a consumer credit game. *Simulation and Games: An International Journal of Theory, Design, and Research*, No. 1 (March, 1970), 43-54.
- Badaracco, M., & Martinez, L. (2011). An intelligent tutoring system architecture for competency-based learning. *KES 2011, Part II, LNAI 6882*, 124–133, Berlin Heidelberg: Springer-Verlag.
- Baker, E. (1968). A pre-civil war simulation for teaching American history. *Simulation Games and Learning*. Beverly Hills: Sage Publications, Inc.
- Baker, R.S., Corbett, A.T., Koedinger, K.R., & Wagner, A.Z. (2004). Off-task behavior in the cognitive tutor classroom: When students “game the system”. *Proceedings of ACM CHI 2004: Computer-Human Interaction*, 383-390.
- Baker R.S., Corbett A.T., & Koedinger K.R. (2004b). Detecting student misuse of intelligent tutoring systems. ITS 2004. *Lecture Notes in Computer Science*, 3220, 531-540, Berlin Heidelberg: Springer-Verlag.
- Barrus, A., Sabo, K., Joseph, S., Atkinson, R., & Perez, R. (2012). Evaluating adaptive, computer-based mathematics tutoring systems: A math improvement and feasibility study. Arizona State University: ANGLE.
- Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., and Chan, D. (2015). School composition and the black–white achievement gap (NCES 2015-018). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved March 27, 2017, from <http://nces.ed.gov/pubsearch/>.

- Camera, L. (2016, January 13). Achievement gap between white and black students still gaping. *U.S. News & World Report*. Retrieved from <https://www.usnews.com/news/blogs/data-mine/2016/01/13/achievement-gap-between-white-and-black-students-still-gaping/>.
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). Effectiveness of reading and mathematics software products: Findings from two student cohorts (NCEE 2009-4041). Washington, DC: NCES & USDE.
- Carpenter, J., & Hanna, R. E. (2006). *Predicting student preparedness in calculus*. Paper presented at 2006 Annual Conference & Exposition, Chicago, Illinois. Retrieved from <https://peer.asee.org/1428/>.
- Christensen, C., Horn, M., & Staker, H. (2013). Is k-12 blended learning disruptive? An introduction to the theory of hybrids. Clayton Christensen Institute, Inc. May 2013. Retrieved from <http://www.christenseninstitute.org/publications/hybrids/>.
- Coburn, C. E.; Hill, H. C.; Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The common core state standards and implementation research. *Educational Researcher*, 45(4), 243-251.
- Conlan, O., O'Keeffe, I., Hampson, C., & Heller, J. (2006). Using knowledge space theory to support learner modeling and personalization. In T. Reeves, & S. Yamashita (Ed.), *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education 2006*, 1912-1919. Chesapeake: AACE.
- Desilver, D. (2017). U.S. students' academic achievement still lags that of their peers in many other countries. Pew Research. Retrieved from

- <http://www.pewresearch.org/fact-tank/2017/02/15/u-s-students-internationally-math-science/>.
- Dickard, N. (2003). *The sustainability challenge: Taking edtech to the next level*. Washington, DC: Benton Foundation.
- Dimitrov, D.M. & Rumrill, Jr., P.D. (2003). Pretest-posttest designs and measurement of change. *Speaking of Research*, 159-165, IOS Press.
- Doignon, J.-P. (1994). Knowledge spaces and skill assignments. *Contributions to mathematical psychology, psychometrics, and methodology*, 111-121. New York, NY: Springer-Verlag.
- Doignon, J.-P. & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin Heidelberg: Springer-Verlag.
- Domenech, D., Sherman, M., & Brown, J.L. (2016). *Personalizing 21st century education: A framework for student success*. San Francisco, CA: Jossey-Bass, A Wiley Brand.
- DreamBox Learning - Online Math Learning. (2017). *DreamBox Learning*. Retrieved March 27, 2017, from <http://www.dreambox.com/>.
- Every Student Succeeds Act (ESSA) | U.S. Department of Education. (2017). *Ed.gov*. Retrieved 27 March 2017, from <https://www.ed.gov/essa?src=rn/>.
- Falmagne, J.-C. & Doignon, J.-P. (2011). *Learning spaces*. Berlin Heidelberg: Springer-Verlag.
- Falmagne, J., Cosyn, E., Doignon, J., & Thiery, N. (2006). The assessment of knowledge, in theory and in practice. *Lecture notes in computer science*, (3874), 61-79.
- Falmagne, J., Cosyn, E., Doble, C., Thiery, N., & Uzun, H. (2007). Assessing mathematical knowledge in a learning space: Validity and/or reliability. *Annual Meeting of the*

- American Educational Research Association (AERA)*, 2007. Retrieved from http://www.stat.cmu.edu/~brian/NCME07/Validity_in_L_Spaces.pdf/.
- Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P. & Johannesen, L. (1990). Introduction to knowledge spaces: How to build test and search them. *Psychological Review*, 97, 201-224.
- Falmagne, J. & Doignon, J. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175-196.
- Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal*, 91(1), 29-42.
- Foster, M. E., Anthony, J. L., Clements, D. H., Sarama, J., & Williams, J. M.(2016). A randomized control trial evaluating the effectiveness of computer assisted instruction in numeracy on math outcomes for monolingual English speaking kindergartners from title 1 schools. *Society for Research on Educational Effectiveness*. Retrieved from <http://files.eric.ed.gov/fulltext/ED566981.pdf/>.
- Frankenberg, E., Lee, C., and Orfield, G. (2003). A multiracial society with segregated schools: Are we losing the dream? Cambridge, MA: The Civil Rights Project Harvard University. Retrieved March 27, 2017 from <http://pages.pomona.edu/~vis04747/h21/readings/ arewelosingthedream.pdf/>.
- Gano, L. R. (2011). Fitting technology to the mathematics pedagogy: Its effect on students' academic achievement. *Journal of College Teaching & Learning*, 8(11), 29-36.
- Garcia-Santillan, A., Escalera-Chavez, M., Lopez-Morales, J., & Rangel, A. (2014). Students, computers and mathematics: The golden trilogy in the teaching-learning

- process. *TOJET: The Turkish Online Journal of Educational Technology*, 13(3), 138-147.
- Gillard, J., Robathan, K., & Wilson, R. (2011). Measuring the effectiveness of a mathematics support service: An e-mail survey. *Teaching Mathematics and Its Applications*, 30, 43-52.
- Gury, N. (2011). Dropping out of higher education in France: A micro-economic approach using survival analysis. *Education Economics*, 19(1), 51-64.
- Haberman, M. (1991). The pedagogy of poverty versus good teaching. *Phi Delta Kappan*, 73(4), 290-294.
- Hagerty, G., Smith, S., & Goodwin, D. (2010). Redesigning college algebra: Combining educational theory and web-based learning to improve student attitudes and performance. *Primus*, 20(5), 418-437.
- Hampikian, J., Guarino, J., Chyung, S. Y., Gardner, J., Moll, A., Pyke, P., & Schrader, C. (2007). Benefits of a tutorial mathematics program for engineering students enrolled in precalculus: A template for assessment. *American Society for Engineering Education*, 12(305), 1-11.
- Hannafin, R. D., & Foshay, W. R. (2006). Computer-based instruction's (CBI) rediscovered role in K-12: An evaluation case study of one high school's use of CBI to improve pass rates on high-stakes tests. *Education Technology Research and Development*, 56, 147-160.
- Harper, M., & Reddy, A. A. (2013). Mathematics placement at the University of Illinois. *Primus*, 23(8), 683-702. doi:10.1080/10511970.2013.801378

Hechinger Report. (2010, April 1). Math achievement: U.S. math education is broken.

Retrieved from <http://hechingerreport.org/u-s-math-education-is-broken/>.

Hockemeyer, C., Held, T., & Albert, D. (1998). RATH – A relational adaptive tutoring

hypertext www-environment based on knowledge space theory. In C. Alvegard (ed.)

Proceedings of CALISCE '98, 4th International Conference on Computer Aided

Learning and Instruction in Science and Engineering. Goteborg, Sweden, June 15-17,

417-423.

Horn, M. B., & Staker, H. (2014). *Blended: Using disruptive innovation to improve schools*.

San Francisco, CA: Jossey-Bass, A Wiley Brand.

Hu, X., Luellen, J. K., Okwumabua, T. M., Xu, Y., & Mo, L. (2008). Observational findings

from a web-based intelligent tutoring system: Elimination of racial disparities in an

undergraduate behavioral statistics course. Paper submitted to *American Educational*

Research Association. Retrieved from

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.518.9582/>.

Hyer, J.M., & Waller, J.L. (2014). Comparison of five analytic techniques for two-group,

pre-post repeated measures designs using SAS. Georgia Regents University, 1798-

2014.

Juan, A. A., Steegmann, C., Huertas, A., Martinez, J., & Simosa, J. (2011). Teaching

mathematics online in the European area of higher education: An instructor's point of

view. *International Journal of Mathematical Education in Science and Technology*,

42(2), 141-153.

Kane, T. J., & Staiger, D. O. (2005). Using imperfect information to identify effective

teachers. Retrieved from

- <https://www.dartmouth.edu/~dstaiger/Papers/2005/Kane%20Staiger%20teacher%20quality%204%2028%2005.pdf/>.
- Kerr, J. C. (2015). US students lag peers in East Asia in math, science. Associated Press. Retrieved from <http://bigstory.ap.org/article/1322b3a25918447d8d6db482f3ae517e/us-students-lag-peers-east-asia-russia-math-science/>.
- Kezar, A. (2011). What is the best way to achieve broader reach of improved practices in higher education? *Springer Science Business Media*, 36, 235-247.
- Koedinger, K. & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, (19), 239–264.
- Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.
- Koedinger, K., Ritter, S., & Wolfson, M. (2007). Revisiting the big city: Validating the contribution of the pat algebra tutor to student learning in mathematics.
- Korbey, H. (2013). A key to understanding math. Retrieved from <http://blogs.kqed.org/mindshift/2013/10/imagery-a-key-to-understanding-math/>.
- Krimski, S. (2012). Do financial conflicts of interest bias research? An inquiry into the “funding effect” hypothesis. *Science, Technology, & Human Values*. Published Online, 20 September 2012, Sage Publications.
- Latham, B., Schneider, C., & Vander Ark, T. (2015). *Smart parents: Parenting for powerful learning*. Federal Way, WA: Getting Smart.

- Lavergne, V. (2007). The effect of the aleks web-based learning system on standardized math scores. Unpublished action research project in Shawnee Mission Schools, Shawnee, KS.
- Loredo, G. (2012). Agenda item VI-G. Retrieved from http://www.thea.state.tx.us/generalpubs/agenda/ag2012_04/VIG/VIGSR.pdf/.
- Lucas, L., Postma, C., & Jay, C. (1974). A comparative study of cognitive retention using simulation-gaming as opposed to lecture-discussion technique. *Peabody Journal of Education*, 261-266.
- Marder, M. (2012). Failure of U.S. public secondary schools in mathematics. *AASA Journal of Scholarship and Practice*, 9(1), 8-24.
- Marzano, R., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.
- Mathematica Policy Research. (2017). *Rapid-Cycle Tech Evaluations Accelerate Decisionmaking*. Retrieved March 27, 2017, from <https://www.mathematica-mpr.com/our-publications-and-findings/projects/rapid-cycle-tech-evaluation/>.
- McKenney, J., & Dill, W. (1966). Influences on learning in simulation games. *American Behavioral Scientist*, 10(2), 28-32.
- MIND Research Institute. (2017). *Mindresearch.org*. Retrieved March 24, 2017, from <http://www.mindresearch.org/>.
- Montague, M., Krawec, J., Enders, C., & Dietz, S. (in press). The effects of cognitive strategy instruction on math problem solving of middle school students of varying ability. *Journal of Educational Psychology*.

- Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). *E-desk: A review of recent evidence on the effectiveness of discrete educational software*. Menlo Park, CA: SRI International. Retrieved from https://www.sri.com/sites/default/files/publications/imports/Task3_FinalReport3.pdf/.
- No Child Left Behind - ED.gov. (2017). *Www2.ed.gov*. Retrieved 27 March 2017, from <https://www2.ed.gov/nclb/landing.jhtml/>.
- Nwaogu, E. (2012). *The effect of Aleks on students' mathematics achievement in an online learning environment and the cognitive complexity of the initial and final assessments*. Dissertation, Georgia State University. Retrieved from http://scholarworks.gsu.edu/msit_diss/94/.
- NWEA | Educational assessment that helps kids learn. (2017). *NWEA*. Retrieved March 27, 2017, from <https://www.nwea.org/>.
- Office of Educational Technology. (2017). *Educational Technology Rapid Cycle Evaluations - Office of Educational Technology*. Retrieved March 27, 2017, from <https://tech.ed.gov/rce/>.
- Orfield, G., Kuscera, J., & Siegel-Hawley, G. (2012). *E pluribus . . . Separation: Deepening double segregation for more students*. Cambridge, MA: The Civil Rights Project Harvard University. Retrieved March 27, 2017, from http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/mlk-national/e-pluribus...separation-deepening-double-segregation-for-more-students/orfield_epluribus_revised_omplete_2012.pdf/.
- Padilla-Oviedo, A., Mundy, M.-A., & Kupczynski, L. (2016). The effects of instructional strategies, college division, and gender on students' performance in college algebra at

- a university in south Texas. *Journal of Advance Research in Mathematics and Mathematical Sciences*, 1(8), 22-34.
- Picciano, A. (2006). Blended learning: Implications for growth and access. *Online Learning Consortium*, 10(3), 95-102.
- Pigott, T. D., Polanin, J. R., Valentine, J. C., Willians, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, 42(8), 424-432.
- Pyke, P., Gardner, J., Hampikian, J., Belcheir, M., & Schrader, C. B. (2007). An innovative method to realistically track engineering student retention and academic progress. *American Society for Engineering Education*, 12(215), 1-17.
- Ravenel, J., Lambeth, D. T., & Spires, B. (2014). Effects of computer-based programs on mathematical achievement scores for fourth-grade students. *Journal on School Educational Technology*, 10(1), 8-21.
- Reasoning Mind – Solving for Every Variable. (2017). *Reasoning Mind*. Retrieved March 27, 2017, from <http://www.reasoningmind.com/>.
- Ringstaff, C. & Kelley, L. (2002). *The learning return on our educational technology investment: A review of findings from research*. San Francisco, CA: WestEd. Retrieved from https://www.wested.org/online_pubs/learning_return.pdf/
- Salerno, C. A. (1995). The effect of time on computer-assisted instruction for at-risk students. *Journal of Research on Computing in Education*, 28(1), 85.
- Sarkis, H. (2004). Cognitive tutor algebra 1 program evaluation: Miami-Dade county public schools. Lighthouse Point, FL: The Reliability Group. Retrieved from [www.carnegielearning.com/web_docs/sarkis_2004 .pdf/](http://www.carnegielearning.com/web_docs/sarkis_2004.pdf/).

- Saxe, K., & Braddy, L. (2015). A common vision for undergraduate mathematical sciences programs in 2025. The Mathematical Association of America, *Notices of the AMS*, June/July 2016. Retrieved from <http://www.maa.org/sites/default/files/pdf/CommonVisionFinal.pdf/>.
- Schmoker, M. (2011). *Focus: Elevating the essentials to radically improve student learning*. Alexandria, VA: ASCD.
- Schofield, J.W. (1995). *Computers and classroom culture*. Cambridge, UK: Cambridge University Press.
- Schwartz, K. (2013). *In teaching algebra, the not-so-secret way to students' hearts*. Retrieved from <http://blogs.kqed.org/mindshift/2013/12/to-learn-algebra-the-not-so-secret-way-to-students-hearts/>.
- Shakerdge, K. (2016, May 08). High failure rates spur universities to overhaul math class. Retrieved March 26, 2017, from <http://hechingerreport.org/high-failure-rates-spur-universities-overhaul-math-class/>.
- Slavin, R., & Lake, C. (2008). Effective programs in mathematics: A best-evidence synthesis. *Review of Educational Research*, 78(3), 427-515.
- Staker, H., & Horn, M. (2012). Classifying k-12 blended learning. Innosight Institute, Inc. Retrieved from <http://www.christenseninstitute.org/?publications=classifying-k-12-blended-learning-2/>.
- Sturgis, C. (2017). Competency education: Competency works. *Competencyworks.org*. Retrieved March 26, 2017, from <http://www.competencyworks.org/about/competency-education/>.

- Taagepera, M., & Noori, S. (2000). Mapping students' thinking patterns in learning organic chemistry by the use of knowledge space theory. *Journal of Chemistry Education*, 77(9), 1224-1229.
- Taylor, J. (2008). The effects of a computerized algebra program on mathematics achievement of college and university freshmen enrolled in a developmental mathematics course. *Journal of College Reading and Learning*, 39(1), Fall 2008.
- Tienken, C., & Wilson, M. (2007). The impact of computer assisted instruction on seventh grade students' mathematics achievement. *Planning and Changing*, (38), 181–190.
- Tuning Project. (2017). *Tuning.unideusto.org*. Retrieved 26 March 2017, from <http://tuning.unideusto.org/tuningal/>.
- U.S. Department of Education. (2012). Definitions. Retrieved from <https://www.ed.gov/racetop/district-competition/definitions/>.
- U.S. Department of Education. (2017). *Www2.ed.gov*. Retrieved 27 March 2017, from <https://www2.ed.gov/programs/racetothetop/>.
- Vander Ark, T. (2013). *The future of learning: Personalized, adaptive, and competency-based*. Bellevue, WA: DreamBox Learning.
- Vander Ark, T. (2012). *Getting smart: How digital learning is changing the world*. San Francisco, CA: Jossey-Bass, A Wiley Brand.
- Vander Ark, T. (2016). Smart machines will eat jobs—except where smart people create them. *Education Week*. October 31, 2016. Retrieved from http://blogs.edweek.org/edweek/on_innovation/2016/10/smart_machines_will_eat_jobs_except_where_smart_people_create_them.html/.

- Varsavsky, C. (2010). Chances of success in and engagement with mathematics for students who enter university with a weak mathematics background. *International Journal of Mathematics in Science and Technology*, 41(8), 1037-1049.
- Villano, M., & Bloom, C. (1992). *Probabilistic student modeling with knowledge space theory*. Brooks Air Force Base: Armstrong Laboratory.
- Vincent, Eric T. (2005). *Foundational skills are key to success in the workplace*. ACT Inc. Retrieved from <http://greatplains.edu/wp-content/uploads/2009/09/FoundationalSkills.pdf/>.
- Watt, M. (2015). *States' implementation of the common core state standards and the australian curriculum: A comparison of the change process in two countries*. Online Submission. ERIC. Retrieved from <http://files.eric.ed.gov/fulltext/ED557593.pdf/>.
- Waxman, H., Lin, M., & Michko, G. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Naperville, IL: Learning Point Associates.
- Wolf, M.A., & Schneiderman, M. (2014). Technology-enabled personalized learning: Findings & recommendations to accelerate implementation. Pre-conference paper presented at the Technology Enabled Personalized Learning Summit, Raleigh, NC. February 2014.
- Woolf, B. (2009). *Building intelligent interactive tutors*. New York, NY: Morgan Kaufman.
- Xu, Y, Meyer, K., & Morgan, D (2009). A mixed-methods assessment of using an online commercial tutoring system to teach introductory statistics. *Journal of Statistics Education*, 17(2). Retrieved from www.amstat.org/publications/jse/v17n2/xu.html/.

VITA

Burak Yilmaz was born on May 11, 1982 in Sinop, Turkey. He graduated from a private high school in Istanbul and earned a full ride scholarship to Bilkent University in Ankara, Turkey. Upon completing his undergraduate studies in mathematics at Bilkent University, Burak moved to the United States to pursue graduate studies. This move was motivated by his desire to become an effective educator and tackle social injustice and inequities in underserved communities. He started teaching middle and high school math at an urban public charter school in Oklahoma City, OK. While working as a teacher, Burak received his master's degree in general education from the University of Central Oklahoma.

Burak taught 4 years in Oklahoma City and became an assistant principal in his last year at the secondary school. Later on, he moved to Kansas City, MO to help start a new elementary charter public school, Frontier School of Innovation, authorized by the University of Missouri – Kansas City. After working one year as the school's first vice principal, Burak was promoted to a principal position at the feeding high school within the same charter school system. He led the school successfully for two years and moved to Houston, TX for another high school principal position at Harmony Public Schools, second largest charter school network in the nation.

In December 2012, Harmony Public Schools was awarded a \$30-million federal Race to the Top – District (RTT-D) grant by the Obama administration and Burak was chosen to lead this four-year grant to design, implement, and scale personalized learning initiatives at Harmony's secondary schools. As RTT-D project director, Burak supported the district's large scale rollout of blended learning and project based learning initiatives, as well as designing actionable data dashboards. Towards the end of the RTT-D grant period, Harmony

Public Schools pursued other large federal grants and Burak was part of a district leadership team that put together another successful federal grant application, which won Harmony a \$27-million Teacher Incentive Fund (TIF) grant in September 2016. Due to his experience and successful leadership with Harmony's Race to the Top program, Burak continued to serve as the Project Director for the new TIF grant. His current work focuses on designing innovative professional development solutions, aligning competency systems and career pathways, and designing equitable performance based compensation systems for teachers and school leaders.