

MISSION ACCEPTED: A CASE STUDY EXAMINING THE RELATIONSHIP OF
KHAN ACADEMY WITH STUDENT LEARNING

by

GEOFFREY BARRETT

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership,
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements for the degree of
Doctor of Education

March 2018

DISSERTATION APPROVAL PAGE

Student: Geoffrey Barrett

Title: Mission Accepted: A Case Study Examining the Relationship of Khan Academy with Student Learning

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Methodology, Policy, and Leadership by:

Kathleen Scalise	Chairperson
Michael D. Bullis	Core Member
Keith Hollenbeck	Core Member
Joanna Goode	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School
----------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2018

© 2018 Geoffrey Barrett

DISSERTATION ABSTRACT

Geoffrey Barrett

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

March 2018

Title: Mission Accepted: A Case Study Examining the Relationship of Khan Academy with Student Learning

This study examined implementing the online website Khan Academy as a primary resource for mathematics instruction. Participants were high school students aged 15-18 years enrolled in the traditional mathematics courses of Algebra 1, Geometry, and Algebra 2. A pre-test/post-test research design was implemented over the course of a six-week period of instruction. I wanted to examine whether Khan Academy was associated with positive learning outcomes over the six-week period as compared to measures of normalized growth.

Additionally, I asked whether a beta program to personalize instruction on Khan Academy was associated with statistically significantly better outcomes compared to the regular Khan Academy course sequences alone. To address my questions, I randomly assigned students into treatment and comparison groups. As a measure of learning growth, I used the Northwest Education Assessment's Measures of Academic Progress (MAP) to establish a pre-treatment baseline and again at the end of the program to measure learning growth. I compared before and after means. Overall, I found that students in both groups showed overall positive growth, statistically significantly different from normal expected growth. However, I did not find a statistically significant difference between the two groups.

In terms of practical implementation, the results of this study suggest that use of Khan Academy as a primary instructional resource is associated with positive learning outcomes in this data set. Further study with larger sample sizes to confirm these preliminary results is recommended.

CURRICULUM VITAE

NAME OF AUTHOR: Geoffrey Barrett

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of New Mexico, Albuquerque
University of Iowa, Iowa City

DEGREES AWARDED:

Doctor of Education, 2018, University of Oregon
M.A., Special Education, 2002, University of New Mexico
B.A., History (Biology minor), 1989, University of Iowa

AREAS OF SPECIAL INTEREST:

Educational Technology
Education of Homeless Students

PROFESSIONAL EXPERIENCE:

Teacher, West Lane Technology Learning Center, 10 years
Teacher, Robert F. Kenney Charter High School, 5 years

ACKNOWLEDGMENTS

I wish to express sincere appreciation to my committee chair, Dr. Kathleen Scalise, whose guidance in the preparation of this manuscript was essential. Also, special thanks are due to Dr. Michael D. Bullis who served for a time as my committee chair and offered vital support and mentorship. In addition, I wish to thank Dr. Keith Hollenbeck and Dr. Joanna Goode for their insights into this project. West Lane Technical Learning Center provided the space and internet resources required to complete this study with support from director, Ron Osibov. Finally, I would not have been able to finish this project without the support of my wife, Lois, and daughters, Faolan and Siobhan.

Dedicated to my wife, Lois Pribble, who inspires me to be better.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION AND LITERATURE REVIEW	1
Background	1
Statement of Problem.....	2
Definitions.....	2
Blended Learning.....	2
Online Learning	2
Mastery-based Learning.....	3
Student-centered Learning	3
Student-directed Learning.....	3
Growth of Online Learning.....	3
Literature Review.....	3
Effectiveness of Online Learning	4
The Pedagogy of Khan Academy	5
Khan Academy Implementation Strategies	8
Effectiveness of Khan Academy.....	10
Research Questions	15
II. METHODS.....	16
Population Sample	18
Random Assignment.....	19
Setting.....	21
Khan Academy Implementation Strategy.....	22

Chapter	Page
Measuring Growth	23
Data Collection.....	24
Data Analysis	25
III. RESULTS AND DISCUSSION	26
Effect of MAP Recommended Practice Pilot	27
Examining Aspects of Attrition	28
Overall Proficiency Growth.....	29
Comparison to Expected Growth Norms.....	29
Comparison to Expected Growth Norms.....	30
Discussion.....	34
Limitations	36
IV. DISCUSSION.....	39
Limitations	41
Implications for Practice.....	43
Conclusions.....	43
APPENDIX	49
Implementation Guide	49
REFERENCES CITED.....	52

LIST OF FIGURES

Figure	Page
1. Example of MAP Recommended Pathway.....	18
2. Initial Frequency of Participant Percentile Ranks	19
3. Khan Academy Learning Pathway	23
4. Percent of participants exceeding normal expected growth	31
5. Percentile Rank Range.....	32
6. Reported satisfaction with learning progress.....	36
7. Likelihood of watching videos.....	37
8. Frequency of seeking instructor assistance.....	37
9. Frequency of note-taking	38
10. Khan Academy Learning Pathway	51

LIST OF TABLES

Table	Page
1. Pre-test/post-test research design	17
2. Comparison vs. treatment, original pairs..	26
3. Comparison vs. treatment, rematched pairs	29
4. Overall pre- and post-test means, t-test results	29
5. Comparison of RIT growth statistical significance (p-value) by quartile	30
6. Observed RIT growth vs. expected growth	33
7. Comparison of RIT growth statistical significance (p-value) by quartile.....	34
8. Relationship between RIT score change and initial score	35

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

The purpose of this study is to examine the relationship of Khan Academy with learning outcomes for high school math students. Although Khan Academy is used world-wide and its influence in educational settings is growing, there have been few examinations of its effectiveness in terms of student learning outcomes. The goal of this study is to make a small contribution to that gap in the literature. To examine the question of Khan Academy's effects on student learning outcomes, a population of high school aged students engaged in a six-week treatment consisting of using the platform as a primary resource for acquiring math skills.

Khan Academy is a non-profit educational organization that offers online tutoring in a variety of subject areas. It is most known for its catalog of mathematics instructional videos. Since its beginnings in around 2011, it has evolved from a YouTube video catalog of discrete mathematical skills into a full-fledged, mastery-based, student-centered tutorial program. Pertaining to secondary education, Khan Academy offers full content courses, referred to as *missions*, in traditional learning pathways such as Algebra 1, Geometry, Algebra 2, and Pre-calculus, as well as in an integrated math approach, Mathematics 1, Mathematics 2, and Mathematics 3. Recently, Khan Academy introduced the MAP Recommended Practice feature which allows teachers, referred to as *coaches*, to personalize individual learning pathways based on student MAP scores.

Presently, Khan Academy is used by millions of learners globally and is often incorporated into primary and secondary school math programs. Despite its widespread use and influence on math instruction around the world, few scholars have focused

exclusively on whether the use of the platform has a positive effect on students' learning outcomes.

Statement of Problem

Educators and researchers have raised concerns about both Khan Academy and distance learning, despite the tremendous growth of both. Some educators criticized the learning platform for lacking a pedagogical foundation and instructional sophistication (Kai, 2012; Strauss, 2012) . Another critique of Khan Academy, is that teachers often do not implement the program in a way that was originally envisioned by the founder Salman Khan (Cargile & Harkness, 2014). Over time, Khan Academy has revamped much of its exercise content and modes of delivery, although many of the videos remain unchanged. This study examined the outcomes of Khan Academy as an instructional tool. The purpose of this study is not to address all the criticisms of Khan Academy, but focuses on student learning outcomes. The bottom line here is: do students gain proficiency through use of the platform as an instructional tool.

Definitions

To avoid confusion, some commonly used terms will be used according to the following definitions.

Blended Learning. Blended learning refers to a classroom situation in which some part of the instruction occurs in a face-to-face setting while some part is computer and internet-based conducted remotely from a brick and mortar location.

Online Learning. Online learning occurs entirely through the internet with no *physical* face-to-face interaction between teacher and student. It should be noted that with technology available at this date, the line between *blended* and *online* learning can

be blurred. Presently available applications support remote face-to-face interactions through video conferencing, voice over the internet protocol (VOIP), and instant messaging. For the purposes of this study, online learning will refer to instruction that is conducted entirely over the internet and not in a physical face-to-face setting.

Mastery-based Learning. Mastery-based learning focuses on each individual student's learning growth. Rather than proceed along a course of study at a uniform pace with all other students, as in a traditional classroom, regardless of whether learning growth has occurred, mastery-based learning requires students to achieve a threshold of proficiency before moving on to the next task or concept (Kulik, Kulik, Bangert, & Bangert-Drowns, 1990). This study will consider mastery-based programs to be instruction that requires a student to demonstrate a specific level of proficiency in order to progress to the next level.

Student-centered Learning. Student-centered learning refers to instruction that is specifically and deliberately designed to focus on the needs of the student, including consideration of past history of success or lack of success in learning, socio-emotional issues impacting performance, and present levels of academic proficiency.

Student-directed Learning. A student-directed approach allows students multiple options in order to reach their learning goals. These options can include how much time to spend on a particular learning task, the sequence of task focus, and types of learning tasks to access.

Literature Review and Conceptual Framework

I conducted a web search for peer-reviewed articles with the search terms *Khan Academy* and *mathematics* utilizing the University of Oregon Library search

function. The search returned 215 potentially relevant articles. After reviewing the abstracts, commentaries and articles dealing with topics unrelated to this study were eliminated. I retained articles reporting original research. There were 21 articles that met my criteria. I then classified each article into the categories Effectiveness of Online Learning, Khan Academy Implementation Strategies, Khan Academy Student Engagement, and Effectiveness of Khan Academy (2). I review each category below.

Effectiveness of Online Learning. Studies support the hypothesis that online learning is at least as effective as face-to-face instruction. Some moderating factors, such as blended learning, may increase the positive effects of online learning, but even in the absence of those factors, online instruction as a means of delivering instruction has positive support in the literature. A 2013 meta-analysis of 45 studies found that “purely online learning has been equivalent to face-to-face instruction in effectiveness (Means, Toyama, Murphy, & Baki, 2013).” That study also found that the moderating influence of blended learning resulted in higher effects. The authors cautioned against interpreting results as suggesting “that online learning is superior as a medium (Means et al., 2013, p. 36).” The authors suggest that varying different kinds of learning activities proved most effective across strategies. One example of a specific quasi-experimental study examining the effects of an online math tutoring program, ASSISTments, found that students using the program statistically significantly outperformed students who were in the comparison school. The effect was greater for students identified as requiring special assistance (Koedinger, Mclaughlin, & Heffernan, 2010).

A key factor to be considered here is the context of this investigation, situated in the growth of online or distance learning for education. The growth of Khan Academy is

embedded in a broader increase in popularity of distance learning. Before the advent of the internet, distance learning was most commonly practiced by correspondence. Today it is more often the case that instruction is delivered via the internet. Many educational companies offer completely online courses. Programs like Odysseyware, Connections Academy, and K12 offer full online curricula for kindergarten through high school. Other course management systems, many of them free, such as Google Classroom, Schoology, and Edmodo, provide platforms for individual teachers to build their own online course content.

Due to the availability of online learning opportunities, high school enrollment in online courses has steadily increased in recent years. According to the U.S. Department of Education (DOE), 1.3 million high school students were enrolled in distance learning classes in 2009-10. Distance learning has also grown in higher education, with an estimated 5.8 million students enrolled in at least one online course and 2.8 million enrolled in exclusively distance education courses (Allen, Seaman, Poulin, & Straut, 2016). Given this growth, it is important to examine the effectiveness of online resources available to teachers and students on the internet.

Effectiveness of Summer Math Programs. One measure of the effectiveness of using Khan Academy in a summer math program is to compare the results to other summer math programs. An obstacle to making a definitive comparison is the lack of research comparing student achievement against normal growth expectations in summer programs. Examples of research that measure learning growth with a pre-/post-test design do exist. One such program conducted by researchers at Indiana University examined the effects of a two-week summer program found that students did experience positive learning

gains (Timme et al., 2013). In that study, students received instruction in physics, AP Physics, pre-calculus, and AP Calculus. The program did not address entire course content but focused on prerequisites necessary for success in each regular high school courses. Researchers found positive results in this study.

It is important to note some differences in the design of that program compared to the design of the program in this study. The Indiana University program did not attempt to deliver a semester long course in a summer program. Also, the courses addressed by the study were higher level courses for high school students, while this study will examine effects of a summer program on growth in required high school math content rather than more advanced content, including Algebra 1, Geometry, and Algebra 2. Finally, and perhaps most important for comparison purposes, the pre-/post-test design of the Indiana study tested students using identical tests containing content directly addressed in the program. The current study compared student growth in the program to normal growth expectations with a widely used, standardized, measurement instrument, the NWEA Measure of Academic Progress. Course content was not specifically tailored to address the standardized test, but consisted of the entire course content for a typical high school mathematics course.

Conceptual Framework: The Pedagogy of Khan Academy. The conceptual framework of Khan Academy instruction is mastery learning, which is defined above as requiring students to achieve a threshold of proficiency before moving on to the next task or concept (Kulik et al., 1990).

The efficacy of mastery-based programs is well-established in the literature. A 1990 meta-analysis of 108 controlled studies found that mastery-based programs had a

positive effect on student assessment performance for upper-elementary through college-level age groups (Kulik et al., 1990). The study did find, however, that students in mastery-based programs often take more time to complete the course of study. However, the meta-analysis found an average growth of 0.5 standard deviations on final examination scores. Further, low-aptitude students benefitted more from mastery-based programs than high-aptitude, although both groups benefitted. Finally, students in mastery-based programs tended to be more satisfied with the learning experience than students in more traditional settings.

Khan Academy is a mastery-based, student-directed learning resource. Instructors have the ability to set pacing recommendations and assign playlists, but within the learning platform students are able to decide for themselves what resources to use and when to use them to achieve mastery. For example, students are able to access videos which can be watched multiple times or opt to not use videos at all. They are also able to consult example solutions, called hints, to determine for themselves what mistakes they are making. Immediate feedback informs the student if they were successful or not. If unsuccessful, the student can access the entire solution and thus learn from his or her mistake. Students demonstrate mastery of a skill by providing the correct answer five times in a row. After that, the specific skill is added to the student's mastery challenges. From that point, each correct response on a mastery challenge raises the student's mastery status by one level. After the student demonstrates consistent competency, the skill is upgraded mastered, indicating the student has acquired this skill and knowledge. Incorrect responses on the mastery challenges result in that skill being returned to a "Needs Practice" classification and the process begins again.

Khan Academy Implementation Strategies. Khan Academy is a web-based program that is available free of charge to anyone. It has two main components that are regularly accessed by learners. First, it supports a catalogue of videos that contain explanations of math concepts as well procedural algorithms for hundreds of skills. Second, it provides problem sets that cover skills and concepts from very beginning math, single digit addition, through calculus. Recently, Khan Academy upgraded secondary course content to include the traditional pathways Algebra 1 and 2, Geometry, Pre-Calculus and Trigonometry, as well as the more internationally integrated math pathways, Mathematics 1, Mathematics 2, and Mathematics 3. The pedagogy of Khan Academy is student-directed, competency based mastery of identified skills in each course content area.

The content and framework of Khan Academy supports several different implementation strategies. Those strategies primarily include a) personalized learning tool, b) supplemental resource, c) flipped classroom method, and d) primary course resource. Flexible teachers are able to combine and evolve strategies to fit the needs of their students (Murphy, Gallagher, Krumm, Mislevy, & Hafter, 2014). I will describe each of these below.

Informal use of Khan Academy as a *personalized learning tool* is likely the most popular strategy. This type of use involves either an instructor making recommendation or individuals independently seeking assistance through the platform. In either case, students access specific videos or problem sets to assist their learning as needed. Recently, Khan Academy added a feature, still in beta form, that allows instructors to match scores obtained on the MAP to playlists generated by the program's engine. That

feature has not been widely used and this study is likely the first academic examination of it.

Use as a *supplemental resource* is likely the most common method of implementation (Murphy et al., 2014). In this strategy, a teacher encourages or schedules regular use of Khan Academy videos and problem sets as a skill building utility. Teacher-led instruction is the primary instructional strategy with Khan Academy as a secondary resource. This implementation strategy is conducive to traditional classrooms with Khan Academy use occurring outside of class time. Students are typically awarded extra credit points for participation.

In the *flipped classroom* method, instructors assign activities on Khan Academy for students to complete prior to the scheduled class meeting. The goal is to focus classroom time on guided and independent practice instead of lecture which is replaced by the outside class use of Khan Academy. The founder of Khan Academy, Salman Khan, initially embraced the flipped classroom as the most effective implementation strategy for the platform.

Implemented as a *primary course resource*, Khan Academy serves as both the primary method of concept introduction as well as providing the problem sets for practice. The Khan Academy website allows for tracking of mastery of skills through practice and intermediate assessment and re-assessment. Utilizing this method, the student learns independently and at their own pace. The center of focus is on the student and the process of learning, rather than on the teacher or style of teaching.

Effectiveness of Khan Academy. Khan Academy is recognized as an important tool available to teachers. For example, one study described the platform as enabling “powerful on-line classes (Ruipérez-Valiente, Muñoz-Merino, Leony, & Delgado Kloos, 2015).” The influence of Khan Academy extends beyond borders. A study conducted in Chile asserted that it is “beneficial for students’ math skills (Light & Pierson, 2014).” Income-strapped India turned to Khan Academy in some communities as a substitute for teachers and books (Learning & Subbarayan, 2012). The perception of Khan Academy as an effective educational tool for the teaching of mathematics is widespread.

Emerging evidence provides some qualified support for that view. For example, a 2014 statewide pilot study involving almost 6,000 students found a positive relationship between use of Khan Academy and proficiency growth. Despite the recognition of Khan Academy and almost universal support for the platform, the literature on its effectiveness remains scant. I will review in detail the largest studies to date, a state-wide pilot conducted in Idaho mentioned above and a similar study conducted in California.

A statewide pilot study conducted in Idaho, *Learning Gets Personal* (hereafter, the Idaho study) found a positive relationship between student use of Khan Academy and proficiency growth as measured by the MAP (cite: Learning Gets Personal). Students who completed a higher portion of their *mission*, defined as the assigned course of study, showed more score gains than students who completed less. Percent of mission completion also positively correlated with percentile rank improvement, demonstrating gains against the normal distribution of MAP test takers.

The Idaho Study was a large-scale pilot project that included more than 5,000 participants in grades 3-8 from 43 different schools throughout the state. The duration of

the study was one school year. Researchers administered pre- and post-assessments using the MAP as the measurement instrument. Instructor participants were early adopters of Khan Academy who agreed to a set of classroom conditions including use of Khan Academy at least one hour per week and use of the MAP to measure growth. In addition, instructors were required to attend a professional development session at the start of the 2013-14 school years and complete weekly surveys on implementation. Beyond those requirements, instructors had wide latitude in terms of implementation strategies adopted, which I will discuss below.

The Idaho Study reported generally positive results for the effectiveness of Khan Academy classroom use. Most notably percent of mission completion showed a positive relationship with learning. Students who completed 0-10% of their assigned mission achieved expected annual growth, those who completed more than 40%, achieved more than 1.5 times their expected annual growth, and those completing more than 60% achieved 1.8 times their expected growth. All groups averaged at least expected growth achievement. These results, though, positive, should be adopted with some caution.

The generalizability of the Idaho Study is limited by a number of confounding factors. First, as mentioned previously, there was little control over implementation strategies. Khan Academy supports many types of classroom use, from watching videos occasionally for extra help to primary classroom resource. The only requirement of the study was that teachers agreed to incorporate use of Khan Academy for one hour per week. Similar to the findings of Koedinger et al., (2010) the results of this study suggest that more exposure to and engagement with the platform results in a larger positive effect.

Secondly, the Idaho Study experienced a high attrition rate. While more than 10,000 students initially participated in the study, 5,304 completed it. The researchers noted several reasons for the high attrition rate: a) some students did not take both the pre- and post-assessments, b) some data was lost due to inability to link Khan Academy data to MAP scores, c) some student MAP scores were disqualified due to invalid results caused by completing the assessment too quickly, and d) some students worked in missions that were outside grades 3-8, which was the focus of the study. These attrition rates and rationale do raise some concern about integrity of the results. I will address this issue by running a smaller scale experiment with tighter controls over attrition.

As a third point, only grades 3-8 were included in the Idaho Study. At the time of that study, Khan Academy did not offer full courses in high school level classes. Since then, a full high school curriculum aligned to the Common Core State Standards has been rolled out. Examining the effectiveness of Khan Academy's course material on high school learning growth would broaden our knowledge of available resources for that age group.

The Idaho study is an important step in furthering our knowledge of the effectiveness of Khan Academy. As such, it provides a baseline for further research. There are still large gaps to fill in our knowledge of Khan Academy's effects on student learning in terms of proficiency growth. One area, in particular, is to focus on the effectiveness of particular implementation strategies. For example, it could be hypothesized that, based on the Idaho Study, the deepest implementation strategy that results in the most exposure and engagement would yield the largest effect.

The results of a study conducted by SRI Education generally support the findings of the Idaho Study. Through funding from the Bill & Melinda Gates Foundation, *Research on the Use of Khan Academy in Schools* examined the implementation and the effectiveness of Khan Academy in mathematics classrooms (Murphy et al., 2014). Like the Idaho Study, the SRI Study found a positive relationship between Khan Academy use and test scores. The SRI Study also collected data on teacher perceptions of the effectiveness of Khan Academy and found that 80% of teachers reported that Khan Academy had a positive impact on students' conceptual understanding of mathematics.

The SRI Study was conducted during school years 2011-12 and 2012-2013. Researchers included seven sites for 2011-2012 and six sites for 2012-2013, with four sites repeating both years. In the first year, 1,694 students participated, increasing to 2,246 in the second year. The study reported that most sites served students from low-income communities and that several specifically used Khan Academy to address the needs of struggling students. The sites included students from three types of schools: regular public, independent, and charter. A majority (1,260 of 1,694) of the sample population attended regular public schools in 2011-12. In the second year, 47% of the participants attended regular public schools. Student participants were in grades 6 through 8.

Similar to the Idaho Study, the SRI Study reported a variety of implementation strategies, identified as a) personalized learning tool, b) supplemental resource, c) flipped classroom model, and d) primary instructional resource. All sites were reported to use a blended learning model. It should be noted that only one school during the two year period of study adopted Khan Academy as a primary instructional resource. Reasons

given for that were lack of adequate computer access, content gaps at specific grade levels, or both. The study reported that variations in implementation occurred within schools, not just across schools, as well as even within single classrooms over time. Many teachers adjusted their implementation strategies as they gained proficiency in using Khan Academy as an instructional tool.

One site, identified as Site 2, provided an example of an implementation strategy. Site 2 adopted a competency-based instructional model that focused on self-pacing and self-directed learning. One of Site 2's goals was to develop students' self-advocacy and independence in preparation for post-secondary educational opportunities. The sample size of Site 2 was 200 student participants, 45% of whom qualified for the federal free lunch program. Students at Site 2 participated in a daily two-hour math block divided evenly between teacher-led instruction and student-directed independent learning. During the independent learning period, students followed Khan Academy playlists to access videos and problem sets for practice. Students were allowed to progress at their own pace. During "core time," teacher-led instruction focused on deepening conceptual understanding and one-on-one support.

The SRI Study, like the Idaho Study found a statistically significant positive relationship between both independent variables of minutes spent and problem sets completed and improved test scores. In the SRI Study, the California Standards Test (CST). This independent finding, using a different measurement instrument, provides support for the Idaho Study's similar findings.

Research Questions

The discussion above has revealed some areas of further study. First, there is little available data on the effects of Khan Academy on student outcomes. Two studies cited in this review did provide some support for the claims that use of Khan Academy enhances outcomes, but neither study attempted to isolate a particular implementation strategy. By isolating Khan Academy as a primary classroom resource, a fuller extent of effectiveness can be captured. Secondly, participant assignment to control and treatment groups can allow for a comparison of two approaches, using Khan Academy's available recommended course playlists versus the new beta feature that allows for greater personalization by using student MAP scores. Third, collection of learning analytics available through the teacher dashboard can shed more light on effective student use of the learning resources accessible on the Khan Academy platform. This study, thus, employed the following research questions:

RQ1: Is there a difference in effects between the control and treatment groups? (Control and treatment groups are defined in the Methods chapter.)

RQ2: Is there a difference in the overall achievement growth between either or both study groups and normalized growth expectations?

RQ3: Is participation in the program associated with positive growth for student learning, for this data set?

RQ4: Is there a relationship between beginning proficiency level and growth in proficiency at the end of the study period?

CHAPTER II

METHODS

This study examined the outcomes of Khan Academy as a primary instructional tool for high school mathematics students participating in a hybrid summer program. Using the NWEA Measure of Academic Progress (MAP), I examined the differences in outcomes between the treatment and a control group assigned to the normal condition, and compared the outcomes to normed growth expectations. Using statistical analyses, I examined the growth of students using Khan Academy, including differences between the comparison and treatment groups. I also conducted a correlational analysis to determine the effect of starting proficiency on the amount of growth achieved during the six-week study period.

Study Design

This study employed a pre-test/post-test design to assess the effects of using Khan Academy as a learning resource. Participants were administered a pre-test using the NWEA MAP assessment before engaging with the treatment, the use of Khan Academy as a learning resource. After six weeks, participants were administered the MAP as a post-test in order to examine the effects of the treatment. Additionally, students were randomly assigned to one of two groups, a comparison group and a treatment group in order to assess effects of a pilot program designed to personalize student course material based on MAP results (see Table 1).

Table 1

Pre-Test/Post-test Study Design

	Pre-Test	Khan Academy	Khan Academy w/MAP recommendations	Post-Test
Comparison	X	X		X
Treatment	X		X	X

Participants assigned to the control group were assigned to a standard course of study available on Khan Academy. Khan Academy offers the traditional courses in Algebra 1, Algebra 2, Geometry, Pre-Calculus, and Trigonometry. In addition to traditional courses, the program also offers an integrated math approach with courses Mathematics I, Mathematics II, and Mathematics III. For the purposes of this study, each student was assigned a course based on the recommendation of their high school counselor. As it turned out, all participants in the control group were assigned to traditional sequences of Algebra 1, Geometry, and Algebra 2.

The treatment group was assigned a course of study recommended by their counselors but adjusted according to Khan Academy's MAP recommended practice tool. That tool is a beta feature on Khan Academy and is based on each individual MAP score. In some cases, the course was adjusted by the researcher to account for sequencing and to assure that the student received instruction meeting the required course content per state standards. In those cases, elements of the recommended course were combined with the required content from their traditional course of study to best scaffold the learning process of the student.

An example of the difference between the comparison and treatment groups is The MAP Recommended Practice tool indicated that a participant whose course of study included sequences would benefit from the exercise: Math Patterns 2. In this specific

case, that participant would have Math Patterns 2 to their course of study which would deviate from the standard course of study (see Figure 1).

Map Recommended Practice	Standard Course of Study
Math Patterns 2	Extend Arithmetic Sequences
Extend Arithmetic Sequences	Use Arithmetic Sequence Formulas
Use Arithmetic Sequence Formulas	Recursive Formulas for Arithmetic Sequences
Recursive Formulas for Arithmetic Sequences	

Figure 1. Example of MAP Recommended Practice vs. Standard Course of Study (Sequences)

A stratified randomized matched pair design was used to assign students to study groups. For the stratification, participants were first matched into pairs based on their pre-test MAP RIT scores and their courses of study. After they were paired one member of each pair was randomly assigned to either the comparison or the treatment group. Note that after the treatment was applied, differential attrition was addressed by examining the characteristics of the completion groups. More details on the randomization approach and the attrition plan are described below in the Population Sample section.

To continue with the study, treatments were applied, data collected from students during the intervention process (which will be discussed below), and then the post-test administered and results analyzed.

Population Sample

Participants in this study were high school aged students who were referred to the summer math program by their school counselors. Students came from a combination of rural schools in Lane County, Oregon and more urban school located in the city of

Eugene, and could be expected to be higher risk students for credit denial in mathematics due to referral to summer intervention by their counselors. All home schools were public with the exception of one private school. Sixty students participated in the summer program. Each student's current level of mathematics achievement was assessed using the MAP which generated a percentile rank range from 5 to 99, indicating a broad range of initial ability with bias toward lower achievement (see Figure 2).

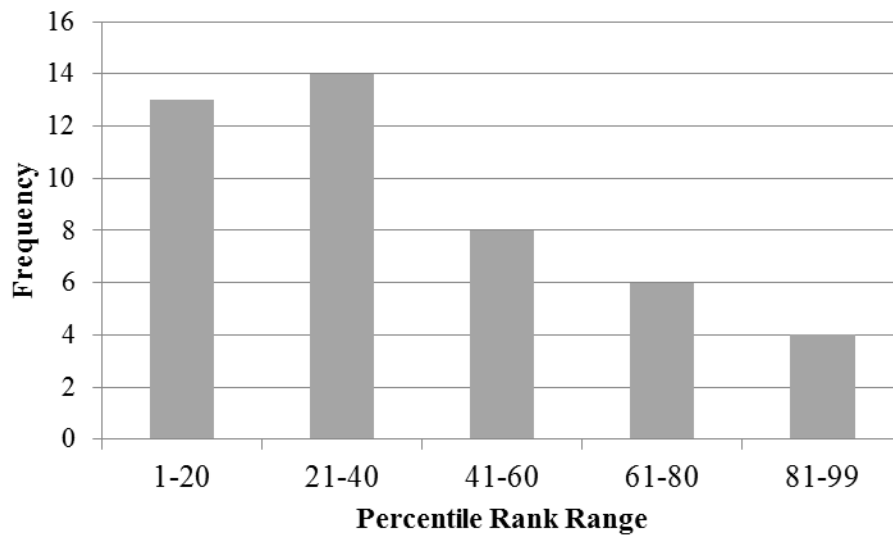


Figure 2. Frequency of Participant Percentile Ranks. Results of MAP pre-test by percentile rank.

Random Assignment. I randomly assigned each participant to the treatment group or the control group. In order to control for initial differences, I matched participants into pairs of similar pre-test MAP scores. All scores were sorted by percentile rank, then paired with a similar score in the order. Gender and course enrollment were considered in matching scores with the similarity in score given the most weight.

In order to allocate students to groups, I conducted a matched pair examination of the results obtained in this study. Participants were first matched into pairs based on their

pre-test MAP RIT scores and course assignments. Note that these pairs were used only for stratification to allocate the sample, and not as a matched-pair design in the analysis. (For analysis, the groups were considered equivalent groups at pretest, see results below.)

After students were paired up to stratify the sample, I randomly assigned one member of each pair to the treatment group. The treatment group was assigned a course of study based on the MAP Recommended Practice. The comparison group was assigned the normal course of study available on Khan Academy.

After assignments were completed, I compared the means of each group to ascertain whether there was an initial statistically significant difference of means between the two groups. The mean pre-test RIT scores for the treatment group ($\mu = 228.95$, $SD = 13.98$) and the comparison group ($\mu = 230.00$, $SD = 14.21$) were not found to be statistically significantly different ($p = 0.82$, $\alpha < 0.05$). Therefore the groups were considered equivalent groups on the variables of interest following random assignment

Ensuring Participant Privacy. All data was de-identified and no individual scores are reported in this study. Also, no disaggregated data with a sample size less than six is reported. I conducted this study in a school environment in which all the requirements of FERPA are rigorously observed.

Potential Effects on Participants. Because all students were exposed to the state-required content standards, there was little potential harm to the students, either in the comparison or the treatment group. At the outset, it was unknown whether one instructional strategy is statistically significantly superior to the other, so there is no known advantage to being assigned to one group over the other. As the instructor, it was my responsibility to ensure that all students received high quality instruction and that

consideration took precedent at all stages of this study. In keeping to that principle, however, there were no instances in which the needs of the study conflicted with my professional responsibilities in the carrying out of this study.

It is possible that knowledge of participating in a study could have a positive effect on student motivation. While that might be a factor that impacts the generalizability of the results, it is a potential risk that I feel is worth taking. In observing students who participated, they did not seem to be concerned about how their work affected the study and, indeed, most seemed to forget that they were participants in a study and were not mindful of that fact on a day-to-day basis.

Setting

The Summer Math Academy (SMA), which is the context of the study here, operates out of a small public charter high school in rural Oregon. The SMA is a hybrid program combining elements of online distance learning with face-to-face instruction. During the course of the study, group instruction did not occur, but one-on-one instructional sessions were normal, particularly with students who required extra assistance. All students were required to spend at least one 3-hour session at the school in person each week; otherwise, they worked from a distant location. Students were able to access one-on-one instruction through internet applications such as Google products and online whiteboards, as well. One-on-one instruction was delivered as needed, either by request of the student or intervention by the instructor. The duration of the program was six weeks with an open computer lab Monday through Thursday from 8:30 to 11:30. The program used Khan Academy as its primary instructional tool throughout.

Khan Academy Implementation Strategy

Khan Academy's pedagogy is mastery-based, student-centered instruction. Students learn through active engagement with the program, accessing videos and hints, to assist in the learning process. Teachers, termed *coaches* act as guides or mentors, intervening as needed. KA provides coaches with a variety of tools to monitor student progress. They can view whether students are achieving success or struggling, the time students spend overall as well as on each problem within an exercise, and which learning tools the student accesses. The coach can focus instructional time on students who are struggling while students doing well are able to proceed at their own pace. While this program has apparent success, it is important to use an outside tool to monitor student proficiency growth to corroborate anecdotal observations.

For the purposes of this study, I employed Khan Academy as the primary classroom resource in a blended hybrid summer math program. Students were assigned their coursework through Khan Academy. All students received a tutorial on the best practices for success on Khan Academy. I instructed students to follow a best practices learning strategy following these steps: a) examine given problem, b) determine whether they have the background knowledge to attempt a solution, c) attempt a solution. If the solution is correct, the student moves on to the next problem. Otherwise, they are encouraged to watch a video first, then attempt a solution. In either case, if the solution is incorrect, the student may consult the hints to determine what their mistake was. If after following the recommended learning pathway, the student still requires assistance, he will be encouraged to solicit assistance from the teacher (see Figure 3). See Appendix A for a full description of the program implementation.

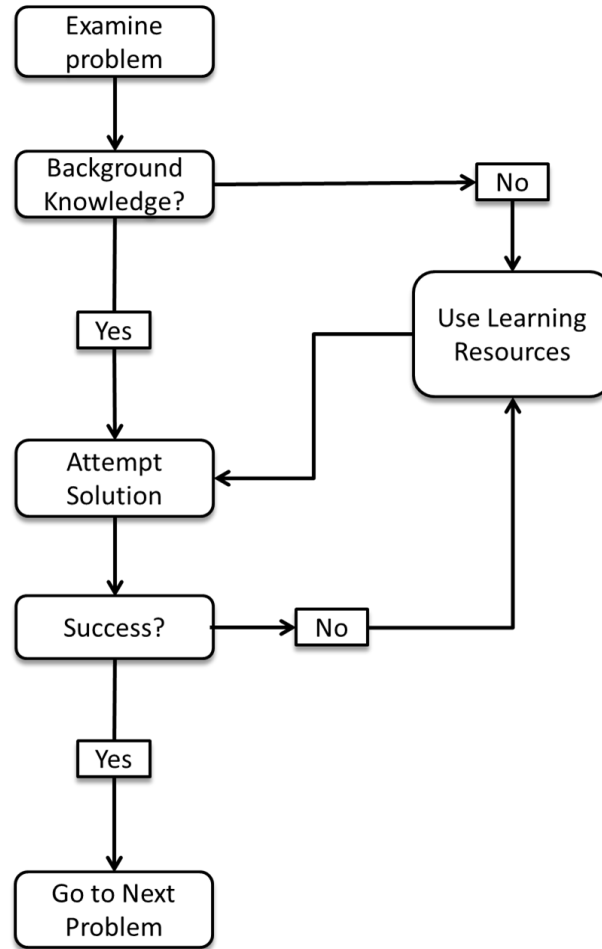


Figure 3. Khan Academy Learning Pathway.

Measuring Growth

This study employed the Northwest Evaluation Association Measures of Academic Progress as the instrument to measure growth during the six-week period of this study. The Northwest Evaluation Association (NWEA) is a “global not-for-profit educational services organization.” NWEA developed the Measure of Academic Progress as an interim assessment to measure student academic growth over time. MAP assessments are computerized adaptive tests (CATs) that report scores based on a linear Rasch Unit (RIT) scale. The MAP RIT scale provides a valid and consistent measure of

academic growth (Wang, Mccall, Jiao, & Harris, 2013). I conducted pre- and post-MAP assessments with participants in this study to measure the degree of growth during the period of the study.

To assess the appropriateness of utilizing MAP for this study, a search for the use of NWEA Measure of Academic Progress (MAP) as a measurement of student academic growth was conducted. That search found 31 articles that matched search parameters on the UO Library Search engine related to MAP as a measure of growth. Of the 31 documents, eight were dissertations that used the MAP as a measure of growth. Only the abstracts of these eight studies were available for review. One study involved the use of MAP to measure the academic growth of students receiving instruction through American Sign Language/English bilingual model (Lange, Lane-Outlaw, Lange, & Sherwood, 2013). Due to its use in several peer reviewed studies, I concluded from my review of the MAP that it is a suitable instrument for measurement of growth in this study.

Data Collection

Each participant completed a MAP mathematics pre-test at the start of the six-week program and a post-test again at the end of the program. Testing conditions followed protocols established for the administering of state tests. Those protocols include no electronic devices in testing area, no discussion or helping on the assessment, use of only materials provided within the computerized testing environment itself, for example, calculators. The treatment period lasted be six weeks from the end of June through the first week of August. I administered the post-test to all participants at the end of the six-week period.

Data Analysis

In order to determine the effects of Khan Academy use as an instructional tool, participant pre-test and post-test scores were compared. The MAP was administered for both the pre-test and the post-test. I examined the difference of means between the treatment and the comparison groups and conducted a t-test comparison as a measure of the statistical significance. Similarly, I compared the growth of all students on the MAP during the six-week period and conducted a t-test comparison of means as a test of statistical significance. Finally, I compared the actual growth found to the expected growth as defined in the *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth*. (Thum & Hauser, 2015).

The analysis applied in most cases was a one-tailed t-test to examine the difference of means between the pre-test and the post-test results. A one-tailed test because only the amount of positive change was of interest. While negative change is possible in some instances, an assumption herein is that students will make either no gain or some gain but are not likely to make negative gains after an application of an instructional treatment. A Pearson's correlational test to investigate the possibility that there is a negative relationship between initial RIT score and change after the application of the treatment was also applied to the data.

CHAPTER III

RESULTS

Quantitative analyses were conducted to address the research questions introduced above. Primarily, results were compared using one-tailed t-test analyses to determine whether the observed differences of means between pre-test and post-test results were statistically significant. A correlational analysis was performed to determine the relationship between the observed change in pre- and post-test scores was related to the initial proficiency of the participants.

Effect of MAP Recommended Practice Pilot.

The first research question addressed was whether there was a difference outcomes between participants in the comparison versus the treatment groups. A one-tailed t-test was conducted to compare the mean growth achieved by each group. The comparison group achieved a higher mean RIT growth ($\mu=7.95$) than the treatment group ($\mu=6.84$). The statistic I obtained from the t-test indicated that there was no statistically significant difference between the means ($p = 0.401, \alpha < 0.05$).

Table 2

Comparison vs. Treatment Pre-test/Post-test Change in Scores, Original Pairs

	n	Mean	SD	Lower	Upper	df	t	Sig.
Comparison	19	7.95	8.59	3.80	12.08	18	0.401	0.345
Treatment	26	6.84	9.08	3.09	10.59	25		

*Statistically significant at $\alpha < .05$

As can be seen from Table 2 and noted in the Methods chapter, the comparison and treatment groups became unbalanced due to attrition.

In order to complete the analysis on the post-test results as described in the Methods section, attrition in the sample was next addressed. Since this was a summer program for students at high risk of credit denial in mathematics, it was to be expected that the sample group would show substantial attrition rates and that non-completion of the program of study would be the case for a substantial number of students (non-finishers), with exact numbers of course not possible to be known prior to treatment. This was found to be the case.

During the course of treatment, 44 students, or 73% of the original sample, completed the entire study, including taking both the pre-test and post-test and participating in the intervention. Sixteen students (non-finishers) opted out of participation, dropped out of the summer math program, or did not take the post-test. A limitation of this study, addressed below and discussed in the Limitations sections, then is attrition (study design threat of mortality within the program of study) and the degree to which attrition might have taken place differentially in the groups in some systematic way that was not a random.

Differential favored the treatment group, with 11 participants attritioning from the comparison group and 5 from the treatment. Possibly of interest for future study is that more students from the comparison group did not finish (11) as compared to the treatment (5); however this study was too small to investigate and interpret meaningfully this difference, which was not part of the study design to examine. In order to examine the potential effects of attrition on the characteristics of the groups following treatment, finisher students who were paired with non-finishers for the comparison assignments were examined for similar characteristics between groups on the elements of

stratification. Appropriate comparison students were identified for all students except for the differential attrition, which could not be addressed due to differences in final sample sizes between the two groups. While samples do not need to be equal for the t-tests applied in the analysis, this remains a limitation because the two groups could have been less equivalent at post-test than originally, introducing some bias in the results. Note that missingness at random or not at random was investigated to the extent possible with some external indicators described below.

Examining Aspects of Attrition. The scores of students who opted out of participation were not included in any calculations. However, in order to help estimate the scope of missingness not at random from attrition of the non-finishers on the final results, I compared initial and final results on the pre- and post-test for finishers and non-finishers. The 13 non-finishers who had post-test scores available achieved an average RIT of 227.26 on the initial MAP assessment as compared to 229.47 for finishers. A t-test was conducted to determine the statistical significance of the difference between the two groups and a result of not statistically significant ($p\text{-value}=.33, \alpha<.05$) was obtained. This helps to support a claim of missing at random between the two groups, but remains some but weak evidence that should be interpreted cautiously because other factors between the two groups may have been different, as well as data from non-finishers who returned for post-test could have been different from non-finishers who did not return for post-test.

Analysis of Rematched Pairs. After accounting for attrition, a second t-test was conducted on the rematched pairs of data. Results of that analysis were similar to the initial test (see Table 3). Again, no statistically significant difference between the comparison group and the treatment group was found ($p = 0.370, \alpha < 0.05$).

Table 3

Comparison vs. Treatment Pre-test/Post-test Change in Scores, Rematched Pairs

	n	Mean	SD	Lower	Upper	df	t	Effect size	Sig.
Comparison	19	7.68	7.42	4.10	11.26	18	0.333	0.12	0.370
Treatment	19	6.79	8.65	2.62	10.96	18			

*Statistically significant at $\alpha < .05$ **Overall Proficiency Growth.**

The second question addressed was the overall effect of the use of Khan Academy on the participants as a whole. To examine that question, the group means of the pre-test scores and post-test results of the entire group were. A difference in the mean of the scores obtained on the pre-test was observed to be higher than that of the post-test. A one-tailed t-test was conducted to determine whether the difference in means was statistically significant.

The results of the t-test indicated that the difference of means was statistically significant. The post-test mean for the entire population of participants ($\mu = 237.53, SD = 11.68$) was observationally higher than the pre-test ($\mu = 230.42, SD = 15.82$). The p -value obtained was 0.008, indicating significance below an $\alpha < 0.05$ (see Table 4). Due to the small population size, I also calculated a Cohen's d statistic to determine the effect of this result. I obtained a Cohen's d of 0.51, which is typically interpreted as a moderate effect.

Table 4

Overall Pre- and Post-test Means, t-test Results

	n	Mean	SD	Lower	Upper	t	df	Effect size	Sig.
Pre-test	45	230.42	15.02	225.85	235.35	-2.48	44	.51	.008*
Post-Test	45	237.53	11.68	234.09	241.55				

*Statistically significant at $\alpha < .05$

The possibility that the difference in scores could be statistically significant for either the comparison or treatment group was also tested. It was observed that the comparison group obtained a higher mean growth than the treatment group. To test the possibility that the either group, particularly the treatment group, could have obtained non-statistically significant results, a separate t-test was conducted. As with the overall population, for both the comparison ($p = .042$, $\alpha < .05$) and the treatment ($p = .041$, $\alpha < .05$) groups there was a statistically significant difference in mean scores (see Table 5) between the pre-test and the post-test.

Table 5
Comparison and Treatment t-test Results, Pre- and Post-test Means

	n	Mean	SD	t	df	Effect size	Sig.
Comparison							
Pre-test	19	234.42	15.49	-1.78	18	.59	.042*
Post-Test	19	242.37	10.98				
Treatment							
Pre-test	26	227.50	13.97	-1.78	25	.52	.041*
Post-Test	26	234.00	10.88				

*Significant at $\alpha < .05$

. Comparison to Expected Growth Norms. To examine the question of whether participants achieved overall growth when compared to normal expectations, results from this study were compared to the normal expected growth as calculated by NWEA. NWEA publishes a norms study periodically that can be used to predict expected growth over varying time periods. Statistics are published for 10th grade expected growth over three time periods, a) fall to winter, b) winter to spring, and c) fall to spring (Thum & Hauser, 2015). Because the participants were high school aged students who entered the summer program immediately at the conclusion of the regular school year the best

comparison period is winter to spring. The rationale was to avoid comparison to a statistic that included recapturing losses from a long summer of no instruction. However, the comparisons to expected norms for all periods were calculated.

Determining Normal Growth. As mentioned above, the most appropriate statistic to use as a comparison for this study is the Winter to Spring expected growth, which NWEA publishes to be 0.85 RIT points. Initially, it was observed that most student (78%), achieved RIT growth higher than the 0.85 expected growth. Additionally, most students (69%) scored higher than expected growth for an entire school year from fall to spring semester (See Figure 4).

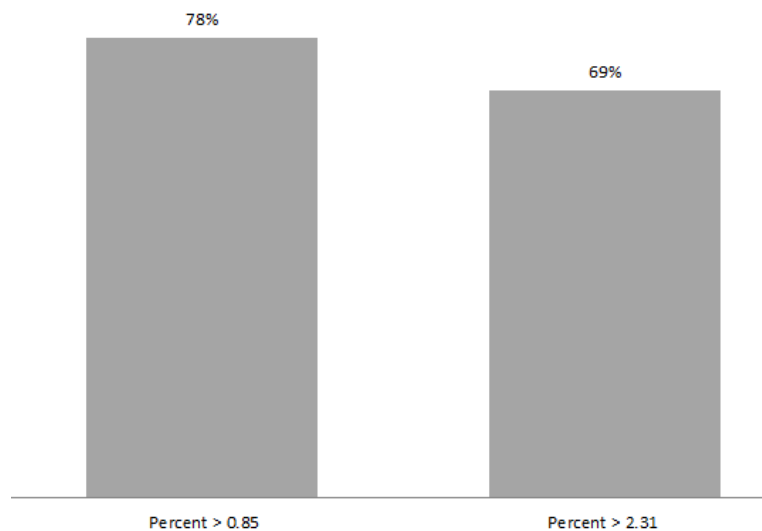


Figure 4. Percent of participants exceeding normal expected growth for winter to spring semester (0.85) and for one school year (2.31).

Observing the change in percentile rank frequency further suggests that use of Khan Academy has a positive effect on student learning outcomes (see Figure 5). The frequencies of percentile ranks have shifted to the right when compared to initial frequencies.

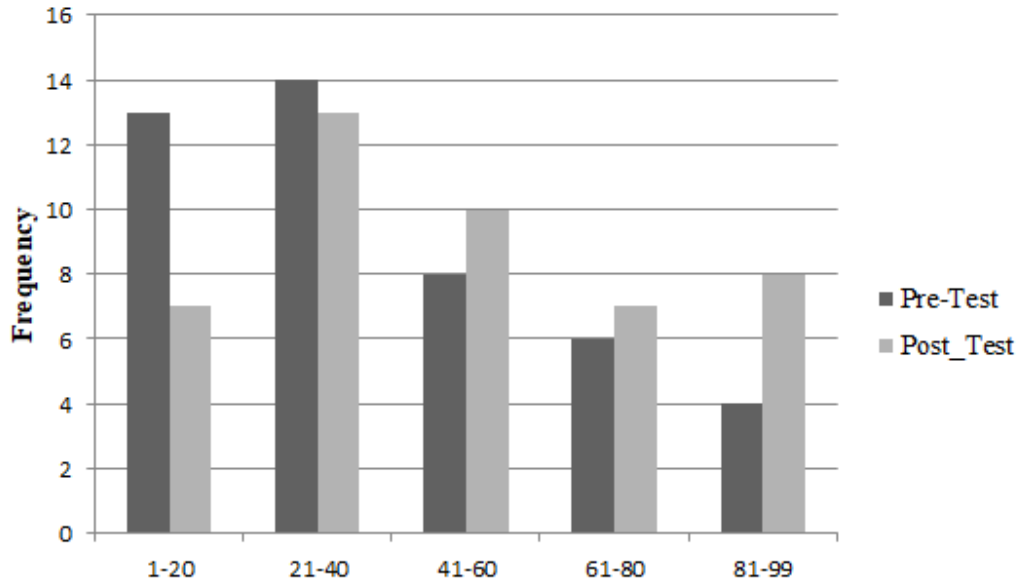


Figure 4. Percentile Rank Range

To determine whether this observation indicated statistically significant growth for the six-week period, a single sample t-test was conducted comparing the mean growth of the overall population and the two assigned groups to the expected growth. The results are summarized in Table 6. Both the comparison group and the treatment group achieved statistically significantly more growth than expected ($p < 0.0001$, $\alpha < 0.05$). This was the same result for all comparisons, including Fall to Winter (1.46) and Fall to Spring (2.31). In each case, there was a statistically significant difference in the mean growth obtained and the expected value ($p < 0.0001$, $\alpha < 0.05$).

Table 6
Observed RIT Growth vs. Expected Growth

	Mean Growth		n	SD	t	Effect size	Sig.*
	Expected	Observed					
Winter to Spring	0.85						
Overall		7.48	45	8.71	4.97	.91	<0.001*
Comparison		7.95	19	8.59	3.50	.99	0.002*
Treatment		6.84	26	9.04	3.23	.80	0.004*
Fall to Winter	1.46						
Overall		7.48	45	8.71	4.51	.80	<0.001*
Comparison		7.95	19	8.59	3.20	.87	0.004*
Treatment		6.84	26	9.04	2.90	.70	0.008*
Fall to Spring	2.31						
Overall		7.48	45	8.71	3.86	.59	<0.001*
Comparison		7.95	19	8.59	2.78	.64	0.012*
Treatment		6.84	26	9.04	2.44	.50	0.022*

*Statistically significant at $\alpha < .05$

Effect of Initial Proficiency on Learning Outcomes

Another important question to consider was whether participants with different levels of math achievement benefited differently from the Khan Academy program. The MAP Recommended Practice should impact students with lower levels of achievement more than students who are at or above proficiency for their level of math instruction. To examine this question, a comparison of participants in the treatment group with a RIT score less than the 10th grade mean of 234, with the similar control group was conducted. Again, the RIT change within the comparison group ($\mu = 13.78$) was higher than the mean for the treatment ($\mu = 8.8$), but the difference was not statistically significant ($p = 0.105$, $\alpha < 0.05$).

Related to the previous question, another consideration is whether participants with lower initial proficiency would benefit more or less than those with higher initial

proficiency. To examine that question, the data were divided into groups based on pre-test RIT scores. Two comparison analyses were conducted. The first compared growth of students who initially scored less than the average 10th graders (RIT score of 234) to those who scored higher. Second, the data were divided into quartiles and again a comparison of the mean growth of each group was conducted. The results of these comparisons are shown in Table 7.

The relationship between initial proficiency and amount of change in RIT score was examined by conducting independent one-tailed t-tests (see Table 7). The difference in means was statistically significant only for students who scored less than 234 on the pre-test. Although positive change in RIT scores was observed for all groups, only students with lower initial proficiency achieved statistically significant results.

Table 7

Comparison of RIT Growth Statistical significance (p-value) by Quartile

	n	Pre-test	Post-test	Mean Change	t	df	Effect size	Sig.
<234	26	220.62	231.15	10.54	-3.82	25	1.08	<0.0001*
234 and more	19	244.79	247.47	2.68	-0.94	18	0.31	0.177
Quartile 1	11	209.45	225.64	16.18	-5.10	10	2.28	<0.0001*
Quartile 2	11	227.55	234.45	6.91	-2.76	10	1.24	0.006*
Quartile 3	11	234.90	238.60	3.55	-1.64	10	0.89	0.573
Quartile 4	12	250.27	253.55	3.27	-1.20	11	0.54	0.122

*Statistically significant at $\alpha < .05$

The results of the comparison discussed above suggested a relationship may exist between initial proficiency and the amount of change experienced by participants. A Pearson's correlation was conducted to compare the relationship between initial RIT scores and RIT score growth (see Table 8). The analysis obtained a statistically

significant negative statistic (-0.622), indicating a negative relationship between initial proficiency level and the amount of positive RIT score change achieved. In other words, students with a lower proficiency are more likely to make statistically significant gains by using the Khan Academy program.

Table 8
Relationship between RIT Score Change and Initial Score

Pre-test RIT Scores	Score Change	
	Pearson Correlation	-0.622**
Sig. (1-tailed)	0.000	
N	44	

**Relationship is statistically significant at the 0.01 level (1-tailed)

In this section, I have analyzed the relationship of initial RIT score levels on learning outcomes. The results have suggested that learning outcomes are related to initial proficiency levels for this data set, in that participants with lower proficiency achieved larger gains than those with higher proficiency.

Participant Perceptions of Learning Progress

Each week students were provided an opportunity to complete a voluntary survey to provide feedback on their perceptions of their learning progress and what tools they utilized to help them acquire new concepts and skills. Generally, students reported satisfaction with their progress (see Figure 6). When asked: How would you rate your learning progress since last reflection, nearly 90% of responses reported being satisfied (42%) or very satisfied (46%).

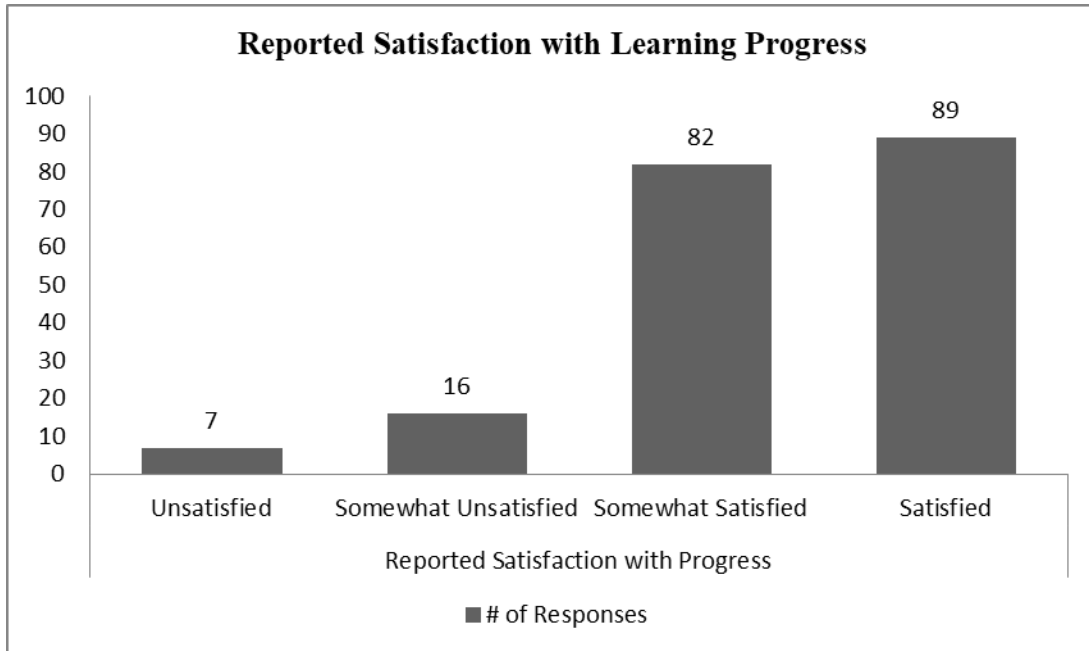


Figure 6. Participant reported satisfaction with learning progress.

Another aspect of student learning examined was how participants utilized Khan Academy to learn the concepts and skills. As noted in Chapter 2, participants were provided an orientation on a specific learning pathway that included using Khan Academy tools before asking the instructor for assistance. In order to examine whether students utilized the proscribed pathway, the weekly survey asked participants to provide feedback on their own learning pathway. Participants reported that they were likely (27%) or very likely (57%) to watch a video when stuck on a problem (see Figure 7). A majority of students reported that they seldom (39%) or never (20%) asked the instructor for help (see Figure 8).

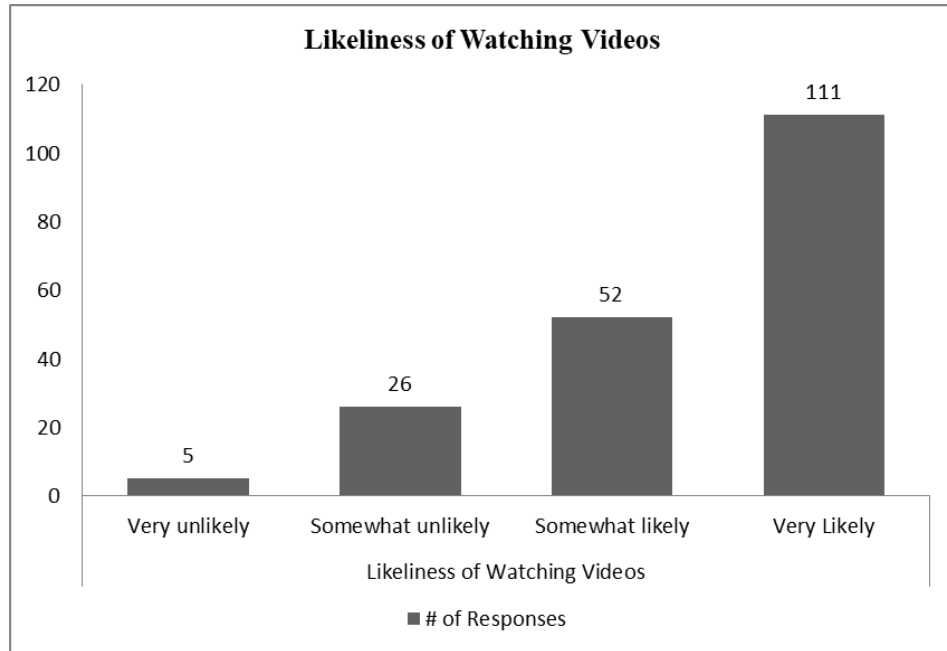


Figure 7. Participant likeliness of watching video as a learning strategy.

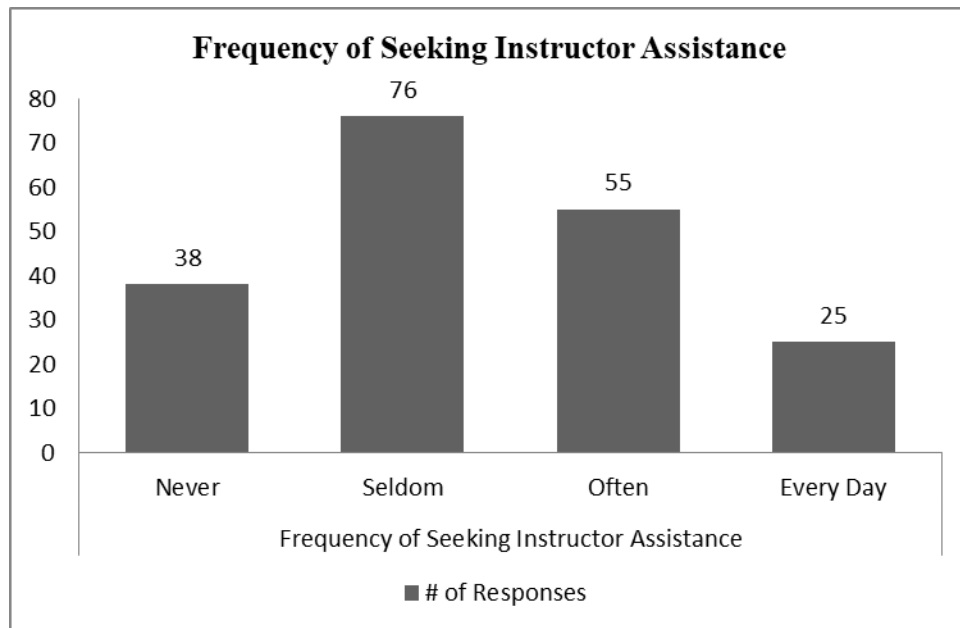


Figure 8. Participant responses, frequency of asking for help

An indicator of engagement with the learning tools on Khan Academy is whether participants were actively interacting with the materials. In order to gain some insight into that question, participants were asked if they regularly took notes from the videos or

the textual explanations. Most respondents indicated that note-taking was a regular activity either often (41%) or always (38%) taking notes (see Figure 9).

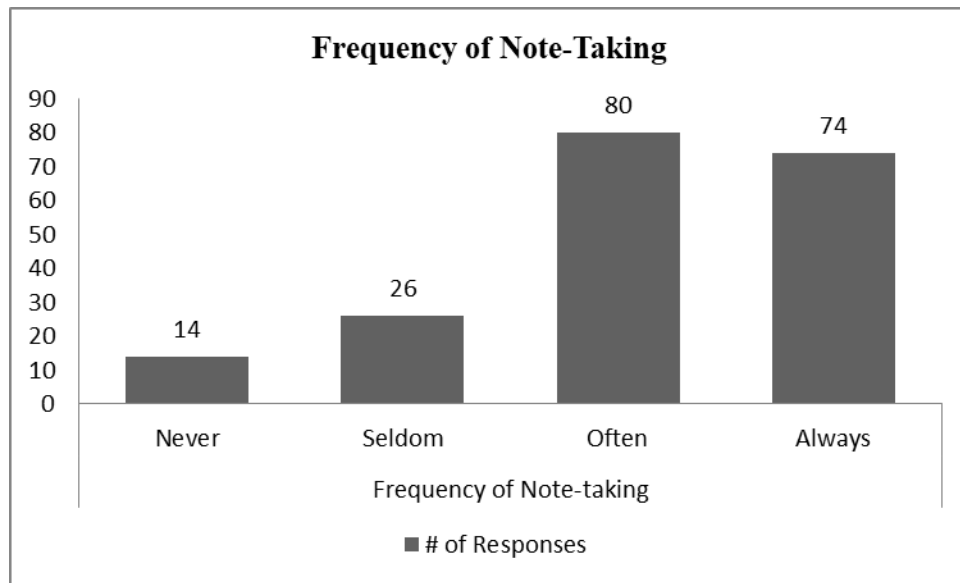


Figure 9. Participant response, frequency of note-taking

In open-ended reflections on their learning experiences, participants commented on obstacles to progress and strategies they could use or did use to overcome them. One participant commented that “taking more notes” was one way to improve their understanding of the math concepts. Several participants echoed that sentiment as well as “slowing down and taking my time.” Other students noted the issue of taking more time to learn the concept and being able to “stop relax and just work on my math” and “take just a little more time and ask for help when needed.” Several participants reflected on their perseverance as an important aspect of learning. For example, one respondent commented that one strategy used was to “stick to a problem until I have succeeded” and another added, “not give up and take a deep breath and ask for help.” Overall, respondents indicated that taking notes, watching videos, using hints on Khan Academy, and asking the instructor for help were all important strategies.

CHAPTER IV

DISCUSSION

Overall, the results of this study suggest that use of Khan Academy is associated with a positive gain in learning outcomes. Used exclusively as the primary classroom resource, most participants in this study showed positive gains that exceeded predicted growth. Participants achieved an average growth of 7.5 RIT points between the pre-test application and the post-test. Compared to a normal growth expectation, as calculated by NWEA, of .85 for a semester of work, this gain is an impressive 8.9 times the expected growth. These results suggest further that use of the Khan Academy platform may be especially beneficial to students who are behind grade level in proficiency.

On the question related to the effectiveness of the alignment with the MAP Recommended Practice Pilot, the results did not support the conclusion that use of the pilot benefited student above using the regular Khan Academy program itself, when the program was employed under best practices with skilled teacher guidance and sufficient teacher time available to do the differentiated instruction manually. There was no statistically significant difference between outcomes of best practice teacher use of differentiated instruction and the automated program. However, since many students may not have access to best practice teacher use of differentiated instruction, or teachers may not have sufficient time to prep differentiation for all students, such as was done for the control group here, the association of the automation with the same level of gains was impressive. It points to use of the new automatically differentiated platform potentially as a support for teachers engaged in mathematics instruction, especially in remediation with high-needs students as in this study that employed a sample of students directed to the

program for additional summer support to improve their limited school-year gains. However, due to the small sample size, no general conclusions should be drawn on the effectiveness of the MAP recommended practice pilot, and, more examination of this question is recommended.

In order to control for the possibility that some students might exert more effort on the post-test than the pre-test, a comparison of results of students who completed sufficient instructional work to earn credit with those who did not was conducted. Earning credit was a potential incentive for all participants. Of the 44 finishers, 36 earned credit and 8 did not. It should be noted that the pool of non-credit earners is very small making conclusions difficult to draw. Nonetheless, a t-test on the difference in mean growth between the two groups, 9.0 for credit earners vs. 1.0 for non-credit earners, found a statistically significant difference. It should be noted that even at an average of 1.0 RIT growth, non-credit earners achieved the expected growth for a semester of instruction. The result should only be tentatively adopted, but it does provide an interesting point for further examination.

Participants' perceptions of their own progress was generally positive and mirrors the actual progress measured. For example, approximately 88% of participants reported being satisfied or very satisfied with their progress while 78% were found to have achieved at least one semester of growth. Perceptions were slightly more favorable than actual observed results, but a finding that high school students who have a history of struggling with mathematics reported satisfaction in their learning results is important. In terms of gaining some insight into the learning pathways of students, many respondents on the weekly feedback surveys expressed that having more time, taking more notes,

watching videos and asking for help were helpful in making progress. Common themes expressed were that taking time, not giving up, and not being afraid to make mistakes were important to learning.

Limitations

There are a number of limitations that should be considered when interpreting the results of this study. Most obviously, this study involved a small number of participants. With only 44 participants, the results are very tentative. In addition, 16 of the initial 60 participants did not complete the study introducing a possible “hardy survivor” effect.

Another limitation is that while this study suggests that Khan Academy is a viable and effective resource for math instruction, there was no comparison to other resources. A possible area for further examination would be to compare Khan Academy use to other online and face-to-face resources and methods.

Participants were in a focused program of primarily math instruction. Some participants were dual enrolled in a second class, but even so, two classes a time is statistically significantly less than what students are normally exposed to in a regular school year environment. This study did not attempt to compare results of students with two classes versus one class. The freedom to focus on just one class could impact these results and further investigation is needed to make any firm conclusions on this point.

Differential attrition between the two study groups, as described in the Results section, was analyzed here to gauge the degree of missingness at random. While evidence of systematic attrition was not found in the approach use, the approach was limited and therefore caution should be exercised in interpreting results. This remains a limitation of

the study that was not possible to address, given the sample and fidelity of outcomes, and would need to be studied in a larger intervention.

Additionally, there was no formal control of the teacher impact on growth. The researcher was the instructor for all participants in the study. The teacher effect in this case was mostly controlled due to the online nature of the program. Instruction was provided primarily through Khan Academy itself and only secondarily and in a support role by the instructor. It should be noted, though, that the courses students engaged in as well as the design of the course including sequencing of activities was determined by the instructor. Further study in this area is difficult due to the responsibility to do what is best for learners and not to subject students to less than optimal practices for the purposes of scientific inquiry.

Another limitation is that no subgroup or other demographic data were collected or analyzed for this study. A suggestion for future studies is to include such data. In particular, it is important to know if English language proficiency is a factor that could lead to statistically significant differences in student learning. Another factor to consider is gender differences and whether males or females respond differently to the treatment.

Finally, it is important to mention that the researcher in this study is also the instructor. While the implementation strategy was purposefully designed to encourage student-directed learning and minimize the instructor role, it is important to keep in mind that the instructor is proficient in the use of Khan Academy as well as integration of web-based learning platforms into classroom instruction. Further study should make attempts to control for instructor effects on learning outcomes.

Implications for Practice

The results of this study suggest several implementation recommendations for schools and districts to consider:

- Remedial programs to boost students currently below grade level.
- Primary instructional resource in alternative education settings.
- Allow sharing students' Khan Academy progress between schools.

An important consideration when implementing a program like this is the role of the instructor. Khan Academy is a resource and an instructional tool. It is not suggested here that this or any other computer or web-based program can replace the role of the instructor in a classroom. While the role for an instructor using Khan Academy as a primary resource may shift from lecturer to guide, it is essential that the instructor continually monitor progress, provide encouragement, and intervene as students navigate the program. As one participant stated when asked what helped them succeed, “by just being there when I need you.”

Conclusions

Khan Academy is a web-based computer application that allows users to learn and practice mathematics skills and concepts. It is a widely known program and is increasingly utilized in classrooms around the world. Despite this popularity, the evidence base for the effectiveness of the platform is lacking. This paper attempted to make a small contribution toward filling that gap in the research.

In addition to the lack of research addressing the effectiveness of Khan Academy, there is concern that teachers have not implemented it in the most effective manner. Sal Khan, the founder of Khan Academy, recommended implementation as part of a flipped

classroom methodology in which students watched videos prior to class meetings then used in class time to practice the skill and receive assistance from the instructor. Many classrooms are unable to implement such a strategy fully because some students do not have access to computers or the internet outside of school hours. This study attempted to look at a more comprehensive implementation strategy: the use of Khan Academy as a primary resource in a student-centered, self-directed classroom.

The previous large-scale studies found promising effects of using Khan Academy but in each case there was no control over implementation strategies. The platform can be used in many different ways and disentangling its effects from those strategies was not a focus of those studies. Also, previous studies found that even within individual classrooms, the implementation strategy was not consistent. In many cases, teachers began employing Khan Academy resources more extensively as they became more familiar with them.

In addition to the above issues, some of the strongest elements of Khan Academy implementation are not accounted for in those studies. The pedagogy of Khan Academy encourages student-centered and student-directed learning. It makes sense that the best use of the platform would be divorced from a regular classroom format and schedule, allowing students to proceed at their own pace. In most cases, the studies cited in this report studied the use of Khan Academy as an extra resource in a classroom, not as a primary instructional tool.

This study sought to address the issues described above. Khan Academy was introduced to participants as a stand-alone, primary resource. Participants were tutored on best practices for using the platform and then engaged in independent learning regularly

over a six-week period with limited guidance, assistance, and direction from the instructor. The results were measured and are herein reported.

This study also examined a beta tool available in Khan Academy called the MAP Recommended Practice. In order to study that question, participants were randomly assigned to a comparison group and a treatment group. The results of those groups were separately compared and analyzed.

This study found that there was an overall positive association of using Khan Academy as a primary instructional resource on learning outcomes for both groups. On average, participants in this study demonstrated statistically significant growth. Generally, students outperformed expected growth norms, even when comparing this six-week program to expected annual growth. That observation is especially true for students who initially assessed at less than a 10th grade achievement level in mathematics. Although students at all levels achieved a measured growth that averaged more than expected growth, the growth rates for students with higher initial levels were not found to be statistically significant.

As to the question of the MAP recommended practice beta tool, there was no statistically significant difference between the achievement results vis-à-vis the comparison group. In fact, overall, the treatment group achieved slightly lower average growth, though the difference was not statistically significant. Small participant size could have played a role in this lack of finding, but without further data no determination can be made.

In addition to the above findings, this study lends support to previous large-scale studies that found students experienced positive achievement growth after using Khan

Academy. Previous studies had found that a majority of students achieve positive results using Khan Academy in a variety of ways. The results of this study support those findings and expand them to a particular implementation strategy: the use of Khan Academy as a primary instructional resource. Additionally, this study suggests that lower achieving students may benefit the most from use of Khan Academy.

These results suggest potential implementation strategies for the educational setting. One use would be as a remedial program to raise students up to grade level. Participants with initial proficiencies in the lowest quartile benefited most from the summer program, regaining on average over 16 RIT points toward grade level. Participants in the second lowest quartile also made statistically significant gains. Of the 11 students in that quartile, six of them went from scoring below grade level to achieving above grade level scores.

While participants in the upper two quartiles did not demonstrate a statistically significant result, it should be noted that change in pre- to post-test scores was generally positive. The mean growth, like that of the lower two quartiles, was higher for both groups than the expected annual growth. In this case, further study is recommended with larger sample sizes to strengthen any conclusions regarding the effects of Khan Academy on this population. I would not interpret these results as suggesting that use of Khan Academy does not benefit higher level students. In fact, the results optimistically suggest the opposite could be true, but further study is required.

It should be noted that this study was carried out in a blended environment combining online learning with face to face, usually one on one, instruction. The role of the teacher could best be described as the *guide on the side* style as opposed to sage on

the stage. The primary motivation for participants in this program was a desire to earn credit, which was directly tied to proficiency achieved in Khan Academy.

This study generated encouraging results for the use of Khan Academy as a mathematical instructional tool. Further study should focus on the effects of use by higher level students, use in different controlled settings, and controlled study focusing on various implementation strategies.

Khan Academy is currently a free web-based resource. If the effects found in this study are replicable, incorporating the use of the platform into mathematics instruction could yield positive results, in particular for students who are behind grade level. A limitation to implementation is the technological infrastructure required but it is possible that the use of expensive textbooks could be reduced or eliminated. The strategy implemented in this study was not a traditional classroom and replicating it might be difficult in traditional school structures.

Another consideration is using Khan Academy with non-traditional students who have difficulty attending school regularly. Khan Academy allows students to access instructional materials without missing lessons due to absences. Also, the instruction is individualized for each student, allowing students to advance at their own pace rather than the regular pacing of a traditional classroom. Access to computers and the internet is a limiting factor, but becoming less so. One possible beneficial use to address the specific needs of students who experience multiple school changes is to allow students to transition their Khan Academy progress from school to school. This option would prevent such students from losing progress or experiencing content discontinuities during transitions.

The conclusion of this study is that Khan Academy is associated with learning gains for this sample that indicate it was an effective tool for learning mathematics, either in the automated differentiation approach, or in best practices teacher differentiation for learning programs. Based on the results of this study, use of Khan Academy was found to be particularly useful for low proficiency students, and students at proficiency levels less than 10th grade level, although benefits are not limited to these categories but encompassed the span of students.

APPENDIX A

Implementation Guide The purpose of this guide is to provide a framework for implementing future replicative studies or using Khan Academy in a school setting. Critical to student success is that the instructor's presence is felt by students daily in terms of feedback on progress and offers of assistance when needed. **Monitoring Progress.** It is advised that instructors use a learning management system as a side-by-side instructional support for the purposes of providing timely feedback, encouragement, as well as monitoring progress. Students can self-monitor progress through Khan Academy. Suggested LMS applications include free web-based applications such as Edmodo, Schoology, or Google Classroom. In this study, Google Classroom and school-based Gmail were used. The selection of an LMS is a matter of instructor preference. **Orientation.** The instructor orients all students either as a group or individually in the use of Khan Academy. Orientation includes technical matters and best practices for using Khan Academy as a learning resource. Technical matters include instruction in accessing lessons, turning in lessons, and tracking progress. Using Khan Academy as a learning resource includes explicitly outlining a procedure for lesson completion. **Best Practices for Learning.** Students are instructed to follow an explicit learning pathway (See Figure 10) that includes:

- Examine the task. Student assess whether they have background knowledge to attempt a problem. If they feel they do, then they make an attempt to solve the problem.
- Watch a video or use hints. If a student decides they need more instruction in order to be successful, they are instructed to either watch a recommended video or

use the hints which provide the student with a step-by-step solution to the problem. After using the learning tools, the student is instructed to make an attempt to solve the problem.

- **Attempt a solution.** If a student assesses that they have the skills or they have watched the videos and used the hints, then they attempt to solve a problem. If they achieve a successful result, they go on to the next problem and start the process again.
- **Study the solution.** If the student makes an unsuccessful attempt, they are instructed to study the solution (using hints) and, if necessary, rewatch a video, and make another attempt. This process repeats through a problem set.
- **Seek assistance.** If a student experiences continual failure, it is imperative that they receive support from the instructor. Students are encouraged to self-advocate, but it is essential that the instructor monitor each student and intervene when necessary even if a student has not requested assistance. Students should be allowed to complete one full problem set (typically between four and seven problems) before instructor intervention occurs.

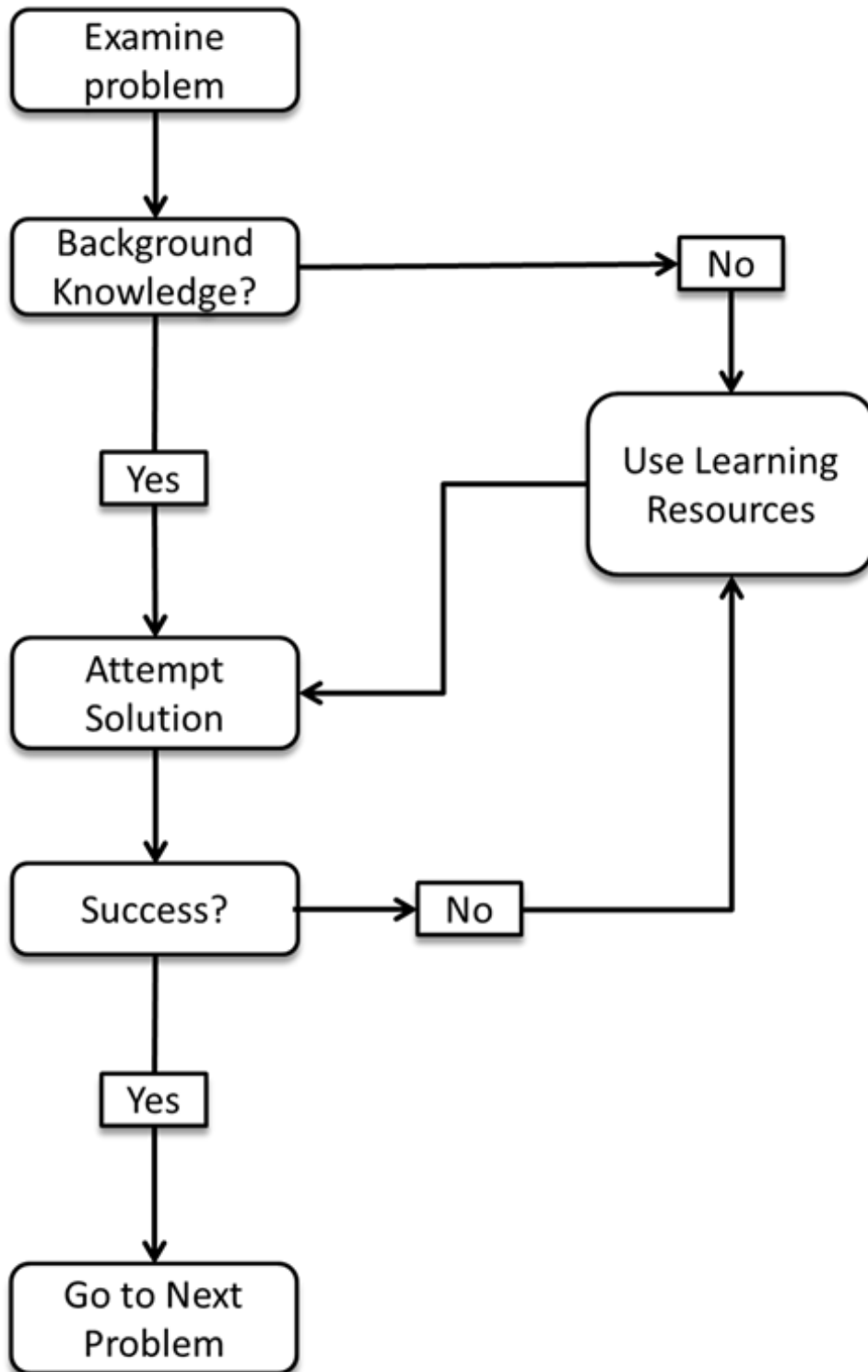


Figure 10.

Student learning pathway.

REFERENCES CITED

- Allen, I. E., Seaman, J., Poulin, R., & Straut, T. T. (2016). Online report card: Tracking online education in the United States. *Sloam Consortium*, 1–4. Retrieved from <http://onlinelearningsurvey.com/reports/onlinereportcard.pdf>
- Cargile, L. A., & Harkness, S. S. (2014). Flip or Flop: Are Math Teachers Using Khan Academy as Envisioned by Sal Khan? *TechTrends*, 59(6), 21–28. <https://doi.org/10.1007/s11528-015-0900-8>
- Kai, K. (2012). Khan Academy: The hype and the reality. *American Educator*, (Fall), 23–25.
- Koedinger, K. R., Mclaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, 43(4), 489–510. <https://doi.org/10.2190/EC.43.4.d>
- Kulik, C.-L. C., Kulik, J. A., Bangert, R. L., & Bangert-Drowns, R. L. (1990). Effectiveness of Mastery Learning Programs: A Meta-Analysis. *Review of Educational Research* Summer, 60(2), 265–299. Retrieved from <http://www.jstor.org/stable/1170612>
- Lange, C. M., Lane-Outlaw, S., Lange, W. E., & Sherwood, D. L. (2013). American sign language/english bilingual model: A longitudinal study of academic growth. *Journal of Deaf Studies and Deaf Education*, 18(4), 532–544. <https://doi.org/10.1093/deafed/ent027>
- Learning, D., & Subbarayan, S. (2012). Lacking Teachers and Textbooks , India's Schools Turn to Khan Academy to Survive. *New York Times*, 10–12. Retrieved from https://india.blogs.nytimes.com/2012/10/15/lacking-teachers-and-textbooks-indias-schools-turn-to-khan-academy-to-survive/?_r=0
- Light, D., & Pierson, E. (2014). Increasing student engagement in math: The study of an intel funded pilot program in chile.
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record*, 115(30303).
- Murphy, R., Gallagher, L., Krumm, A., Mislevy, J., & Hafter, A. (2014). *Khan Academy in Schools*. SRI Education. Retrieved from www.sri.com/education

- Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado Kloos, C. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, 47, 139–148. <https://doi.org/10.1016/j.chb.2014.07.002>
- Strauss, V. (2012). Does the Khan Academy know how to teach? Retrieved January 1, 2001, from https://www.washingtonpost.com/blogs/answer-sheet/post/how-well-does-khan-academy-teach/2012/07/27/gJQA9bWEAX_blog.html
- Thum, Y. M., & Hauser, C. H. (2015). *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth*. Port.
- Timme, N., Baird, M., Bennett, J., Fry, J., Garrison, L., & Maltese, A. (2013). A summer math and physics program for high school students: Student performance and lessons learned in the second year. *The Physics Teacher*, 51(5), 280–285. <https://doi.org/10.1119/1.4801354>
- Wang, S., McCall, M., Jiao, H., & Harris, G. (2013). Construct Validity and Measurement Invariance of Computerized Adaptive Testing: Application to Measures of Academic Progress (MAP) Using Confirmatory Factor Analysis. *Journal of Educational and Developmental Psychology*, 3(1). <https://doi.org/10.5539/jedp.v3n1p88>