ABSTRACT

ACTION RESEARCH ON THE EFFECTIVENESS OF KHAN ACADEMY AS A TIER I INTERVENTION

by David Lee Adams

The present study examined the effectiveness of Khan Academy on math achievement using an action research framework. Three fifth-grade classrooms (N=70), two of which utilized Accelerated Math and the third utilized Khan Academy, were examined. Performance on a test of math achievement was compared across the groups using an Analysis of Covariance (ANCOVA). Data regarding student engagement with and teacher perception of Khan Academy were also collected and presented. Results yielded no significant difference between the group that utilized Khan Academy and the other two groups. Discussion on the implications for practice and future research is presented.
ACTION RESEARCH ON THE EFFECTIVENESS OF KHAN ACADEMY AS A TIER I INTERVENTION

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Specialist of Education

by
David Lee Adams
Miami University
Oxford, Ohio
2016

Advisor: Dr. Jason Abbitt
Reader: Dr. Michael Woodin
Reader: Dr. Michael Edwards

©2016 David Lee Adams
This Thesis titled

ACTION RESEARCH ON THE EFFECTIVENESS OF KHAN ACADEMY AS A TIER I INTERVENTION

by

David Lee Adams

has been approved for publication by

The School of Education, Health and Society

and

Department of Educational Psychology

_____________________________________________
Dr. Jason Abbitt

_____________________________________________
Dr. Michael Woodin

_____________________________________________
Dr. Michael Edwards
# Table of Contents

List of Tables .......................................................................................................................... iv
Introduction ............................................................................................................................ 1

## Literature Review

- Response to intervention ................................................................................................. 1
- Evidence-based interventions ........................................................................................... 2
- Action research .................................................................................................................. 3
- Khan Academy .................................................................................................................. 4
- Mastery learning ............................................................................................................... 5
- Data-based decision making ............................................................................................. 6
- Student-directed intervention ........................................................................................... 6
- Khan Academy research ................................................................................................... 7

## Methods

- Research design ................................................................................................................ 9
- Participants ....................................................................................................................... 9
- Materials .......................................................................................................................... 9
- Procedure ......................................................................................................................... 10
- Results ............................................................................................................................. 10
- Discussion ......................................................................................................................... 13
- Limitations ....................................................................................................................... 14
- Considerations for continued research ........................................................................... 14
- References ......................................................................................................................... 15
List of Tables

Table 1: Descriptive statistics for fall achievement...........................................11
Table 2: Descriptive statistics for spring achievement........................................11
Table 3: Descriptive statistics for time spent on Khan Academy........................11
Table 4: Teacher-participant interview results..................................................12
Introduction

The waters of education are rarely placid. The tides of politics and ideology constantly push and pull our education system in various directions. Paramount among these tidal forces are economic factors. The education system is, at best, at the mercy of the whim of the taxpayer and, at worst, subjected to the brunt of a failing economy. Given this precarious position, administrators and educators often must make difficult choices due to limited resources. These choices, in turn, can have massive implications on the students they serve.

One such choice is in which materials and interventions a school district should invest. Top-quality materials and interventions are rarely cheap, meaning the decision for products usually results in choosing one over another. Moreover, in the 2002 resigning of the Education and Secondary Education Act, commonly known as No Child Left Behind (recently resigned as the Every Student Succeeds Act, 2015), a mandate was placed on schools that stated if they were to receive federal funds for products, scientifically-based research was required to support those products (Caruthers, 2009). Reviewers of academic interventions, such as the What Works in Education Clearinghouse and the Best Evidence Encyclopedia, document whether or not an intervention has enough of the best scientifically-based research to support it (Slavin & Lake, 2007; What Works Clearinghouse, 2014). Within our capitalist marketplace, many of the interventions and materials reviewed by these sites are developed and sold by education publishers which means they carry a price tag that may be quite costly. Such companies possess the resources to effectively market their products and to catalyze internal research for them. In turn, this raises the price of these materials.

However alternatives exist. One such intervention is Khan Academy, a free web-based math platform. The fact that this product is free and requires nothing more than an internet-connected computer makes it attractive to educators, policy makers, and parents. All too often though, the old adage of “you get what you pay for” rings true. Therefore, Khan Academy must be subjected to the same scientific rigor as interventions supplied by education publishers. If this platform is shown to be effective, it could potentially reduce a school’s materials budget while also introducing a level of competition into a somewhat insular market.

The central research question of this study is “is Khan Academy as effective as a widely used, research supported intervention when implemented in a classroom setting?” The study will adopt a quasi-experimental action research design to examine this question.

Literature Review

The following sections will outline the literature relevant to the current study. This review will begin by describing the framework within which the study was conducted. This will include a brief review of the delivery model used and a short description of how the research framework fits within this model. Following this, research relevant to the effectiveness of Khan Academy will be presented and critiqued.

Response to intervention.

P.L. 94-142, or the Individuals with Disabilities Education Improvement Act (IDEA 2004), outlined changes in how students with disabilities are identified and served (Johnson, Mellard, Fuchs, & McKnight, 2006). Among these changes include the option to identify a student with a specific learning disability based on his or her responsiveness to scientifically-
based intervention (Richards, Pavri, Golez, Canges, & Murphy, 2007; Zirkel, 2012). The framework for conducting this process has become known as Response to Intervention (RTI).

RTI is a general education initiative that has gained much attention over the last decade. The general premise behind RTI is to screen children for potential difficulties, to provide interventions that are targeted and individualized to these difficulties, and to systematically monitor progress to ensure that these programs are working (Johnson et al., 2006; National Center on Response to, 2010). The National Research Center on Learning Disabilities (NRCLD) states a definition of RTI as “an assessment and intervention process for systematically monitoring student progress and making decisions about the need for instructional modifications or increasingly intensified services using progress monitoring data” (Johnson et al., 2006, pp I.2). Other authors have stated that the definition of RTI marks a paradigm shift from a medical “within child” view of learning difficulties to a more ecologically-driven framework; that is to say, discovering which instructional practices work best for the student (Greenwood & Kim, 2012). This process allows educators to assist children without having to enter the special education process, thus saving time and resources while also delivering targeted teaching within the general education classroom (Johnson et al., 2006; Fuchs, Fuchs, & Compton, 2012).

The RTI model generally follows a three-tiered approach to general education. The primary tier consists of all students in general education classrooms, and the research-based teaching approach they receive which is referred to as the general curriculum (Johnson et al., 2006; National Center on Response to Intervention, 2010). The secondary tier includes education within the general curriculum but also introduces a more targeted and intensive intervention regimen for students who are at-risk for failing. The tertiary tier consists of the most intensive interventions directed at general education students who are most in need of targeted instruction, thought to represent at least 5% of the classroom (National Center on Response to Intervention, 2010; Johnson et. al. 2006). Using a systematic progress monitoring method, educators can identify when to move a student into or out of a certain tier (Johnson et al., 2006). If the children identified within Tiers II and III are successful in responding to the interventions provided, either defined as meeting a set criterion on a progress monitoring assessment or by individual analysis from a problem solving team, the student is moved back into Tier I (Carney & Stiefel, 2008). If he or she does not respond to the interventions used, school personnel will either adjust the interventions used, select new interventions, problem-solve the particular case, and/or refer the student for services designed to determine their eligibility for special education services.

Evidence-based interventions. A foundation of the RTI model is the evidence-based interventions that are used to accelerate student learning. Students in an RTI framework are continually assessed to determine their progress, or responsiveness, with respect to the particular interventions provided (Fuchs & Fuchs, 2006; Johnson et al., 2006). Thus, intervention in an RTI model serves more than simply effective instruction for a student; it also serves as what Fuchs and Fuchs (2006) call the test-stimulus. Essentially, this term refers to the fact that in the RTI model, students are measured against an intervention much like they would be measured against a test in a traditional model. Since interventions play multiple roles within the model, especially this test-stimulus role, it is imperative that there be evidence to show the validity of said interventions (Fuchs & Fuchs, 2006).
Such concerns as listed above beg the question: “What is an evidence-based intervention?” A report from the Office of Special Education and Rehabilitative Services defines scientific or evidence based interventions as interventions that are supported by research that used empirical methods, rigorous data analyses, valid measures or observations, and has been accepted by a peer-reviewed journal (Copenhaver, Rudio, & Utah State Univ, 2005). However, research can come in many forms, some of which provide results that are more generalizable than others. The What Works Clearinghouse and the Best Evidence Encyclopedia, two organizations that evaluate the purported evidence for academic interventions, both regard randomized control trials and regression discontinuity studies as the highest forms of evidence for an intervention (What Works Clearinghouse, 2014; Best Evidence Encyclopedia, 2014). However, research has indicated that only a small fraction of research on the effectiveness of reading and math programs actually utilize a randomized control trial design (Seethaler & Fuchs, 2005).

In addition, as Fuchs and Fuchs (2006) point out, an intervention must not only be evidence-based, but also implemented with fidelity if it is to be used in the test-stimulus role. Implementing an intervention with fidelity means using the intervention in accord with the standard of how that intervention is supposed to be utilized (Hulleman & Cordray, 2009). When an intervention is implemented with high fidelity, it increases both the chances of success of the intervention and the benefit for the participants (Forman et al., 2013). Further complicating the issue, the aforementioned randomized control trial design used as the highest form of evidence for an evidence-based intervention is often conducted with strict constraints to control threats to validity, thus increasing confidence that the measured difference is, in fact, due to the intervention (Hulleman & Cordray, 2009). Therefore, this more controlled setting varies greatly from the classroom where the intervention will be implemented.

In a study to examine how the effectiveness of an intervention is often lost in the field, Hulleman and Cordray (2009) tested the effects of a motivation intervention on students’ perception and learning. Hulleman and Cordray discovered that the relative strength of the intervention decreased significantly when the intervention was moved from the laboratory to the field even when fidelity is controlled. Hulleman and Cordray also noted that the results of the intervention were significantly better than those of the control group, indicating that the previous laboratory research contained some ecological validity. This research leads another question: If the methods used to establish validity of an evidence-based intervention are not wholly valid in the classroom, how does one assess the validity of an intervention? One possible solution is action research.

**Action research.**

Action research involves efficient, investigatory techniques to solve real, situational problems (Yasmeen, 2008; Zambo & Isai, 2013). Action research is most often conducted by teachers to determine the most effective means of instruction for their particular classroom (Yasmeen, 2008). However, because action research is conducted in these real situations, controlling for threats to validity in the traditional sense is impossible, even in experimental or quasi-experimental designs (Feldman, 2007). Thus, some researchers have noted that the traditional view of validity may not apply to action research (Heikkinen, Huttunen, Syrjala, & Pesonen, 2012). Instead, these researchers propose other principles for validation which include historical continuity and workability (Heikkinen et al., 2012). These principles focus on the
outcomes of research (e.g., did it improve classroom instruction) rather than on whether or not the research measures an objective reality (Heikkinen et al., 2012).

In contrast to these principles, other action researchers have noted that validity still plays a central role in determining the quality of the research (Feldman, 2007). Feldman (2007) notes that in addition to these principles, action researchers must also provide an explanation that is grounded in theory of why their actions led to the outcomes described. Feldman states that in so doing, one is able to discern whether or not a researcher’s methods and results are valid with regard to his or her research question.

Thus, action research, being conducted in real situations such as the classroom, could be a possible solution to the effectiveness gap between the laboratory and the field that Hulleman and Cordray (2009) discuss. Action research on an intervention can help determine not only ecological validity for the intervention, but also identify barriers to implementation (Forman et al., 2013). Therefore, when providing the empirical evidence for an RTI intervention, as Fuchs and Fuchs (2006) noted was imperative, action research should be a part of that evidence. Using action research, the present study will examine one such intervention, Khan Academy, to determine its efficacy. However, in keeping with Feldman’s (2007) aforementioned principle, the authors will ground this intervention in theory.

**Khan Academy.**

Khan Academy is a browser-based educational application that focuses primarily on mathematics (Khan Academy, 2014). The website initially began as a collection of YouTube videos, but has since morphed into an interactive site that utilizes practice problems, analytics, and game mechanics (Khan, 2012). Users create a profile for which they receive an individualized “learning dashboard” that breaks down content into discrete skills ranging from single digit addition to calculus (Khan Academy, 2014; Khan, 2012). These skills can be organized by grade level or listed in complete order. Utilizing the user profile, the site records what skills the student has worked on, what they have “mastered,” and provides recommendations as to what to work on next (Khan, 2012). The program randomly generates the problems the students work on in each skill. The criterion for mastery is to answer correctly ten problems in a single skill (Khan, 2012). However, after the first five problems, the program waits 16 hours before issuing the next problem to ensure that the student has retained the information over time.

Since its inception, the Khan Academy website has garnered much attention and praise, endorsed by organizations such as The Bill and Melinda Gates Foundation and Google (Khan Academy, 2014). The program has been adopted for use by whole school districts such as the Los Altos School District and, in partnership with the J. A. and Kathryn Albertson Foundation, was piloted across the state of Idaho (Khan Academy, 2014). While the list of supporters for Khan Academy is impressive, an intervention must be supported by empirical evidence to be considered an evidence-based intervention. In keeping with the aforementioned literature, this study will review the theory behind Khan Academy to ground the results of the action research project. In his book, *The One World Schoolhouse*, Salman Khan (2012) notes that the website is grounded in educational theory, relying heavily on mastery learning, student-directed learning, and data-based decision making (Khan, 2012). The following sections will review these concepts as well as other research on Khan Academy.
Mastery learning.

Mastery learning has its theoretical roots in the methods of Carleton Washburne in the 1920s (Kulick, Kulick, & Bangert-Drowns, 1990). This method was then refined and expanded by John B. Carrol and, later, Benjamin S. Bloom (Guskey & Jung, 2011). While not often used, research indicates that mastery learning does lead to positive student outcomes (Kulick et al., 1997). A more current iteration of this mastery learning concept, called explicit instruction, is predicated on this notion of task analysis (Piper, Marchand-Martella, & Martella, 2010). Much like mastery learning, research on explicit instruction in mathematics suggests that it is effective, especially for basic skills (Kroesbergen, Van Luit, & Maas, 2004; Piper et al., 2010). Two key assumptions involved in mastery learning are that skills can be broken down into a hierarchy of discrete tasks and that nearly all students can learn if given adequate time (Bloom, 1968).

The first assumption in mastery learning is that all students, given enough time, can achieve mastery (Bloom, 1968). Carroll (1963, as cited in Bloom, 1968) states this fact eloquently by pointing out that the goal of education is to eliminate the relationship between aptitude and achievement through instruction. In other words, the premise of teaching is that one can have students demonstrate the same level of achievement regardless of initial aptitude. Support for this idea has been shown through grade norms for standardized tests (Bloom, 1968). These norms show that a criterion met by the top students in one grade is achieved by nearly all students of another grade (Bloom, 1968). Therefore, those students who did not meet the criterion early on required more time to reach the criterion. The essence of this idea is the notion of self-paced learning, where students learn independently at their own rate. Mitra (2005) has shown evidence of the effectiveness of independent, self-paced learning in underprivileged children in India. Mitra (2005) provided underprivileged students in India with a computer that had material pre-installed on it. Mitra (2005) then tested these students and found significant growth in their achievement through this self-paced and self-organized learning environment. Therefore, if a student has not mastered a certain skill, factoring for instance, what that student needs is more time with that skill rather than moving on to a new skill and hoping that he or she will “get it.”

The difficulty then becomes, where does this time come from? Salman Khan proposed that technology may provide a solution (Khan, 2012). With technology that delivers individualized instruction, students are free to move at their own pace (Khan, 2012). This type of self-paced, virtual learning has been shown to be effective in past research (Zhang, Zhao, Zhou, & Nunamaker, 2004). In essence, what this technology allows to happen is access to each student’s zone of proximal development, the area where a student’s learning should be focused to promote optimal learning (Vygotsky, 1978). For a single teacher to access each student’s zone of proximal development would be daunting. Thus, using technology as a tool to not only support, but also challenge students can greatly facilitate learning.

The second claim, that skills can be broken down into discrete tasks, seems to have face-validity; as most of us master skills or lessons in discrete steps. Teachers often divide their lesson plans into discrete units to be moved through and, hopefully, mastered (Guskey & Jung, 2011). However, one can point out two flaws in this assumption: how does one determine when to move on to the next unit and what happens to students who did not learn the material? Bloom’s Learning for Mastery approach attempts to solve this by making a criterion for moving on to the next unit and providing remediation for students who have not met the criterion (Bloom, 1968). However, simply noting whether or not a student has met criterion without providing feedback is not best-practice (Guskey & Jung, 2011). One must provide feedback in
the form of formative assessment, that is, assessment which provides guidance and feedback throughout the unit (Lalley & Gentile, 2009).

Moreover, some scholars have noted similarities between mastery learning and the RTI framework (Guskey & Jung, 2011). Guskey and Jung (2011) state that the RTI framework and mastery learning share core concepts. Among these are individualized instruction, appropriate interventions, and use of progress monitoring. Since the Khan Academy predicates itself on a mastery learning framework, Guskey and Jung’s descriptions indicate that it may also be suitable for RTI.

**Data-based decision making.**

In the RTI framework, it is imperative to have accurate and reliable data that indicate how well a student is progressing. These data are often gathered through curriculum-based assessment of student performance on the local curriculum (Hintze, Christ, & Methe, 2006). Curriculum-based assessment can further be broken down into mastery and general outcome monitoring (Deno et al., 2002; Hintze et al., 2006). Mastery monitoring lends itself to mastery learning and consists of tracking which tasks in the unit the student has completed or not completed (Deno et al., 2002). This type of evaluation consists of creating an interdependent task hierarchy (Deno et al., 2002). The evaluator then determines a student’s mastery of a task based on a test that examines whether or not the student has achieved the standard or criterion for mastery (Deno et al., 2002). Mastery monitoring, also referred to as criterion-referenced, acknowledges whether or not a student has met criterion and, thus, determines a child’s ability relative to the criterion and not relative to his peers (Lalley & Gentile, 2009). On the other hand, general outcome monitoring involves repeatedly sampling a child on an identical task to assess growth on that particular task (Deno et al., 2002). The tasks are chosen so that they are reliable predictors of a child’s academic success in a particular area (Deno et al., 2002).

While both mastery monitoring and general outcome monitoring have their advantages and disadvantages, some scholars suggest an application of both models can be highly beneficial (Kelley, Hosp, & Howell, 2008). Kelly et al. (2008) state that the use of curriculum-based measurement (a type of general outcome measure) within the framework of curriculum-based evaluation (a type of mastery monitoring) can improve the math achievement of students. Kelly et al. note that this procedure involves a problem solving model in which information from the CBM informs the teacher of progress. If progress is insufficient, a task-analysis using a mastery measure indicates what skill is lacking and an intervention targets that skill (Kelly et al., 2008).

This use of data to monitor progress and inform instruction is a central tenant of the RTI model (Fuchs & Fuchs, 2006; Guskey & Jung, 2011; Johnson et al., 2006). The Khan Academy, while not itself a validated progress monitoring tool, does utilize technology to collect data on students that is much akin to the mastery monitoring framework (Khan, 2012). This effortless collection of data may facilitate this process of data-based decision making, if the data collection techniques employed by the program are both reliable and valid.

**Student-directed intervention.**

While being scientifically-based and verified is the most crucial component of an intervention, considering how intrusive or time-consuming an intervention can be is also of great importance (Hulac, Dejong, & Benson, 2012; Witt, 1986). These additional factors can affect how well the intervention is implemented and its overall effectiveness (Forman et al., 2013; Hulac et al., 2012). Hulac et al. (2012) note that one way to alleviate intrusiveness of an
intervention is to employ student-directed interventions. Hulac et al. tested such an intervention for math fluency and found that the students were not only able to self-administer the intervention, but also show more growth during the intervention phase relative to the baseline phase.

Moreover, the use of technology can also help facilitate student-directed learning. Due to the vast amount of information that may be procured via the Internet, students today are able to access information on demand and at a rate that is unprecedented (Ng, 2008). The aforementioned Mitra (2005) study is a prime example of the benefits of integrating technology with student-directed learning. Furthermore, Ng (2008) notes that students prefer the use of self-directed learning websites over more traditional teaching styles. Ng also states that the use of multimedia on these websites can enhance instruction in ways that a traditional approach cannot.

Much like mastery learning, the Khan Academy predicates its usefulness on the fact that it is student-directed (Khan, 2012). Researchers have noted problems with student-directed learning via the internet include students meandering off-topic (Ng, 2008). Since the Khan Academy is self-contained on one website, it may overcome this challenge. Further, Salman Khan states that the student-direction on the site is the only way to successfully integrate mastery learning. In so doing, students are free to move through the course material at their own pace and choose what is suitable for their level of mastery (Khan, 2012).

Khan Academy research.

Quality research on the Khan Academy is sparse. Although endorsements and anecdotal evidence abound, there have only been a few research projects devoted to the program. Part of this may be due to the fact that Khan Academy is a non-profit organization that does not sell itself to schools. However, it is precisely because the program is free and readily accessible that makes it relevant in this age of tight budgets and limited resources.

The most noted research on Khan Academy are two pilot studies conducted by the Los Altos School District, an affluent district located in California (Kronholz, 2012). In a narrative piece, Kronholz (2012) documented the use of Khan Academy in these schools. At Los Altos, Kronholz reports that the program was piloted in two fifth grade classrooms and two seventh grade classrooms. According to Kronholz, Los Altos reported that among the seventh graders who used the platform, all were remedial students and 41 percent scored “proficient” or “advanced” on the California Standards Test. This compares to 23 percent the year before. Kronholz adds that among the fifth graders, 96 percent of those using Khan Academy were “proficient” or “advanced” compared to 91 percent of the rest of the district. While this report is interesting, especially in regard to the remedial students, the narrative does not provide much context to evaluate the methodology. In the seventh grade classroom, the notion that there were more students proficient using Khan Academy than the prior year is meaningless if we do not know the context. Remedial students can vary greatly; therefore, this year’s group may have had students who simply began at a higher level. If it is a repeated measures analysis, the results may also be attributed to maturation. For the fifth grade classrooms, it is obvious that the entire district is achieving excellent rates of proficiency. The added five percent for the Khan classrooms may be attributed to many things, not the least of which could be teacher enthusiasm for the new program.

In an independent study project for the University of Southern California, Barman (2013) evaluated the effectiveness of Khan Academy videos in teaching mathematics. Barman provided the Khan Academy videos to a South African high school mathematics course for a total of three
weeks. Barman also did individual case studies to determine the effectiveness of the videos themselves. Barman showed the video of how to calculate slope, then administered three test questions relating to the video. Barman scored the results as 0, 1, 2, or 3 to represent the number of questions answered correctly. Barman found that the mean score was 2.36 for this video. Barman then conducted this same approach with a video for finding percent. The mean for this was 1.1. Based on this and other data, Barman concluded that the Khan Academy videos are effective for low-level math skills, but ineffective for higher-level math skills. However, Barman does note that the standard for a passing grade at this school was thirty percent. This may indicate that these students do not possess the more basic skills that are involved in some of the higher-level math skills. Furthermore, this study only used the Khan Academy videos and therefore does not indicate the effectiveness of the website platform on any level of math skills.

In a thesis project, Martin (2013) examined Khan Academy’s impact on students’ motivation, confidence, and skill level in mathematics. Martin conducted an eight week, qualitative action research project on 27 fifth grade students to examine this question. In the study, “pathways” of skill progression were created by identifying Khan Academy math modules that met the California Mathematics Standards. Students were then assigned a pathway and worked through the modules utilizing the Khan Academy program. Martin utilized self-made questionnaires and data from the Khan Academy program to examine the impacts on students’ motivation, confidence, and skill level. Martin concluded that the program had a beneficial impact on motivation and confidence, which in turn had a positive impact on skill level. Martin did note, however, that the students seemed to struggle with particular modules, much like the results from the aforementioned Barman study. While this study shows the practical impacts of Khan Academy on motivation and confidence, the question of Khan Academy’s impact on mathematical skill level, as measured by a reliable, valid achievement measure, remains unanswered.

In addition, Light and Pierson (2014) examined the effect Khan Academy had on student engagement with math. They observed Chilean math classes and reported observational information on the differences observed in classes that utilized Khan Academy and those that did not. Light and Pierson (2014) noted that Khan Academy increased student engagement to math tasks. They attributed this increase in engagement to the fact that Khan Academy presented many more math problems to the students than the traditional classroom setting. In addition, they noted that the gamification aspects of Khan Academy as well as the student-driven aspects of it led to more student engagement. Gamification refers to the incorporation of game mechanics and design to other domains. Lastly, they stated that the platform also changed teaching styles by shifting assessment focus to more of a mastery learning and differentiated teaching approach.

Lastly, to determine how Khan Academy was being implemented, Cargile and Harkness (2015) surveyed high school students who were utilizing the platform. Results indicated that the students were not utilizing the platform as envisioned by Salman Khan (2012). This included a lack of project-based learning and active learning strategies; rather, the students were mostly using the platform for drill and practice. Cargile and Harkness (2015) noted that implementation guides were not available at the time of the study. Guides have since been made available on Khan Academy’s website under the Coach Resources section.

The present study seeks to both expand the literature on Khan Academy and to inform practice on its implementation. Given that Khan Academy has little academic literature on its effectiveness, especially when compared against other computer-assisted educational programs,
the present study will help to fill a gap in the literature by answering the following research question: “Do students who utilize Khan Academy have significantly different rates of mathematic achievement when compared to students who use a different, evidence-based math intervention?” In addition, one of the claims made by Khan Academy is that by using its mastery-learning and student-directed approaches, students of varying achievement levels will remain engaged with the program (Khan, 2012). Thus, a subsequent research question will be, “Do students of varying academic achievement levels engage with the Khan Academy platform at comparable levels?”

**Methods**

**Research design.**

This study examined the math achievement of three separate fifth grade classrooms (N=70) in a southwestern Ohio elementary school using a quantitative action research approach. The school district initiated the action research as part of a program evaluation to determine what mathematics intervention would be most appropriate for the school. Two of the classrooms used the school’s already implemented math intervention, Accelerated Math. While both the What Works Clearinghouse and Best Evidence Encyclopedia state there is limited evidence for Accelerated Math (Slavin & Lake, 2007; What Works Clearinghouse, 2010), there have been published studies in support of the intervention (Ysseldyke & Tardrew, 2007). The remaining class implemented Khan Academy solely. The design utilized random sampling via the cluster sampling method, assigned by the school’s administrator. This study analyzed data from this action research to answer the research questions.

**Participants.**

Participants in this study were all fifth grade students from an elementary school in southwestern Ohio who attend the general education classroom. The study examined all three fifth grade classrooms in the school which each contained approximately 25 students, with a total sample size of 70 students. Due to the geographic limitations of the study and small sample size, the participants will only be reflective of this particular school. In spite of this, the goal of the study is to improve practice and this lack of generalizability is not of major concern.

**Materials.**

Participants in the study utilized either Accelerated Math or Khan Academy as a math intervention. Accelerated Math (AM) is a published, scientifically-based intervention for mathematics (Renaissance Learning, 2014). The program is a computer assisted intervention that begins by screening a student and then provides targeted problems that he or she can work on based on that student’s assessed ability level (Renaissance Learning, 2014). Numerous studies have been published that support the effectiveness of Accelerated Math as an intervention (Renaissance Learning, 2014; Ysseldyke & Tardrew, 2007).

Khan Academy is a browser-based, computer assisted program for assisting with mathematics and other subjects (Khan Academy, 2014). While similar to other computer-assisted interventions, Khan Academy differs with its function of allowing students to choose any topic to explore or lesson to try (Khan, 2012). Khan Academy states that this ability to allow for self-paced learning is key to student success (Khan, 2012). Moreover since it is browser-based, Khan Academy is able to be accessed by students from anywhere with an internet connection (Khan Academy, 2014).
Academic achievement was measured by the students’ scores on the STAR Math assessment. The STAR Math assessment is a widely used and researched instrument for measuring student achievement in math (Renaissance Learning, 2014) with reliability coefficients ranging from .79 to .83 (National Center on Intensive Interventions, 2012). The instrument is based on item response theory and is computer adaptive to the student’s achievement level (Renaissance Learning, 2014). This is significant because, as Kronholz (2012) points out, the Los Altos pilot utilized the California Standards Test which may have not differentiated well between high achieving students. Thus, comparing the students on the STAR Assessment may be better suited at teasing out these differences among high achieving students.

Engagement was measured by the amount of minutes a student has spent on Khan Academy. The data for this measure were collected using the Khan Academy Coach Tab, which shows a log of how many minutes each student had spent on the program (Khan Academy, 2014).

Achievement level was defined as low, average, and high and will be determined by examining the students’ scores on the fall STAR assessment. Low was defined as students who scored below the 25th percentile, average between the 25th and 75th percentiles, and high was above the 75th percentile.

Procedure.

This study analyzed data from an action research project that was initiated and implemented by the school district. At the beginning of the school year, the school assigned its fifth grade classrooms to either one of two control groups (AM) or to the treatment group (Khan Academy). The students utilized the interventions during a 45 minute intervention block in which tier I interventions were administered. A university graduate student monitored the implementation of the groups weekly and also noted any difficulties in implementation.

Students were administered the STAR Math assessment three times per year, as per the school’s typical assessment schedule. Data from all three assessments were collected after the school had removed any identifiers from the data for the two comparison groups. The treatment group’s data were collected with identifiers and then given a numerical, non-identifying code. Then, data from the Khan Academy website regarding minutes spent on the program were collected and linked to the treatment group via the coding method. Following this, all identifiers were removed and destroyed.

Results.

To answer the first research question (is there a difference in math achievement between students who utilize Khan Academy or another computer-assisted intervention?), data from the last STAR assessment were analyzed using an analysis of covariance to determine if there was a significant difference between the groups. The data from the first STAR assessment were used as a covariate to mitigate pre-existing differences in the groups. For an analysis of covariance to be conducted, several key assumptions must be met. These include the assumptions for an analysis of variance, normality of the datasets, homogeneity of variances, and independence of observations, as well as linearity of regression and homogeneity of group regression coefficients (Keselman et al., 1998). Descriptive statistics shown for the data sets is presented in Tables 1 and 2.
Table 1
Descriptive Statistics for Fall Achievement

<table>
<thead>
<tr>
<th>Teaching Condition</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan Academy</td>
<td>.4300</td>
<td>.82981</td>
<td>23</td>
</tr>
<tr>
<td>Accelerated Math</td>
<td>.3615</td>
<td>.71951</td>
<td>23</td>
</tr>
<tr>
<td>Accelerated Math</td>
<td>.5209</td>
<td>.76123</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2
Descriptive Statistics for Spring Achievement

<table>
<thead>
<tr>
<th>Teaching Condition</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan Academy</td>
<td>.7734</td>
<td>.75970</td>
<td>23</td>
</tr>
<tr>
<td>Accelerated Math</td>
<td>.7896</td>
<td>.71590</td>
<td>23</td>
</tr>
<tr>
<td>Accelerated Math</td>
<td>.7859</td>
<td>.67583</td>
<td>24</td>
</tr>
</tbody>
</table>

Since the means and standard deviations of the STAR Math assessments are unpublished, the authors transformed the percentile rank of each student into a z-score with a mean of zero and a standard deviation of one. After conducting a one-way analysis of covariance, there is not a significant difference between students who used Accelerated Math and those who used Khan Academy on math achievement controlling for initial math achievement, $F(2, 70) = .447$, $p = .670$.

Furthermore, within the group that utilized Khan Academy, a Spearman rank-order correlation coefficient was computed to determine the relationship between the amount of time spent using Khan Academy and math achievement on the final STAR Math assessment. A nonparametric test was chosen for this analysis as the amount of time students spent on Khan Academy was not normally distributed. The analysis indicated that there was a positive correlation between time spent using Khan Academy and math achievement, $r_s (23) = .535$, $p = .009$. However, a Spearman rank-order correlation coefficient found that there was no correlation between time spent on Khan Academy and initial math achievement, $r_s (23) = .371$, $p = .081$.

To answer the second research question (does Khan Academy equally engage students of differing math ability levels?), descriptive statistics were collected from the three achievement groups on their minutes spent on Khan Academy. These results are displayed in Table 3. However, due to the small sample size for each ability level, further analysis and interpretation is not possible. The ability group of students who scored less than the 25th percentile contained only one student. As such, no statistical data is reported from this group.

Table 3.
Descriptive Statistics for Time Spent on Khan Academy

<table>
<thead>
<tr>
<th>Ability Group</th>
<th>N</th>
<th>Mean (minutes)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average- 25th-75th Percentile</td>
<td>14</td>
<td>828.07</td>
<td>511.06</td>
</tr>
<tr>
<td>High- &gt; 75th Percentile</td>
<td>8</td>
<td>1483.13</td>
<td>911.69</td>
</tr>
</tbody>
</table>

Those who began the school year at the 75th percentile or above for math achievement ended up spending more time on average than those within the average group. However, the large standard deviations for each group suggest that there is much overlap in the two distributions.
Interviews with the implementers of the programs yielded results that showed insight into the ease of implementation of the interventions and were used to address the third research question (what is the level of difficulty for implementing Khan Academy?). A brief summary of the reported main advantages and disadvantages is presented in Table 4. The teacher-participants that implemented Accelerated Math noted that they had received training in the form of a webinar in the program; however, both noted that they felt that they required more training to be fully proficient in the program. In their view, the program did well at differentiating instruction for their students, particularly those with very high and very low math achievement. While they noted that the program allows for some customization such as manipulating a student’s placement, the teacher-participants stated that they preferred to utilize the placements generated by the program. In addition, the teacher-participants had mixed reactions to this fact. On the one hand, they stated that it was an advantage that the Accelerated Math program was largely predetermined and required little customization; however, they also added that, if time was allowed, they would enjoy having more freedom to better align the content with their particular lessons.

The teacher participant that implemented Khan Academy noted that no formal training in its implementation was given. However, a guided overview of the program was provided by a district employee who was familiar with the program. The major advantage of Khan Academy that was reported was its open nature for students to explore topics. This was perceived as allowing students to access material that was on their ability level for both low and high math achieving students. It was also noted that the program was well received by the students. Disadvantages also included the openness of the program which created difficulty in keeping students on a particular task or area of difficulty to work on. In addition, Khan Academy was frequently updated during the study. While these updates were useful and brought welcomed changes, they also required adjustments in using the program. For instance, over the course of the study, Khan Academy began curating the skills by grade level rather than by skills (e.g. 3rd grade skills as opposed to a series of multiplication skill sets). This did not change the fundamental function of the program, but did change how the skills were organized and curated. Thus, the teacher-participant reported that Khan Academy required more experimentation outside of the classroom to determine the best use of its content than other programs previously utilized.

Table 4.
Teacher-Participant Interview Results

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan Academy</td>
<td>Open for students to freely explore topics.</td>
<td>Difficult to keep students focused on one skill.</td>
</tr>
<tr>
<td></td>
<td>Enjoyed by students.</td>
<td>Continuously updated.</td>
</tr>
<tr>
<td></td>
<td>Data is teacher-friendly.</td>
<td></td>
</tr>
<tr>
<td>Accelerated Math</td>
<td>Differentiates students based on ability level well.</td>
<td>Unable to modify/experiment with the program.</td>
</tr>
<tr>
<td></td>
<td>Non-intrusive to teacher time.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple implementation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Links to assessment well.</td>
<td></td>
</tr>
</tbody>
</table>
Discussion.

Khan Academy is predicated on the theoretical frameworks of mastery learning, student-directed learning, and data-based decision making (Khan, 2012). By creating a program that uses these notions as its basis, Khan Academy suggests that it will increase student achievement in math (Khan, 2012). In addition, the idea of Khan Academy being student-directed has been claimed to make the program engaging for students of all ability levels (Khan, 2012). The present study examined these claims using an action research framework to attempt to answer these research questions: 1. Do students who utilize Khan Academy have differing rates of math achievement than those who utilize another, computer-assisted intervention? 2. Do students of differing math achievement levels, as measured by a standardized, math-achievement test, show different rates of engagement with Khan Academy? Lastly, teacher interviews were conducted to answer a third research question: 3. What are the major advantages/disadvantages of implementing computer-assisted interventions?

For the first research question, results from the ANCOVA did not demonstrate a statistically significant difference between the Khan Academy group and the Accelerated Math groups on math achievement. This suggests that those who utilized Khan Academy did not show any added benefit from the program above what their peers experienced, similar to the finding of Martin (2013). Although, this is in contrast to the gains found in the Los Altos School District pilot study reported by Kronholz (2012). This contrast may be explained by many factors that would require additional research including implementation of the program, as there is no standard implementation procedure, student populations, However, it also suggests that the Khan Academy group did not perform significantly less well on math achievement tests than the Accelerated Math group. This may be significant since, as mentioned above, Accelerated Math does have a research base that demonstrates its effectiveness (Ysseldyke & Tardrew, 2007). While this study does not show that Khan Academy is equivalent to Accelerated Math in improving math achievement, it does show that, for this particular classroom, utilizing Khan Academy did not cause these students to perform significantly worse in the area of math achievement.

Moreover, the positive, moderate correlation between time spent on Khan Academy and math achievement at the end of the school year is an interesting finding that will require more research. The fact that this correlation is not present for initial math achievement suggests that some factor influenced this relationship over the course of the study. Further research into the nature of this relationship is needed to determine precisely Khan Academy’s affect on those who utilize the program more than others.

For the second research question, the descriptive statistics reported above for average time spent on Khan Academy per ability group revealed that those who were initially high-achieving spent more time on the program than those who were initially low achieving. Research suggests that both computer-aided interventions and student-directed interventions increase student engagement (Ng, 2008). While the results from this study could not be analyzed to make inferences, it suggests another avenue of research to determine whether the differences in time spent on Khan Academy that were observed are significant. It did not appear that Khan Academy had the impact in this classroom that was seen in the research by Light and Pierson (2014). However, in the classrooms that Light and Pierson (2014) observed, Khan Academy replaced the general education math instruction rather than used as an intervention. Thus, these students would have had much more exposure to Khan Academy and it may have played a larger role in their general classroom assessment.
Finally, the teacher interviews revealed some insight into the nature of implementing computer-assisted interventions. Perhaps most interesting was the fact that all of the teachers appeared to prefer interventions that are more scripted and do not have an “openness” to them like Khan Academy. This is consistent with research suggesting that when one reduces the intrusiveness of an intervention in the classroom, it is easier to maintain fidelity and increases overall effectiveness (Hulac, Dejong, & Benson, 2012; Witt, 1986; Forman et al., 2013). Thus, these teachers may already be aware of this fact and prefer interventions that are not intrusive. In addition, Khan Academy states that the “openness” of its platform allows students to be self-directed and allows the intervention to differentiate by student ability (Khan, 2012). However, the teacher interviews showed that the openness of the platform was more of hindrance and hassle for implementation. This finding appears consistent with the research by Cargile and Harkness (2015). They reported that Khan Academy was not being implemented as envisioned by Salman Khan and they attributed this somewhat to the lack of implementation materials. Since both their study and the present study, additional resources have been added to the Khan Academy website to assist educators in implementing the program. These implementation guides are fertile ground for more research to test Khan Academy’s effectiveness when implemented in the suggested manner.

**Limitations.**

Limitations to this study included a lack of randomization and an inability to control for all potentially contributing variables. Which teacher’s classroom that would utilize Khan Academy was a decision that was determined by the school administration. In addition, students were assigned to each class as the normal process of assigning students that the school had utilized in the past. Therefore, students were not randomly assigned to each condition nor was the condition randomly assigned.

Moreover, the study only had one teacher who utilized Khan Academy. Thus, any findings that could have been attributed to Khan Academy may also have been the result of the teacher who implemented it. However, as an action research study, the aims of this research was to determine Khan Academy’s effectiveness in this particular teacher’s classroom. Therefore, understanding Khan Academy’s objective effectiveness was not a priority for this study.

**Considerations for continued research.**

The present study was part of an action research program to determine the effectiveness of Khan Academy as a tier I intervention. The study found that there was no significant difference between the class that utilized Khan Academy and the other two classes that did not. Further research highlighted by this study is a continued look at how Khan Academy impacts students of varying ability levels, strategies for effective use of student-directed, computer-assisted interventions, and the effects of using Khan Academy as a supplemental program for home-school connections. In addition, examination of the effectiveness of Khan Academy as implemented by the new suggestions on their website would be another avenue of research. This may include research into the effect of further teacher training and fidelity with these implementation guides. Unfortunately, due to district changes, continuation of this action research project was not possible.
References


