

June – 2017

Khan Academy as Supplemental Instruction: A Controlled Study of a Computer-Based Mathematics Intervention



Daniel P. Kelly and Teomara Rutherford
NC State University

Abstract

Khan Academy is a large and popular open educational resource (OER) with little empirical study into its impact on student achievement in mathematics when used in schools. In this study, we examined the use of Khan Academy as a mathematics intervention among seventh grade students over a 4-week period versus a control group. We also compared differences between students who had supplemental mathematics instruction and those who had not. In both cases, we found no statistically significant differences in student test scores. Khan Academy has several internal metrics used to track student performance and use. We found significant relationships between these metrics and student test scores in this study. Khan Academy and other OER provide access to information and knowledge to large numbers of the population. This research adds to the discourse methods by which Khan Academy and other OER may affect learners.

Keywords: Khan Academy, mathematics education, open educational resources, supplemental instruction

Introduction

Open educational resources (OER) represent a broad swath of online materials that reside in the public domain and/or are offered to the general public (Atkins, Brown, & Hammond, 2007). These materials can include online courses such as massive open online courses (MOOCs), video-based instruction, textbooks, or any resource that supports access to knowledge. It is clear that large numbers are using OER, and, although it is difficult to pinpoint exact numbers, metrics from some of the larger distributors are available. For example, MIT's OpenCourseWare project has had over 113 million unique visitors, and Coursera, which offers free online courses from many top universities, has over 13 million registered users (Weiland, 2015). As of 2014, the online math learning resource Khan Academy had more than 10 million unique users per month and was projected to have greater than 12 million per month in 2016 (Murphy, Gallagher, Krumm,

Mislevy, & Hafter, 2014). Information on access and use of OER is largely dependent on web analytics such as unique website visitors, downloads, and number of views. The potential for untracked re-distribution makes the full extent of OER use even more difficult to appreciate. It is clear from the reported number of visitors and users that OER use and access numbers in the hundreds of millions. The current study focuses on the OER Khan Academy, which boasts “A free, world-class education for anyone, anywhere” and, according to Khan’s website, reaches over one million teachers and ten million students (Khan Academy, 2015). Although Khan Academy has its roots in mathematics instruction, it has expanded to cover topics such as science, history, and economics. The lessons provide video-based lecture and problems for students to solve that coincide with those videos. Participants are rewarded with points and badges that serve to provide positive reinforcement and encouragement as they work through each lesson. Student acquisition of these digital badges in other programs has been shown to motivate students towards knowledge and skill development (Gibson, Ostashewski, Flintoff, Grant, & Knight, 2015).

Although marketing and anecdotal claims of Khan Academy’s success are numerous (e.g., Noer, 2012; Thompson, 2011), little empirical research regarding how the use of Khan Academy relates to student learning gains exists in the extant literature. The existing research on Khan Academy focuses on explanations of implementation and correlations with use. For example, Murphy and colleagues (2014) examined student and teacher perception of Khan Academy, their mathematics attitudes, and student test scores in a study of Khan Academy implemented as part of the math classes in nine schools. Use of Khan Academy was positively associated with a reduction in math anxiety and improved math self-beliefs among the students in the study. Although the study focused on how Khan Academy was implemented in the schools, the authors did note that more Khan Academy use was associated with positive achievement outcomes. These claims were merely exploratory given the nature of the study—investigation into voluntarily implementing schools. In a different study, Light and Pierson (2014) examined Khan Academy in Chilean schools through qualitative data collection and analysis. Their focus was on use and student and teacher perceptions but did not address student achievement. They reported that administrators and teachers felt Khan Academy was useful in improving procedural skills, but that face-to-face student-teacher interaction was the best mode for difficult concepts and promoting a deeper understanding of mathematics. One study in South Africa reported an increase in student learning when Khan Academy and other computer assisted learning was included in instruction; however, Khan Academy was used as part of an after-school supplemental mathematics program (Böhmer, Burns, & Crowley, 2014). It is unclear from this study whether it was Khan Academy or the additional instructional resources and time that accounted for the increase in student outcomes.

Purpose

This study seeks to examine the impact of Khan Academy on student mathematics understanding. The study also seeks to add to the discourse regarding whether and how Khan Academy and similar products can be used to supplement or supplant instruction in K-12 classrooms. Supplemental instruction, or instruction provided in addition to regular classwork, has been shown to increase grades, material retention, and motivation (Coletti, Wisniewski, Shapiro, & DiMilla, 2014). There is also evidence that educational technology applications in mathematics can produce positive effects on student achievement

(Cheung & Slavin, 2013). To investigate potential effects of Khan Academy when used as an educational tool for supplemental instruction we asked the following questions:

1. What differences in scores on a standardized assessment exist among students who received teacher-led supplemental mathematics instruction, used Khan Academy as supplemental instruction, and those students who received no supplemental mathematics instruction?
2. Are the internal metrics tracked and reported within Khan Academy (e.g., time spent, points earned, topics mastered) associated with student outcomes?

Method

Context

The lead author was originally hired in the school discussed in this study as a technology education teacher whose responsibilities included the teaching of math enrichment classes detailed below. As part of the class, he included the regular use of Khan Academy and received positive feedback from the students, administration, and parents. In order to determine the efficacy of his decision to use Khan Academy and justify continued use, he compared his students' math class grades to those in other enrichment classes to determine if Khan Academy was an effective means of supplemental instruction. The results of that comparison showed no statistically significant differences between students using Khan Academy, students in other math enrichment classes, and those in English language arts (ELA) enrichment classes. In assessing the data, the lead author discovered that the students' grades in math class were heavily skewed with a high number of students with course grades at or above 100%.

Due to the potential ceiling effects of the previously-used measure, the lead author designed a study with the explicit intent of investigating the association between Khan Academy use and an assessment that would result in more normally distributed scores suitable for statistical analysis. To avoid other confounds, the lead author also ensured that all students within the study received their main math instruction from the same teacher, none of the control groups were using Khan Academy as supplemental instruction, and the access to and activities within Khan Academy were monitored.

The data used in this study were collected prior to the lead author being affiliated with a university and were intended for use as a measure of instructional effectiveness. As such, Institutional Review Board (IRB) approval was not required for the original data collection. IRB approval was sought to use the de-identified student data reported herein, and exempt status was obtained retroactively.

Participants and Setting

This study was conducted in a large suburban charter school in North Carolina and consisted of 114 total participants in three groups. Students at the school had above average scores on standardized, state-wide, end-of-grade mathematics assessments when compared to both the state and the county in which it operates. All of the school's seventh-grade students in general mathematics classes participated in the study.

All of these students received mathematics instruction from the same teacher. The sample included students identified as gifted/talented and/or as special education students as long as they participated in the general mathematics class.

Math enrichment classes were for general supplemental math instruction and not planned in collaboration with the math teacher. Students were assigned their enrichment class based on their elective course schedule and not on their math class or ability and alternated between math and ELA enrichment on a quarterly basis. In this respect, although the distribution of students across treatment and control groups is far from random, there is little reason to believe it is related to student achievement.

Design

A post-test only control-group quasi-experimental design was chosen for this study due to factors related to convenience and accessibility. Classes of students were selected for control ($n=75$, both math and ELA enrichment courses) and treatment ($n=39$) groups based on whether the classes were taught by the lead author.

Procedure

Students in the treatment group worked with Khan Academy for a minimum of 30 minutes per class over a four-week period. This time was tracked by the lead author/classroom teacher and was used as their daily “participation” grade for the class. Students were not given specific topics to work on within Khan Academy. This was intentional and meant to simulate the typical use of Khan Academy as a self-paced and self-directed supplement to classroom mathematics instruction by students working from home.

Control group students attended their math or ELA enrichment class with instruction given at the teacher’s discretion. Teachers of control group classes were asked not to expose their students to Khan Academy or any other computer-based interventions.

Measures

Math performance. At the end of the four-week period, all students (control and treatment) were given a common assessment in their regular math classes. The assessment contained 21 multiple-choice questions. The assessment items were derived from a quasi-random stratified selection of questions from a previously administered, released copy of a North Carolina End-of-Grade mathematics assessment (grade six). Students were given a paper copy of the assessment and recorded their answers on a separate answer sheet. Students also indicated their current Academic Enrichment teacher for classification purposes.

Khan Academy metrics. Khan Academy tracks and reports progress and use of the application by the students. These metrics provided more data from which to analyze student outcomes. Khan Academy tracks factors such as time (in minutes) spent using the program, points attained for accomplishing tasks, and topics mastered. To further investigate Khan Academy and its use in formal and informal educational settings, the relationship between these factors was examined to see what, if any, patterns emerged.

Comparison of Groups

Table 1 displays the mean test scores and standard deviations of the ELA enrichment, mathematics enrichment, and treatment groups, as well as the number of students in each group who were male or female. An independent-samples *t*-test was run to determine if there were differences in the post-assessment scores after the use of Khan Academy. Assessment scores for both groups were normally distributed (assessed by Shapiro-Wilk's test), and there was homogeneity of variances (assessed by Levene's test for equality of variances, $p = .949$). Analysis revealed unremarkable differences in mean post-assessment scores between the combined math and ELA supplement control ($M = 72.22$, $SD = 14.750$) and treatment ($M = 73.75$, $SD = 14.280$) groups. Although the treatment group scored slightly higher ($MD = 1.53$), differences did not rise to a level of statistical significance ($p = .596$).

Two other comparisons were relevant. Khan Academy may be thought of as an alternative to traditional supplemental mathematics instruction that may benefit students by producing higher achievement and/or schools by increasing efficiency in the use of instructional resources. With this question, we would want to know whether Khan Academy instruction was better than or at least no worse than traditional supplemental math instruction. In comparing the Khan Academy class and traditional math supplemental class, there was no difference in test scores, $t(60) = -1.009$, $p = .842$. Given the lack of difference between these two methods of supplemental math instruction, it is worth asking whether extra math assistance is helping the students at all. As such, we compared the combined math supplemental group to the ELA supplemental group. Again, there was no statistically significant difference, $t(112) = .649$, $p = .259$.

Table 1

Descriptive Statistics by Group

<i>Group</i>	<i>N</i>	<i>%</i>	<i>M</i>	<i>SD</i>
Female	60	53	72.15	14.98
Male	54	47	71.59	15.80
ELA	52	46	70.08	17.66
Math	23	20	73.78	10.60
Treatment	39	34	73.18	14.32
<i>Total</i>	<i>114</i>	<i>100</i>	<i>71.89</i>	<i>15.31</i>

Associations between Khan Academy Metrics and Test Score

Correlational relationships within the treatment group were examined using Pearson's *r* (Table 2) between the participants' assessment scores and several metrics provided by the Khan Academy program: minutes spent, topics mastered, and points attained. Analyses found several statistically significant positive correlations with respect to student use of Khan Academy. Interpretation of correlational values considered .30-.50 to be weak, .50-.70 to be moderate, and >.70 to be high (Hinkle, Wiersma, & Jurs, 2003).

Table 2

Khan Academy Metrics Correlations and Descriptive Statistics

N=39	<i>M</i>	<i>SD</i>	<i>Assessment score</i>	<i>Topics mastered</i>	<i>Points attained</i>	<i>Minutes spent</i>
Assessment score	73.18	14.32	–	–	–	–
Topics mastered	111.67	37.71	.51	–	–	–
Points attained	168K	71K	.41	.71	–	–
Minutes spent	667.59	201.47	.13	.76	.58	–

Note. Bolded correlation coefficients are significant at the 0.05 level (2-tailed).

Within Khan Academy, metrics were positively correlated at a moderate to strong relation. Minutes Spent was the metric arguably most unrelated to prior student ability and therefore the least biased test of the potential impact of Khan Academy use. It was not associated with student test score, $r(37) = .132, p = .422$ even though there was a strong positive association between minutes spent using Khan Academy and topics mastered, $r(37) = .76, p < .001$. Student assessment scores were positively correlated with both topics mastered ($r(37) = .51, p = .001$) and points attained ($r(37) = .41, p = .009$).

Discussion

This study adds to the extant literature on the associations between use of Khan Academy and in-class mathematics performance. As contributions, our study presents results of the implementation of Khan Academy in a naturalistic classroom setting, its comparison to two different control groups, and its situation within supplemental mathematics instruction. Although there is no doubt as to the sphere of influence Khan Academy has with respect to the use of technology for instructional purposes (Thompson, 2011), the effect of that influence has not been subject to rigorous scholarly investigation. Although the current study is merely exploratory, we feel that given the limited landscape of research on Khan Academy, our failure to find an association between Khan Academy and math test scores is of note.

This is not to say that Khan Academy is without merit or value. The metrics (minutes spent, points attained, and topics mastered) within the program may have value as predictors of student achievement. However, it is possible that student factors, such as prior mathematics ability or achievement, may be both driving these metrics and test scores. Further study including relevant covariates or introducing random variation in Khan Academy use is needed in order to draw conclusions or make generalizable assumptions.

It is of note that although we found no difference between scores in our treatment and control groups, we also did not find statistically significant differences between those students who participated in supplemental mathematics and those who did not. These results may be due to a general lack of structure in enrichment classes. There may be other outcomes not investigated within the bounds of this study that Khan Academy and/or supplemental mathematics instruction have significant and measurable impacts on,

such as math self-concept and math interest, as described by Murphy et al., (2014). It is also of note that although we found no statically significant differences between groups in this study, the use of Khan Academy was found to be no worse than more traditional forms of supplemental instruction.

This study does not present generalizable data and is limited in scope. More research with greater demographic diversity may present a clearer picture of the effects of Khan Academy and possibly similar programming. This study examined only the use of Khan Academy as it may be used by a student at home without a teacher's guidance or knowledge. As noted previously, further study that reduces threats to internal validity is necessary, as is an extension of studies to more diverse populations.

Conclusion

Open educational resources continue to grow in both number and influence. Khan Academy is no exception, and as its influence and resources grow, the number of topics within its purview grows as well. Khan Academy has expanded beyond mathematics content and now includes science, art, history, and economics.

As technology evolves and becomes increasingly ubiquitous in education, the number and scale of programs such as Khan Academy will be evermore commonplace. Research into the applications themselves, but more importantly the methods employed by such applications, is of paramount importance if we continue to use them in classroom instruction. OER provide access to information and knowledge to a broad swath of the population and may impact learners in a variety of ways. This study did not find associations between use of Khan Academy as supplemental instruction and higher math test scores; however, only one parameter of student success was measured. Continued research into Khan Academy and other Open educational resources is needed to determine the true impact of these resources and determine best practices for the use of these programs both in and out of the classroom.

References

- Atkins, D. E., Brown, J. S., Hammond, A. L. (2007). A review of the open educational resources (OER) movement achievements, challenges, and new opportunities. Retrieved September 2, 2016, from <http://www.hewlett.org/wp-content/uploads/2016/08/ReviewoftheOERMovement.pdf>
- Böhmer, B., Burns, J., & Crowley, L. (2014). *Testing numeric: Evidence from a randomized controlled trial of a computer based mathematics intervention in Cape Town high schools*. Paper presented at the African Education Conference, Addis Ababa, Ethiopia.
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88-113. doi:10.1016/j.edurev.2013.01.001

**Khan Academy as Supplemental Instruction: A Controlled Study of a Computer-Based Mathematics Intervention
Kelly and Rutherford**

- Coletti, K. B., Wisniewski, M. E. O., Shapiro, M. R., & DiMilla, P. A. (2014). Correlating freshman engineers' performance in a general chemistry course to their use of supplemental instruction. In *Proceedings of the American Society for Engineering Education 2014 Annual Conference and Exhibition*, Indianapolis, IN.
- Gibson, D., Ostashewski, N., Flintoff, K., Grant, S., & Knight, E. (2015). Digital badges in education. *Education and Information Technologies*, 20(2), 403-410. doi:10.1007/s10639-013-9291-7
- Hinkle, D., Wiersma, W., & Jurs, H. (2003). *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin Company.
- Khan Academy. (2015). [Website Index]. Retrieved August 1, 2015, from www.khanacademy.org/
- Light, D., & Pierson, E. (2014). *The use of Khan Academy in Chilean classrooms: Study of an Intel funded pilot program in Chile*. Paper presented at the Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on Advanced Learning Technologies, Athens, Greece. doi:10.1109/icalt.2014.65
- Murphy, R., Gallagher, L., Krumm, A., Mislevy, J., & Hafter, A. (2014). *Research on the use of Khan Academy in schools*. Menlo Park, CA: SRI Education.
- Noer, M. (2012 November). One man, one computer, 10 million students: How Khan Academy is reinventing education. *Forbes*, 398-407.
- Thompson, C. (2011). How Khan Academy is changing the rules of education. *Wired Magazine*, (126). Retrieved from: http://www.wired.com/2011/07/ff_khan/
- Weiland, S. (2015). Open educational resources: American ideals, global questions. *Global Education Review*, 2(3), 4-22.

