# Master's Thesis

## Testing Numeric: Evidence from a randomized controlled trial of a computer based mathematics intervention in Cape Town High Schools

29 May 2014

RESEARCHER

**Ms. Bianca Böhmer**

**BHMBIA001**

School of Economics

University of Cape Town

SUPERVISOR

**Dr. Justine Burns**

Associate Professor

School of Economics

University of Cape Town



## Abstract

This thesis presents the results of randomized controlled trial conducted to evaluate a Grade 8 after-school mathematics intervention. The programme employed student coaches to facilitate classes in which Khan Academy resources were used to teach basic numeracy. Large gains of 0.321 standard deviations were observed on basic numeracy outcomes for learners who were selected to be on the programme. Similarly, learners in treatment also scored 0.246 standard deviations higher on core Grade 8 curriculum questions at endline. The improvements in mathematics outcomes were evident for learners throughout the distribution and treatment learners outperformed control group learners on every subsection of the mathematics test. There was also no significant differential treatment effect by gender, race, home language, baseline typing speed or cognitive development. However, treatment learners with better English literacy at baseline scored significantly higher than learners in the bottom third on core grade 8 curriculum questions. Additionally, despite close contact between control and treatment learners, no statistically significant evidence of spillover was detected.

## Acknowledgements

I would like to thank the Potter Foundation and RUBEN who generously provided the funding for the evaluation. Much gratitude to my supervisor, Justine Burns, for the valuable input, guidance and encouragement you have given me over the last two years. Also, a big thank you to Luke Crowley for all of your insights, new ideas and ability to inspire action. To my parents, thank you for always supporting me in any and all endeavours, especially over the course of my Masters. I am also very grateful to Thomas MBewu and Anthea Field for their faith in me at the very beginning of this project and for helping whenever any help was needed. Thanks to Megan Blair for the many inspiring and helpful conversations. A thank you to the whole research team for the excellent research assistance provided with a particular thanks to Cecelia Dube, Salma Kagee, William Crawford, Matamela Netshikulwe and Kim Ingle. To the dedicated liaison teachers and administrators at the partner schools as well as the learners who participated in the programme and the evaluation, thank you for the all of the assistance during the year. Many thanks to the programme managers at Numeric who helped in numerous ways and were always a pleasure to talk to. Lastly, a huge thank you to Andrew Einhorn, without whom this project would never have been.

## 1. Introduction

Education is viewed as an economic and social priority in almost every country in the world. Human capital formation through education increases the productivity of the labour force and is therefore widely regarded as one of the chief drivers of economic growth (Mankiw, Romer & Weil, 1992, Barro, 2001). Not only does education improve economic outcomes, but also contributes to the development of the country more generally by developing a political and social consciousness and increasing civic engagement (Galston, 2001).

Starting in the 1960's, the provision of primary education became an important goal in developing countries. Since then there has been a rapid expansion in access to primary schooling in the developing world (Glewwe & Kremer, 2002). Despite the successes in increasing school enrollment in developing countries, corresponding increases in learning did not take place. This resulted in a shift away from the primary preoccupation of increasing the quantity of education provided to improving quality. In economic literature a higher quality education is linked with an increased probability of employment and higher wages, thereby linking the increase in human capital - through the medium of public education - with economic outcomes (Speakman & Welch, 2006).

There are a number of inputs which are generally viewed as contributing factors in the provision of quality schooling: knowledgeable and invested teachers, a conducive schooling environment, a low learner-educator ratio, good administration and leadership at the school, well-maintained facilities, functional infrastructure, and possibly also the availability of extra-curricular and extension activities. These inputs form a part of the technology of education. However particularly in developing countries the process is not fully understood (Speakman & Welsch, 2006). Nor is an analysis of schooling inputs by any means an exhaustive indication of inputs to a child's education and human capital formation more generally. Here the impact of the home environment, parent's education and attitudes, family income, community norms, and health and nutrition on human capital need to be considered along with the impact of traditional schooling (Heckman, 2006).

The poor quality of schooling in developing countries is clearly evident in the results of international standardised tests which show that the literacy, mathematics and science scores of learners in developing countries are much lower than those achieved by learners in developed countries (Gonzales et. al., 2004, Barro & Lee, 2001). Other indicators such as high levels of grade repetition confirm this fact. This underperformance is not necessarily due to underinvestment or a lack of political will to further education. Instead, many education departments have been shown to operate inefficiently, whilst simultaneously lacking the knowledge to implement improvements (Glewwe, 2002).

The advent of computers promised a fundamental change to the organization of education worldwide and Information and Communication Technology (ICT) in the form of computer assisted learning (CAL) holds particular promise for education in developing countries. Although inadequate infrastructure and weak management would pose challenges to the introduction of computers in schools, introducing computer assisted learning is one of the proposed means of dealing with some of the problems related to teaching and learning in developing countries.

There are numerous reasons for introducing computers into classrooms in the developing world. The first is to help bridge the 'digital divide'. Computers provide access to a wealth of information and make it easier to complete certain academic tasks. Furthermore, providing access to computers may directly increase computer skills and general cognitive skills which are sought after in the labour market (Malamud & Pop-Eleches (2011). Secondly, CAL can be used to improve access and provide quality instruction by utilising remote instruction or digital media as a substitute for traditional teaching in areas where the availability of quality teachers is low. The third reason is that it allows for learner centered instruction, where the pace of instruction can be determined at an individual level. In developing countries this feature is particularly useful. Schooling is a cumulative and continuous process and as such, high levels of absenteeism on the part of learners and teachers and frequent disruptions during teaching time can result in knowledge gaps. Particularly in Mathematics and the Natural Sciences, these gaps in knowledge compound over time if the fundamental concepts are never covered in class, are not fully grasped or are misunderstood by the learner. As CAL allows learners to pick up where they left off, it is particularly well suited to such an environment. The last aim is to bring about a change in pedagogical approach which has been observed in pilot programmes in a number of developed countries. The instant feedback which is available with CAL allows learning to be more interactive. This frees up teacher time to concentrate on helping and encouraging struggling learners or challenging more advanced learners with more advanced exercises or projects as opposed to focusing only on instruction and assessment (Schonfield, Eurich-Fulcer & Britt, 1994). It also allows for less traditional approaches such as 'flipping the classroom' in which lecture style instruction and readings are assigned as homework and practice examples are completed in the classroom. These changes allow for more frequent feedback, enable the learners to engage more actively with the work and opens up the possibility of group work, all of which has been shown to enhance learning (Roschelle et. al., 2000).

The South African education system is faced with a plethora of problems that are commonly encountered in developing countries. As a result, even though enrollment rates are relatively high, the quality of education that the majority of learners receive is exceedingly poor. Consequently, education is high on the political agenda. A number of curriculum reforms have been implemented by the Department of Basic education over the last ten years in an attempt to improve the quality of education. Nevertheless, there is still little consensus about the best way forward. Additionally, there are hundreds of non-profit organizations working to improve the quality of education on offer.

This thesis evaluates one such organization, Numeric. Numeric is an after-school mathematics intervention which focused on improving numeracy levels using computer assisted learning. A pilot programme was run, in the form of a randomized controlled trial, to evaluate the impact of the programme over the course of one year. The evaluation took place in 9 schools in low income areas in the Cape Town Metropolitan Area, targeting learners in their first year of secondary school. Mathematics tests conducted at endline in November revealed that the programme had a positive effect on mathematics outcomes. The basic numeracy scores of learners that had been selected to participate in the prorgamme were found to be 0.321 standard deviations higher than non-participating learners. Similarly, the test scores of learners in treatment on Grade 8 curriculum material were 0.246 standard deviations above those of learners in the control group.

This thesis is divided into 9 sections. Section 2 investigates a series of education interventions from around the world which aim to improve the quality of education. Neither increasing attendance, nor improving access to resources necessarily results in an improvement in outcomes, but programmes utilising computer assisted learning have been shown to be effective at improving test scores. Section 3 provides a brief overview of the education system in South Africa. Section 4 introduces the Numeric programme and explains features of the Khan Academy website. Section 5 expounds on details of the experimental design including the school selection procedure as well as the learner application and selection process. The limits on the external validity of evaluation as a result of the design are discussed and characteristics of the group of selected schools and learner applicants are examined. Section 6 contains particulars about the data collection process and the tests that were administered, whilst Section 7 comprises of information about the dataset itself. Section 8 presents baseline summary statistics on the control and treatment groups and shows that the random selection of learners into treatment and control was successful. Section 9 presents the results and analysis of the results. A positive treatment effect, of 0.321 standard deviations on basic numeracy and 0.246 standard deviations on core grade 8 curriculum material is found. Section 10 presents a summary and discussion of the results. Additionally, self-selection during the application process and the after-school nature of the Numeric programme are unpacked and the implication on the evaluation are discussed. Finally, Section 11 concludes.

## 2. Quality Education through Technology

In a similar fashion to many other countries around the world, school enrollment in South Africa has expanded quickly over the past 30 years to provide access to education both at the primary and secondary level. Currently, the adjusted net primary enrolment rate[1] is 98.8 percent, and even access to secondary and pre-primary education is high by developing country standards (Statistics South Africa, 2010b, Spaull, 2011). In effect, access to primary schooling in South Africa is universal. However the quality[2] of the education received by the median learner is exceedingly poor. An added burden in South Africa is that the distribution in the quality of education is highly skewed by income, and is by almost any measure very unequal (Spaull, 2011).

This lack of quality education can be seen in South Africa's comparatively poor performance relative to developed and middle income countries on international standardised tests. One such example is the 2003 Trends in International Mathematics and Science Study (TIMMS)[3] where South Africa scored last on the

[1] The net primary enrollment rate is defined at the proportion of children aged 7-13 that are attending school.

[2] The term quality of education is used to refer to both the scope and content of the educational material as well as the overall experience and environment. Here the term is used to refer to the learning outcomes of students as measured by standardised tests of particular cognitive skills, such as Mathematics and Science scores on the TIMMS or numeracy and literacy measures of SACMEQ. Such scores are widely available, well defined and easily measurable.

[3] TIMMS is an international comparative study which tests the Mathematics and Science knowledge of learners in the 4th and the 8th grade. It also includes a set of school, educator and learner surveys.

list of participating countries (Gonzales et. al., 2004). Even within the Southern African Development Community (SADC) the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SAQMEC)[4] results show that test scores in South Africa[5] are far below those of Botswana, a country which spends a similar amount on education per learner (Spaull, 2011). Equally telling are the low secondary school completion rates. Spaull (2013) notes that out of every 100 students that enroll in first grade only 40 succeed in passing their National Senior Certificate[6] (NSC). The NSC is the South African matriculation certificate, which is a pre-requisite for tertiary study and is usually required to obtain formal employment. As such the failure to complete secondary school has far reaching consequences for the individual. This trend of widespread access with a low mean quality of education is not unique to South Africa. Whilst 27 percent of Grade 6's in South Africa were found to be functionally illiterate and 40 percent functionally innumerate (Spaull, 2011), a country wide survey in India found that 44 percent of students between the ages of 7 and 12 were unable to read a basic paragraph, whilst 50 percent could not perform a simple subtraction (Banerjee et. al., 2007).

In Kenya, a surprisingly effective deworming intervention raised primary school participation by an average of 7.5 percentage points. However despite the substantial decrease in learner absenteeism no significant effect on learner test scores was detected (Miguel & Kremer, 2004). This suggests that once a learner is enrolled, the marginal effect of attending school for an additional few weeks a year is minimal or negligible. In South Africa school enrollment is high, thus working to increase the quantity of schooling obtained is futile. Instead, policies or interventions that specifically target learner effort, improve the quality of the inputs or provide better instruction need to be considered. One possible approach involves introducing new technologies designed to enhance learning, into the homes and classrooms of learners in developing countries.

A number of projects have sought to address unequal education inputs, by providing access to home computers. Computers are a valuable resource in most academic endeavours and as such the presence of a home computer might be linked with academic achievement. Simply providing an individual with computers and educational software, however, does not automatically translate into higher scores on cognitive tests of mathematics and language. No effect on mathematics or language test scores was found in Peru, where 902 000 laptops were distributed to school children in rural areas between 2008 and 2009, or in low-income schools in the US, where free computers for home use were randomly assigned to half of the learners in the sample (Crista et. al., 2012, Fairlie & Robinson, 2012). In Romania, experiments showed that a voucher system subsidising home computers had a negative effect on Mathematics, Romanian and English standardised tests scores, with learners in families that received the voucher scoring between 0.25 and 0.33 of a standard deviation lower than learners in control (Malamud & Pop-Eleches, 2011).

---

[4] SAQMEC is a set of standardised tests of numeracy and literacy that are conducted with learners in 6th grade in 14 African countries (Spaull, 2011).

[5] South Africa : Percentage of Grade 6 students functionally illiterate, 27.26%, and functionally innumerate, 40.17%, Botswana: Percentage of Grade 6 students functionally illiterate, 10.62%, and functionally innumerate, 22.48%. (Spaull, 2011)

[6] The National Senior Certificate is normally referred to as matric in South Africa.

The subsequent assertion that providing free computers has no effect on human capital is, however, less certain. The outcome variable thus far has been student performance on standardised mathematics and language tests. However, there are numerous other skills which may be imparted through the use of a computer that are not detected on standardised tests of school subjects, but may nevertheless be useful in the labour market. Malamud and Pop-Eleches (2011) report such gains in computer skills and cognitive development, measured using the Raven's Progressive Matrices in Romania. Similar improvements in general cognitive skill and computer literacy were also observed in Peru (Crista et. al., 2012).

The lack of impact on standardised test scores given an increase in computer ownership is perhaps understandable, as the computers are provided unconditionally, with no guarantee that they will be used for academic purposes. However, a more surprising finding is that providing schools with additional resources, intended and used solely for academic purposes, does not necessarily lead to an improvement in the school outcomes either. This applies to various traditional inputs to the education production function as well as computer aided instruction (CAI). In Kenya neither providing additional textbooks nor flipcharts had a significant effect on average test scores (Kremer & Holla, 2009). In the textbooks experiment it was discovered that there was a statistically significant improvement for learners from the pre-test top quintile. The hypothesis is that other learners were not using the textbooks effectively either because the standard of English was too high for the average learner to derive any value from the books or because only top learners chose to use them (Glewwe, Kremer & Moulin, 2002).

Similarly, many state funded programmes, in developed and developing countries, have focused on increasing access to technology in low income schools by providing classroom computers or computer laboratories. This is usually coupled with intensive teacher training and support (Angrist & Lavy, 2002, Louw, Muller & Tredoux, 2008, Barrara-Osorio & Linden, 2009). Ultimately, however, the success of computer assisted learning depends critically on the ability of teachers to integrate the new technology into the classroom (Barrara-Osorio & Linden, 2009). The Computers for Education programme in Columbia received refurbished computers from the private sector that were then installed in public schools. A randomized controlled trial of the programme in 97 schools showed that despite receiving training and technical support, survey responses indicated that teachers failed to make use of the computers provided. Thus, there was no significant impact on language outcomes (Barrara-Osorio & Linden, 2009). Even when teachers utilise computers in the classroom this does not guarantee that cognitive outcomes will improve. Angrist and Lavy (2002) investigated the impact of the Tomorrow-98 programme which oversaw the installation of computers in primary and middle schools funded by the Israeli State Lottery. They found that the number of teachers using CAI increased through the programme, but there was no evidence of a simultaneous increase in mathematics and Hebrew test scores of learners in the $4^{th}$ and $8^{th}$ grade.

The results of two other randomized controlled trials in which computer aided instruction was used as a supplement in low performance schools were much more positive. In Vadodara in India, Banerjee, Cole, Duflo and Linden (2007) evaluated a programme implemented by an NGO in local schools which used computers provided by the state to set up a weekly class, in which Grade 4 learners played mathematics games on a shared computer for 2 hours a week. The improvement in test scores after a year on the programme stood at 0.35 standard deviations. Additionally, the effect was stronger for children at the bottom of the distribution. A similar finding is one made by Barrow, Markman and Rouse (2008). They looked at a randomized controlled trial in three large underachieving urban educational districts in the US and find that students assigned to classes that were randomly selected to receive instruction on a computer in Pre-Algebra and Algebra scored 0.17 standard deviations higher than learners in the control group. Surprisingly both programmes disproportionately benefited weaker learners. This is in stark contrast to other resource programmes such as the provision of textbooks to 25 schools in Kenya (Glewwe, Kremer & Moulin, 2002). Here, the test scores of the average learner were unaffected, although learners in the top quintile outperformed the control school learners. This suggests the possibility of using educational software as a remedial tool.

Additionally, Barrow, Markman and Rouse (2008) note that learners in large classes, classes with a high variation in ability or high absentee rates, benefited more from computer aided instruction in algebra and pre-algebra. This finding is consistent with the hypothesis that CAL enables each learner to advance at their own pace, and is particularly relevant in developing countries, where large class sizes and high absentee rates are real constraints which impede teachers and administrators in providing access to a quality education.

Resource constraints are a real barrier to education in developing countries, however, evidence suggests that a simple transfer or injection of resources is not sufficient to raise the quality of the education received by learners. A wide range of reforms and innovative solutions will be required to extend a quality education to learners in poor districts and remote areas and results from randomized controlled trials suggest that introducing CAL is one approach that has contributed to improving learning outcomes in developed countries.

## 3. Education in South Africa

Schooling in South Africa is divided into pre-school, then seven years of primary school, followed by five years of secondary school. The South African Department of Basic Education (DBE) is responsible for provision of basic education through the provision of learning materials, teacher development, tracking progress through national standardised assessments and development of school management and leadership (Department of Basic Education, 2013a). The DBE is administratively supported by 9 provincial education departments, of which one is the Western Cape Education Department (WCED), which is the province in which the intervention is located.

A number of the challenges facing the South African schooling system are an enduring legacy of the apartheid regime. During this period, the South African education system was segregated along racial lines and non-personnel expenditure was heavily skewed towards white schools and learners (Van der Berg, 2006). Starting in the 1980’s changes in policy gradually reduced the spending differentials and the adoption of Norms and Standards in 2000 ratified pro-poor non-personal government expenditure on education. However, despite these shifts in allocation of resources, huge discrepancies in achievement by race and socio-economic status are still evident. Even in the early years of primary school the literacy and numeracy scores of poorer students are significantly lower (Van der Berg, 2007, Spaull, 2011). This difference in educational attainment increases throughout primary and secondary school and is clearly evident in the comparatively low matriculation rates of previously disadvantaged learners compared to those of white learners (Van der Berg, 2007).

Although multiple factors contribute to schools' inability to use this increase in resources to improve educational outcomes, one important factor is a lack of human resources in poorer schools. Good leadership and efficient school management as well as the availability of teachers with adequate content knowledge who take the time to convey this knowledge are paramount to providing a quality education (Taylor, 2011). However, the content knowledge of teachers in South Africa has both a low mean[7] and a high variance (Spaull, 2011, Spaull, 2013). Additionally, the mean level of self-reported absenteeism of South African teachers is high[8]. Furthermore, the average teacher at a quintile 1to 3 school reported being absent twice as often as the mean number reported by teachers at quintile 5 schools. The combination of low teacher content knowledge and high rates of absenteeism is likely to lead to disruptions in teaching and learning and an incomplete coverage of the curriculum. The resulting learning gaps then hamper the future progress of learners.

One of the attempts at improving the quality of education in South Africa came in the form of curriculum reform. A comprehensive overhaul of the school curriculum began in 1997, after the change in government, with the announcement of Curriculum 2005. It consisted of a shift from a content-based to an outcome-based approach to teaching and learning (Cross, Mungadi & Rouhani, 2002). The policy roots of the new curriculum, the lack of appropriateness to the South African context and the implementation process of outcomes-based education were all widely criticized (Cross, Mungadi & Rouhani, 2002). Recently, in 2012, the Department of Basic Education introduced the new Curriculum and Assessment Statement (CAPS) (Department of Basic Education, 2011a). The new curriculum aimed to streamline the curriculum and make content and assessment standards transparent to the learners and educators (Department of Basic Education, 2011a). It is much more detailed than the previous curriculum, with a teaching plan per term that dictates the order and pace of sections covered. Whilst this ensures that minimum standards are met and the full curriculum is covered, educator autonomy is limited as it does not give educators as much freedom to change timeframes or alter the order of topics to address the needs of a particular class or to assist struggling learners (Ramatlapana & Makonye, 2012).

Home language instruction is encouraged, particularly during foundation phase (Grade 1- 3) in South Africa (Department of Basic Education, 2010). However at secondary school level most schools choose either English or Afrikaans as the official language of teaching and learning. The language of teaching and learning at all nine schools on the programme was English, with the exception of a single class at one

[7] For example, the mean number of multiple choice questions that Grade 6 teachers answered correctly on the Grade 6 SAQMECIII learner numeracy test was 9 out of 16 items (Spaull, 2013).

[8] Whist teachers in South Africa reported being absent between 11.4 and 23.4 days on average (mean for quintile 5 and mean for quintile 1). Comparatively, the range across the 5 quintiles for Namibia is 9.2 to 9.7 days, in Botswana it is 10.3 to 10.9 days and in Mozambique it is 5.9 to 7.2 days (Spaull, 2011). Although it is possible that some of the difference may be due to a higher rate of systematic underreporting in other countries it is unlikely to account for such a large difference.

of the schools where the official medium of instruction was Afrikaans. However, only two of the learners who applied to be on the programme were from this class. Despite English being the official medium of instruction, many of the schools still teach, sometimes predominantly, in a language that is not English, to the detriment of the learners during national assessments.

## 4. Intervention: Numeric and Khan Academy

Numeric is a non-profit organization that aims to teach basic numeracy and inspire a wider interest in mathematics. Their after-school classes use CAL extensively in the form of free online and offline resources from the Khan Academy website. A pilot of the intervention, which is described fully later in this section, was run in nine schools in the Western Cape Metropolitan Area in 2013.

### Khan Academy

Khan Academy is a non-profit organization founded in 2008, whose mission is summarized as "providing a free world-class education for anyone anywhere" (Khan Academy, 2014). The website consists of three main components: educational videos, the knowledge map and the coach resources.

The first resource that the website offers is over five thousand short video clips on a range of topics including Mathematics, Science, Economics, History and Art History (Khan Academy, 2014). Despite that variety of videos available, the only videos that were used on the after-school programme were mathematics related. Watching videos is the main method of instruction on Khan Academy and it enables learners to learn new concepts or revise by watching the relevant video.

The second feature of the website is the knowledge map. The knowledge map is a mathematics exercise framework which allows the learner to practice specific mathematics skills. The exercise at the top of the knowledge map is 1-digit addition. Suggested exercises following 1-digit addition are 1-digit subtraction, 2-digit addition or the number line. This allows the learner to logically progress from one topic to the next all the way through to topics that are usually taught in the first year of University. There are no restrictions governing the order in which exercises need to be completed, so each individual has full autonomy over which exercises they attempt.

Gamefication is used to incentivize and engage the learners. Once a learner has selected an exercise, they are presented with a problem. When the learner answers questions correctly they earn energy points. Conversely, if the learner is struggling with a concept they have the option of getting a hint or they can be directed to an appropriate video. Learners need to answer 5 questions correctly in a row to progress to the first level. They then need to complete two further levels of competency until finally the learner only needs to pass a Mastery Challenge in order to achieve "Mastery" in a particular skill[9].

[9] Each skill on Khan Academy represents a very specific competency. Two examples of skills are: Adding two-digit numbers and Multi-digit division without remainders. Mastering a skill requires that a learner complete a practice exercise (to successfully complete a practice exercise you answer 5 of the same problem correctly in a row) and then pass this question three times in a Mastery challenge (In a Mastery challenge a learner is given problems from 8 different exercises to complete. If they answer correctly, they move up a level in that particular skill).

The third resource that the website provides is a coach tools page. This feature is useful for teachers and facilitators that are tracking the progress of learners in their class. It provides detailed statistics on each learner by skill, including the time spent on a problem, number of problems completed and the level achieved. The summary of student progress allows the coach to instantly see the progress of the class as a whole as well as highlighting the skills learners are struggling to master.

## Numeric

The Numeric after-school intervention was the programme under evaluation. There were a total of 11 classes in the pilot from 9 schools in the Cape Town Metropolitan Area. The implications of the school selection process on the external validity of the results are discussed in Section 5. The intervention was centrally managed by a small team and treatment was consistent across all 11 classes. This section highlights a number of important programme characteristics.

The biggest technical innovation was the creation of the offline video browser (OVB). The OVB took the material provided by Khan Academy and packaged it in a form that was both user friendly and accessible to users in a South African context. Due to bandwidth constraints in South Africa it is infeasible, at most schools, to stream videos over the Internet in a classroom environment. Instead all of the videos that are on the Khan Academy website were centrally downloaded and are then stored locally on a server or hard drive. Users are then able to search the library of videos using the OVB and can watch any video by streaming it from the server.

The programme took place after-school during the week or on a Saturday morning. The learners attended Numeric classes for an hour and a half on a bi-weekly basis. The school computer laboratory was used as the venue for the sessions. Classes were kept small with a maximum class size of 20 learners, to ensure that the latency times[10] were not too high if all learners were present. As a result, there were limited spaces available and learners needed to apply to be considered for selection. The application process was intentionally lengthy to discourage any insufficiently motivated learners from applying. This was an attempt to select the learners that were most likely to remain committed to the programme.

The coaches who facilitate the classes are all students, generally mathematics teachers-in-training. They were hired and trained by Numeric at the beginning of the year. The training included an introduction to Khan Academy and a detailed lesson on how to use the website. The coaches were also given tools and ideas for alternative mathematics activities in situations where the Internet was offline. Each class was assigned one coach for the entire year and out of the 11 classes 9 had coaches that remained with their class for the full duration of the programme, whilst the other two started in the second term and remained until November. If for any reason a coach was unable to attend a class, the programme managers[11] would

---

[10] Latency (or lag) time refers to the time period between a stimulus and the response. In this context latency is high when the internet connection speeds are low. A situation with high latency is frustrating for the learners, as they need to wait to get feedback on their answer or for the next question to load, and is also inefficient as class time is wasted waiting for a response from the computer.

[11] A programme manager is responsible for coach training, communication with schools and parents, material development and general administration for 5-7 schools.

arrange that another coach substitute for them on that day, or if none of the other coaches was available then one of the programme managers themselves would substitute for the coach. Only if there was no one available was the session cancelled.

A typical Numeric class had the following structure. Prior to the start of every class, the learners were given a snack, provided by the schools[12]. In an average session the learners would start or end with arithmetic drills. Such an arithmetic drill test involves completing 60 simple arithmetic sums within an individually determined time limit that decreases as the learner improves[13]. The learners were also encouraged to watch a video at least once per session, but the majority of class time was spent doing mathematics exercises online using the Khan Academy Knowledge Map. The online platform provided the learners with instant feedback on problems completed and allowed for detailed tracking of individual learners by the coach. Although the order of the exercises can be determined by the learner, a guideline (in the form of a postcard sized progress card with suggested exercises) was provided by Numeric to help the learners and coaches structure their term and to keep tangible track of their progress. Once a leaner had mastered 25, 50, 75, 100 or 150 skills then they also qualified for a certificate, which got handed out at the bi-annual awards ceremony.

Occasionally classes took a break from the computers to play mathematics games, such as Math24[14], or attempt Maths Olympiad problems. These additional resources were also used as contingency material in case the schools experienced internet connectivity issues on the day of a Numeric class.

In contrast with most school related mathematics interventions the Numeric programme is not curriculum based. They do not align their schedule with the schools schedule and at no point during the year do they deliberately set aside time to tutor or revise sections covered in class. Instead they start out the year with basic arithmetic at the top of the knowledge map and learners progress down the map from that point.

Another important characteristic of the programme was parental involvement. Learner reports were compiled at the end of each term and sent home to the parents to keep them up to date on the progress their child was making on Khan Academy. The attendance requirement of 75 percent was stringently enforced. Whenever a learner missed a session, the learner's parents were contacted the following day. This was done to establish the reason for the child's absence as well as keeping the parents informed. If a learner continued to miss classes the liaison teacher[15] at the school would also be asked to intervene by

[12] There is little reason to think that this would have led to an improvement in the overall nutrition of learners on treatment vs. learners in control. One reason for this is that many of the schools already have school feeding schemes which provide learners with a lunch every day. Quite often this same system was used to provide learners with their afternoon snack. Food nevertheless is important to the learners as much of the feedback centers around food.

[13] All learners start out with 5 minutes to complete their 60 sums. As soon as a learner is able to complete 60 problems in 5 minutes, they move to a 4 minute time limit. This continues all the way down to 1 minute.

[14] For Maths24 learners are presented with a card with four numbers. The aim is to find a mathematical expression ,,using all four numbers only once, that evaluates to 24.

[15] Each school chose a staff member to be the Numeric liaison teacher at the beginning of the year. This person was responsible for all of the organization and administration of the programme for the school. They were usually a Mathematics teacher or head of department. As such it would have been very unlikely that the liaison teacher was

talking to the learner and calling the parents. Ultimately learners who did not meet the 75 percent attendance requirement were dropped from the programme[16]. The programme managers and coaches also attempted to involve parents at various points during the year in the form of parents' evenings and award ceremonies.

The learners were required to pay a nominal participation fee on a quarterly basis. This was done as the Numeric management felt the parents would take the programme more seriously if it was not free. The participation fee differed between schools and was decided based on school fees. A fee waiver option was available to all applicants, to avoid excluding learners on financial grounds[17]. All it required was that the applicant complete a simple one page financial aid form.

The Numeric programme is a combination of all of the features mentioned. Although using computer assisted learning as a tool to teach and practice mathematics is the core objective of Numeric, any effect needs to be viewed as an outcome of the programme as a whole.

## Organizational Structure

Figure 1 presents a diagram depicting all of the major parties involved in the Numeric pilot and associated evaluation as well as the relationship between these parties. They can be grouped into three broad groups: the implementing organization, Numeric, the UCT evaluation team and the learners and liaison teachers at the schools. The implementing team consisted of Numeric the parent organization, the programme managers, the coaches and the Numeric technical team. They are all shaded orange in Figure 1. On the evaluation side there was UCT, a consultant from IPA, as well as the Independent Assessment Committee. All parties that were involved in the evaluation are show in blue. The last party consisted of the learners and liaison teachers at the nine schools. They are depicted in green in the diagram.

The bulk of the daily operations and communications involved the programme managers and coaches from Numeric who were in close contact with the liaison teachers and learners at the schools. The evaluation team from UCT worked with the liaison teachers to set up testing dates and venues at the schools, they received test and survey data from the learners, and also received programme and attendance data from the programme managers.

The UCT Evaluations team was responsible for the overall design, test administration and the marking and capturing of the data and analysis. Additionally, an IPA consultant also provided input on the experimental design and the construction of various tests, whilst the Independent Assessment Committee

---

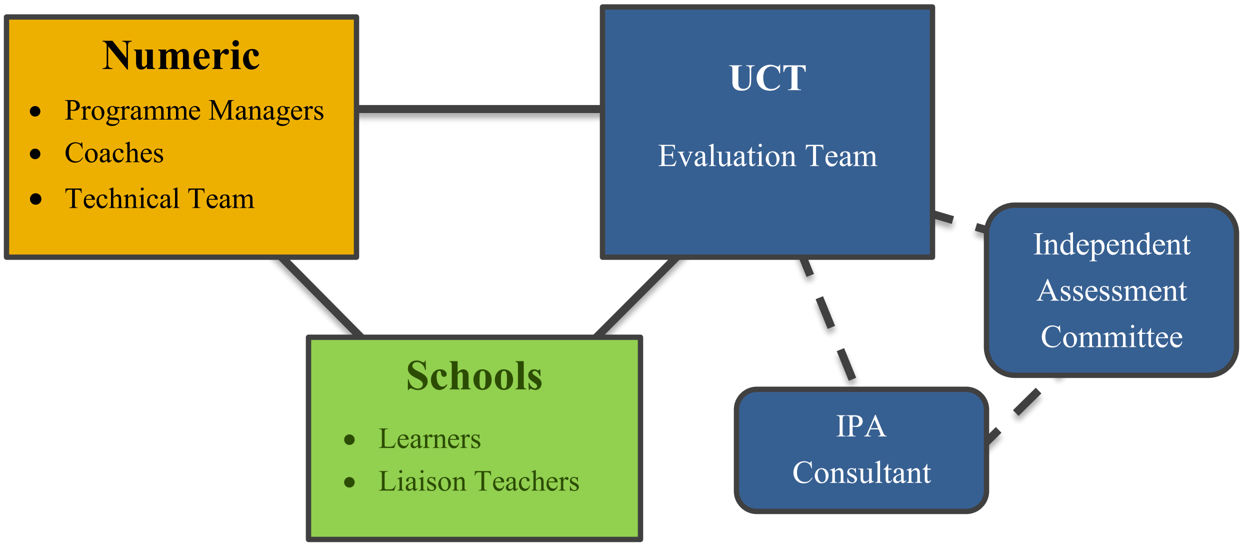
also a Grade 8 class teacher. It is however possible that they may have taught one of the Grade 8 classes mathematics.

[16] Although the programme managers were very strict, this was not always perfectly enforced. In some cases at the discretion of the coach and the programme managers learners were retained even if a learner's attendance rate fell below 75 percent in a particular term. By August 68 percent of the learners that had officially left the programme had either just stopped attending or had dropped out because of poor attendance.

[17] It is possible that some learners decided against applying to the programme as a result of the quarterly fee, but every effort was made to encourage learners to apply regardless of their family's financial situation. Additionally, no child was later excluded from the programme for non-payment. Application rates were also proportionally higher at non-fee paying schools than at schools which charge a nominal school fee, which suggests that this was not a large deterrent in the application process.

was responsible for the creation of the standardised mathematics test which was used as the main outcome measure.

Figure 1: Figure to show the relationships between various organizations, individuals and groups involved in the Numeric pilot




## Experimental Context

The Numeric intervention requires access to a networked computer laboratory, with a connection of at least 4 MB/s. This is a fairly stringent requirement in a country where access to computers is still low and broadband is still slow and prohibitively expensive for the majority of the population (World Bank, 2014, Mybroadband, 2014).

Access to computers at schools is slowly improving. In 2002 39.2 percent of all schools in South Africa had computers, although only 26.5 percent were for teaching and learning (Department of Basic Education, 2004). Whereas in the 2009 Census at School survey the percentage of learners with access to a computer at school was 53 percent, however, only 20 percent of learners in Grade 3 to 12 reported that they had access to the Internet (Statistics South Africa, 2010a). The 2004 White Paper on e-Education stated that computer access and internet connectivity are necessities at both primary and secondary level. However, there is no clear national policy aimed at providing schools with computers, nor is there a consensus on how to fund such an initiative. As such, there has been no coordinated national programme. This lack of centralisation has resulted in huge variation in availability of computers and access to the internet across provinces (Department of Basic Education, 2004).

In the Western Cape a programme named Khanya was spearheaded by the Western Cape Education Department between 2002 and 2012. Through Khanya every secondary school in the province was equipped with a computer laboratory by 2006 (Western Cape Education Department, 2009). Teacher training and technical support were also provided over the lifespan of the initiative. Thus even in poorer areas some schools still had the required infrastructure in place in 2012 or could easily upgrade their facilities to meet the requirements for the Numeric programme. This made the Western Cape an ideal environment to test the programme without incurring the initial expense of setting up computer laboratories with a local area network.

## 5. Experimental Design

Numeric ran a pilot of its programme from January to November of 2013 in nine schools in the Cape Town Metro. The pilot was coupled with a quantitative evaluation of the effects of the programme on standardised mathematics test results.

The evaluation was designed to take the form of a randomized controlled trial[18], with the unit of randomisation at the individual level. Individual leaners were chosen as the unit of observation as opposed to schools or classes, as the treatment was not indiscriminately applied within a treatment school. An oversubscription design was used, in that only learners that applied to be on the programme were eligible for selection into treatment. As this decision was taken at an individual level, random selection into treatment was also made at the learner level. It also greatly increased the statistical power.

The selection of participants took place in two stages. In the first stage schools applied and were selected. This was followed by a set of baseline tests and surveys which were conducted in the selected schools. Thereafter a learner application and selection process took place. Applicants at each school were then either selected to be in the treatment or control groups or were placed on a waiting list.

### School Selection

#### School Selection Process

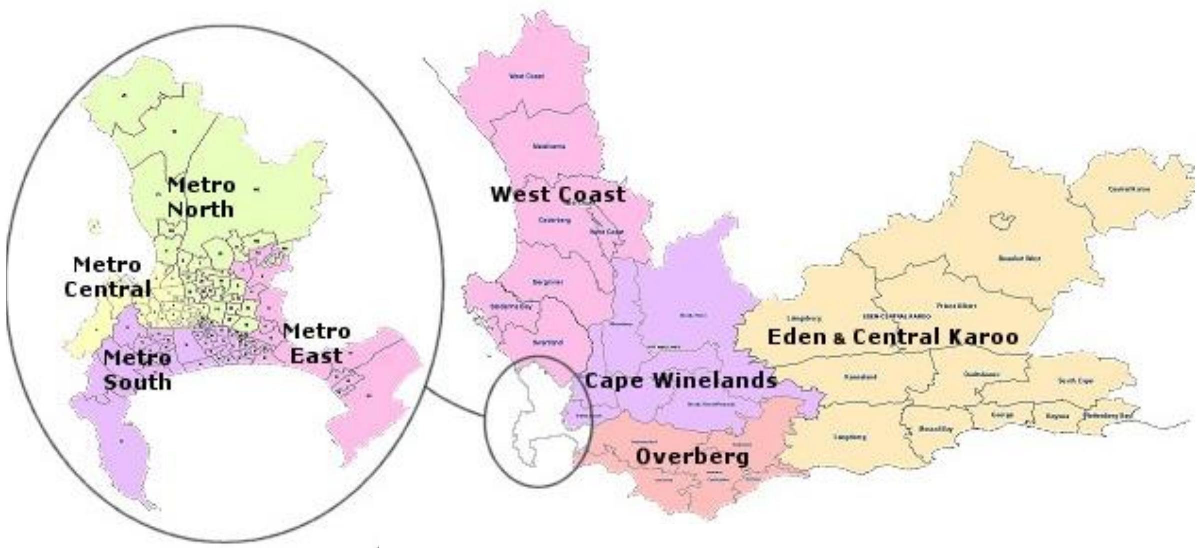
The schools were selected by the implementing organization in collaboration with the circuit managers of the Western Cape Department of Education (WCED). Within the Western Cape schools are divided into 8 administrative districts as shown in Figure 2. Each district is then further broken up into circuits. The process started by identifying poor areas which also had a high density of secondary schools. A circuit manager acts as a link between the Education Department and the governing body of about six to ten schools within a locality. They therefore had reasonable insight into the quality of the leadership, facilities and resources at each school. The circuit managers were asked to recommend any schools in their circuits

[18] A randomized controlled trial (RCT) was chosen as the method of evaluation, as selection into such a programme is a determinant of observable and unobservable learner characteristics. It would therefore be unwise simply to compare the outcomes of learners on the programme and learners that chose not to participate, without correcting for this selection bias. Although econometric methods such a two stage least square estimation or matching can be used for such corrections, these often require very large and diverse samples and the existence of a valid instrumental variable. Randomized controlled trials instead allow for the calculation of an unbiased treatment effect through the creation of a control and treatment group which are comparable- on average- at baseline. An average treatment effect can then easily be calculated, without further corrections, which is free of selection bias despite the presence of unobservables.

that might qualify for the programme given the following two criteria for participation: good management and a working computer laboratory.

A shortlist of schools was compiled by Numeric based on the recommendations of the circuit managers, and the principal, deputy-principals and mathematics teachers at these schools were invited to a meeting. At this meeting the vision for the programme was presented to them and the school was given a detailed list of technical and administrative requirements for the programme. Initial inquiries were made about the computer facilities at the school and an application form for the school was given to the principal.

Figure 2: Map to show the 8 School Districts in the Western Cape




Source: Western Cape Education Department website.

Once the school applications had been received, the Numeric technical team conducted a second thorough inspection of the computer labs and internet connectivity at each school. Upgrades to their system or an increase in their internet connectivity speeds were recommended where applicable. Out of the thirteen schools that applied only two were not accepted into the programme. The first was rejected because the computer laboratory was badly maintained and too small to accommodate a full class, whilst the second was not taken on due to capacity constraints on Numeric's side. This left eleven schools that were initially selected to be a part of the after-school pilot.

Two schools later dropped out of the pilot. The first school requested that Khan based classes be provided to all Grade 8 learners instead of restricting it to a small group. They achieved this by rearranging the school timetable to enable them to incorporate it into their weekly class schedules given the constraint of a single computer laboratory. The second school was scheduled to have an upgrade to their computing facilities in January of 2013. They were conditionally accepted, provisional on the new computer laboratory being installed by the end of February. Unfortunately, construction of the laboratory was only completed at the end of April and as a result they too were excluded from the pilot.

### Characteristics of Selected Schools

The location of the schools within Cape Town was non-random. Of the final nine schools five were located in the Metro Central district, with the remaining four coming from Metro North, South and East. The approach used to select participating schools resulted in a sample of schools which is not representative of the income brackets of schools in the Western Cape Metro District. Table 1 shows the distribution of schools across quintiles[19] in the Western Cape Metro. The information for Table 2 is obtained from the DBE's Education Management Information System (EMIS) database. Column 2 shows that 50 percent of schools in the Cape Town Metro are in the top quintile. In contrast, two thirds of the schools in the sample are quintile 5 schools. This distribution appears skewed in favour of schools in purportedly wealthier areas, which is unusual for a development programme. A Pearson's $\chi^2$ test comparing column 2 and column 4 of Table 1 produces a p-value of 0.683. Thus we fail to reject the null hypothesis that the frequency distribution of selected schools is different to the underlying population distribution within the Western Cape.

Table 1: Proportion of secondary schools in each quintile in the Cape Metro and the distribution of participating schools

| Quintile | Cape Metro (%) | 11 initial schools (%) | 9 Sample Schools (%) |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 2 | 9 | 9 | 0 |
| 3 | 15 | 18 | 22 |
| 4 | 25 | 18 | 11 |
| 5 | 50 | 55 | 67 |
| **Total** | **100** | **100** | **100** |

Source: Western Cape Ordinary Schools Masterlist from the EMIS Database (Department of Basic Education, 2013c) and own data (2013)

It is telling however, that the only school that was provisionally selected, but was unable to participate came from quintile 2. This shows that a lack of appropriate human and physical resources is a barrier to participation in the programme. Additionally, weaker management may negatively affect the efficacy of the programme if such schools were included. If the programme is expanded to areas in South Africa where the necessary infrastructure does not already exist within a school, extensive work would need to be done prior to the start of the programme. An appropriate facility would need to be built and technical support set up. A maintenance plan would also need to be designed in collaboration with the school.

Table 2 shows some demographic and socio-economic characteristics of the schools that were selected to be a part of the programme in 2013. The schools are ranked by average annual school fees in ascending

[19] A school is assigned a quintile by the national Education department according to the socio-economic characteristics of the area around the school as well as the infrastructure and resources available to the school. School in quintile 5 are the richest 20 percent of schools in South Africa, whilst the schools in quintile 1 are classified as the poorest schools. All quintile 1-3 schools are automatically required to be no-fee schools (Western Cape Education Department, 2013a). The funds allocated by the department are meted out according to quintiles, thus this ranking is sometimes contended by the schools in question.

order[20]. Two of the sample schools were no-fee-schools, whilst average annual school fees across the nine schools was R 1 124.09. Half of the learners at sample schools were Coloured, whilst 47 percent were African. English was the most widely spoken home language at 44 percent, closely followed by isiXhosa, which 42 percent of learners in sample schools spoke at home, with Afrikaans making up the remaining 13 percent.

Table 2: Demographic and socio-economic information and baseline mathematics performance of participating schools

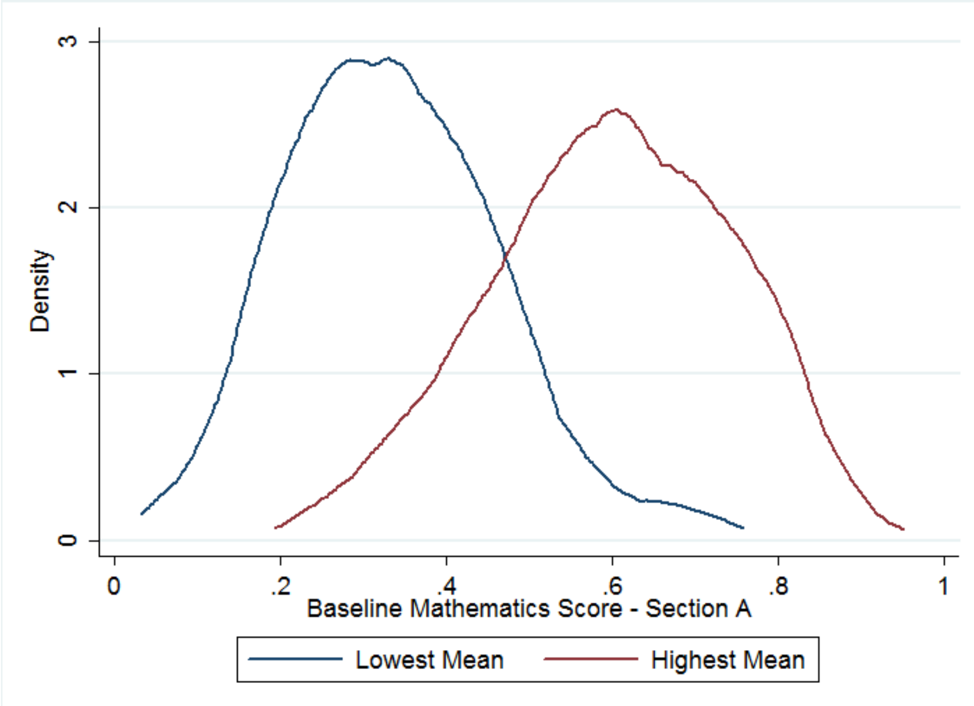
| School | Quintile | Home Language | | | Population Group | | | | Mean Baseline Mathematics Score |
|---|---|---|---|---|---|---|---|---|---|
| | | Afrikaans | English | isiXhosa | African | Asian/ Indian | Coloured | White | |
| 106 | 3 | 0.02 | 0.00 | 0.95 | 0.99 | 0.00 | 0.01 | 0.00 | 0.34 |
| 107 | 3 | 0.00 | 0.00 | 0.99 | 1.00 | 0.00 | 0.00 | 0.00 | 0.33 |
| 105 | 5 | 0.31 | 0.43 | 0.25 | 0.34 | 0 | 0.64 | 0.01 | 0.35 |
| 102 | 4 | 0.08 | 0.36 | 0.55 | 0.58 | 0 | 0.42 | 0 | 0.42 |
| 110 | 5 | 0.38 | 0.37 | 0.24 | 0.3 | 0.01 | 0.69 | 0.01 | 0.42 |
| 104 | 5 | 0.13 | 0.5 | 0.36 | 0.4 | 0 | 0.6 | 0.01 | 0.46 |
| 111 | 5 | 0.09 | 0.87 | 0.04 | 0.08 | 0.01 | 0.92 | 0 | 0.6 |
| 101 | 5 | 0.12 | 0.85 | 0.03 | 0.08 | 0.03 | 0.88 | 0.02 | 0.51 |
| 109 | 5 | 0.06 | 0.91 | 0.02 | 0.05 | 0.17 | 0.77 | 0.01 | 0.6 |
| **Total** | | **0.13** | **0.44** | **0.42** | **0.47** | **0.02** | **0.5** | **0.01** | **0.44** |

Source: Department of Basic Education's (2013) EMIS Ordinary Schools Masterlist and own data (2013)

It is evident in Table 2 that there is a great deal of variation in the demographics and income levels in the nine selected schools. Correspondingly, the difference in mathematics ability of learners entering Grade 8 is also pronounced, as can be seen in the last column of Table 2. The mean scores on the baseline basic numeracy test range from 33 percent to 60 percent. Figure 3 shows the distribution of baseline test scores at the school with the highest mean and the school which had the worst performing learners. The contrast is stark. Only 10 percent of learners at the worst performing school scored above 50 percent, whereas 75 percent of learners at the top school achieved a passing mark of 50 percent on the baseline numeracy test. The remaining 7 schools range between these two extremes.

[20] The exact amounts are not stated to maintain the anonymity of the schools.

Figure 3: The distribution of baseline Numeracy scores of all grade 8 learners at the schools with the highest and lowest means at baseline




Source: Own data collected in 2013

### Implications for the External Validity of the programme

It is important to note at this point that the school selection process has definite implications for the external validity of the evaluation. The Western Cape province is not representative of the country as a whole. The Western Cape is a relatively wealthy province with a high proportion of urban households (Finn, Leibbrandt & Woolard, 2009). It has also consistently been one of the two provinces achieving the highest pass rates in the annual National Senior Certificate Examinations[21]. Only the Gauteng province has traditionally achieved analogous pass rates to the Western Cape (Department of Basic Education, 2013). As the province is reasonably affluent, this meant that the resources and infrastructure required for such a technology intensive programme were already largely in place. It does however somewhat diminish the external validity of the experiment within South Africa, as schools in the Western Cape are not representative of South African schools as a whole.

Furthermore, logistical considerations resulted in a set of schools that were all in the Cape Metro Area, most of which were located less than 15 km from the city center. Thus the selected schools were chosen from a set of well managed urban schools, from poor areas in Cape Town who had the required

[21] For the first time in a number of years, neither Gauteng nor the Western Cape had the highest pass rates in January 2014. Instead they were ranked 3rd and 4th respectively, whilst the Free State and Northern Cape were place 1st and 2nd, sparking debate around learner dropout which questioned the validity of these figures (John, 2014)

technology to run an internet based intervention. It is not immediately obvious that similar effect sizes would be expected in other parts of the country, or even in more rural parts of the province. Nor is it clear whether a positive effect would be an understatement or an overstatement of the theoretical effect size that would be expected if this experiment was repeated on a representative sample of poor South African schools.

There were three distinguishing characteristics of selected schools that differentiated them from a representative sample within the Western Cape. The schools were urban, all were well managed with strong leaders and they had functioning computer laboratories with a working internet connection.

It is difficult to speculate on the direction of the bias as a result of excluding rural schools. In rural areas the availability of coaches may be lower and it is possible that language may be more of a barrier there. Internet connectivity is central to the programme and in urban areas connections are more stable and technical support is more easily available, which results in fewer disruptions to the programme due to loss of connectivity. However many urban learners travel far distances to school and as a result the safety of the learners travelling home later in the day is a huge concern. Community violence and transport strikes are one of the major reasons for learner absenteeism and eventually drop-out (Royker and Thompson, personal communication, 2014, January 14). Such concerns would be absent or only minor considerations in a rural context.

It is also not immediately obvious whether a sample of better managed and potentially slightly wealthier schools would bias the results upwards or downwards. Whilst the programme may potentially not be able to run as efficiently in a less well managed or financed school, this also means that the opportunities for improvement at such institutions are greater (Barrow, Markman & Rouse, 2009). However, bureaucracy or worse management in poorer schools could well affect the efficacy of the programme.

The fact that a school was required to have a functional computer laboratory does not have as great a repercussion for the external validity of the programme. This is because the Khanya project was an external shock, implemented by the Western Cape Education Department. Thus due to this government intervention the socio-economic status of leaners at a school is not necessarily an indicator of the digital resources available to the school. Consequently, the third characteristic of a functional computer laboratory would be correlated with better management insofar as the computers would need to be serviced and maintained, which is more likely to take place in a well-managed school. However this selection criterion is not necessarily a good gauge of other school or learner characteristics.

## Learner Application and Selection

The Numeric programme is an optional after-school activity. Additionally, due to resource constraints there were a limited number of places available on the programme. As a result, learners had to apply to be considered. From this group of applicants, learners were then randomly selected to be on the programme or in the control group, whilst the remainder was placed on a waiting list. This section describes this application and selection process in detail and discusses the implication of selection on the results. Lastly differences in characteristics between applicants and non-applicants are examined.

## Learner Application Process

Early in the school year both learners and parents were introduced to the programme. On the first day an information leaflet was sent home to all Grade 8 parents. A Numeric representative subsequently visited each of the schools during the first week of term where they told the learners about Numeric and introduced them to Khan Academy. After the special Grade 8 assembly interested learners could pick up application forms from the Numeric representative. Class teachers were given additional forms, were tasked with encouraging the learners to apply, and reminding them to return their forms. The completed application forms were collected a week after the presentation.

The Numeric management team divided the applications into three groups based on how comprehensively the application form had been completed: Accepted, maybe and rejected[22]. As a Numeric class contains a maximum of 20 learners. In order to for a school to qualify for a single class on the pilot, a minimum of 50 applications classified as "Accepted" or "Maybe" had to be received[23], whist this number was 90 if a school desired to run two classes. Out of the 50 applicants, 40 were selected to be in the evaluated sample. The combined group of learners in treatment and control is referred to as the evaluated sample to distinguish it from the full sample of all Grade 8 learners. This process of selecting the evaluated sample is described in the next subsection and the implication on the external validity of the results is discussed. The remaining 10 applicants were put on a waiting list. If one of the treatment learners dropped out of the programme, then a learner from the waiting list could then be offered their place on the programme. Table 3 shows the total number of learners in Grade 8 at each school[24]. Column 2 of Table 3 shows the number of applications received at each school. The next three columns show the how many of those applicants were accepted, the number that were assigned a maybe and the number of rejections. It is evident that the minimum requirement of 50 or 90 learner applicants was not strictly followed as four of the nine schools do not have the required number of applicants. Also at schools where the number of applicants was initially low, a second or even third round of applications was accepted[25].

---

[22] This stage of the selection process was done at the discretion of the programme manager. On the front page of the form the learners were asked to fill it in as completely as possible. The programme managers based their decisions on how much of the application form had been filled out as opposed to the quality of the answers. A form was 4 pages long, with an additional two pages that had to be signed by a legal guardian. By observation the rule of thumb seemed to be that a learner was rejected if about 50% or less of the form had been completed. Learners obtained 'Maybe' status if more than half had been filled in, but a fair a number of the questions had not been answered, whereas a learners was 'Accepted' if about three or less questions had been left out.

[23] There was one school in particular, 109, for whom an exception was made. Although they did not receive the minimum number of applications the implementing team in consultation with the school decided that there was still enough interest at the school to go ahead with the intervention.

[24] As many of the schools did not have reliable learner numbers at the beginning of the year, class registers from the end of the second term were used to calculate the learner numbers in the first column of Table 3.

[25] Table 3 shows that there were nine sample schools in total. Two only accepted a single round of application, three asked for a second round of applicants, whilst four also accepted a third and final round of applications. Of the total number of applications received, 73 percent were first round applications, 24 percent were from a second round and only 3 percent of the applications were from third round candidates.

Table 3: Table to show the number of learners, number of applicants and acceptance status at each of the nine schools

| School | Number of Learners in Grade 8 | Applications Received | Acceptance Status | | | Number of Application Rounds | Number of Learners in the Evaluated Sample |
|---|---|---|---|---|---|---|---|
| | | | No | Maybe | Yes | | |
| 101 | 197 | 43 | 0 | 4 | 39 | 2 | 40 |
| 102 | 161 | 45 | 0 | 9 | 36 | 1 | 40 |
| 104 | 183 | 51 | 2 | 5 | 44 | 3 | 40 |
| 105 | 272 | 101 | 2 | 11 | 88 | 2 | 80 |
| 106 | 266 | 83 | 2 | 9 | 72 | 2 | 80 |
| 107 | 285 | 65 | 3 | 6 | 56 | 3 | 40 |
| 109 | 230 | 34 | 2 | 1 | 31 | 3 | 32 |
| 110 | 170 | 53 | 1 | 4 | 48 | 1 | 40 |
| 111 | 244 | 93 | 4 | 13 | 76 | 3 | 80 |
| **Total** | **2008** | **568** | **16** | **62** | **490** | | **472** |

Source: Own data (2013)

Successful applicants received a letter of acceptance a few days prior to the first Numeric class at their school. The starting dates of the nine schools were staggered over five weeks, starting the week of the 11th February and ending the week of the 11th March. This represents a full month's difference between the starting date of the first and last school, and equates to a difference of about 12 hours of contact time between the groups that started last compared to a group that had started in mid-February[26].

### Randomisation

As the implementation of the programme took place in stages, the randomisation also happened in three rounds[27]. The same procedure was followed in all three rounds.

The 40 or 80 sample learners for each school were selected by first excluding learners with applications who had been assigned a 'Maybe' status, starting with learners who lived furthest away from the school. This was done as these learners were the ones most likely to struggle to arrange the required transport and for whom the safety concerns were greatest. If the group was still too large after having excluded all 'Maybe' type applications, as was the case in four schools (see Table 3), then again learners who lived furthest away from the school were excluded first. This method is unlikely to bias the result in any significant manner unless there is reason to believe that learners who travel further to get to school are likely to do systematically worse or better as a result of the programme[28]. Learners who attend a

[26] Although this potentially had some effect on dropout rates, a later starting date does not seem to correspond to a lower treatment effect.
[27] In the first round schools included 105 and 111. The second randomisation was done from schools 101, 102, 104, 109 and 110 whilst the third round concluded with schools 106 and 107.
[28] There is not much evidence to suggest that travel time to school affects Mathematics outcomes of *non-applicants*. This was verified by regressing travel time to school on endline Section A results controlling for School, Age and Gender. The estimate on travel time was 0.009 percentage points and highly insignificant, which showed that there was no significant impact of travel time to school on endline outcomes of non-applicants to the programme. Furthermore the distribution of travel time for non-applicants and applicants is not significantly different (p-value =

secondary school that is far from home usually do so for one of two reasons. Either the school they attend is close to the workplace of one of their parents or their parents pay for transport to send their child to what they perceive to be a better school. Thus it is conceivable that children with more motivated parents are excluded. Whilst this may mean that the child is also highly motivated to achieve, it could also mean that they were pressured into applying. Irrespective of the child's reason for applying, distance from school is a barrier to attendance. The most likely manner in which this approach may have resulted in a bias is that it may have slightly reduced the number of learners who dropped out of the programme.

From this sample group of 40 or 80 learners, individuals were randomly selected to be in treatment or control using an iterated stratification and balancing process. Control and treatment groups were stratified by school and above and below median baseline numeracy scores. Stratification by baseline numeracy was included to ensure that the distribution of baseline mathematics ability was similar between treatment and control.

Additionally the sample was balanced on gender, home computer access, distance from the school, above and below median baseline English literacy, above and below median baseline scores on the standard progressive matrices, above and below median typing scores and on 5 sub-categories of the baseline mathematics test (arithmetic, fractions, decimals, number patterns and word problems).

### Applicant Characteristics

Every Grade 8 learner at all participating schools was given the opportunity to apply to the programme. Moreover, the mathematics tests and learner surveys at baseline and endline were administered to all Grade 8s irrespective of their application status. This allows for a comparison between applicants and non-applicants on mathematics test scores and a variety of other demographic characteristics.

Table 4 presents some of the differences in observable characteristics of applicants. The first row shows that the average numeracy score of applicants is 2.2 percent higher than non-applicants. This difference is statistically significant, although it is unexpectedly small.

The gender of the learners plays an important role in the decision to apply which can be seen by the large number of female applicants in row 2 of Table 4. Whereas 60 percent of applicants are female, only 52 percent of learners in Grade 8 at the nine participating schools where female. The gender difference between the applicant and the non-applicant group is highly significant.

The age difference between applicants and non-applicants is also highly significant, with the average age of applicants 2.4 months lower than that of non-applicants (Table 4, Row 3). Part of the explanation for this difference is due to grade repetition. It is evident that a non-applicant is 15.5 percent more likely to have repeated a grade than an applicant. However it is likely that there are factors other than grade repetition, both observable and unobservable, that contribute to this difference in the mean ages.

---

0.589 using a Pearson $\chi^2$ test). Thus it seems likely that travel time also has very little effect on endline outcomes for *applicants*. Whilst it is possible to conceive that there might be a differential effect on outcomes by travel time between applicants and non-applicants, it seems unlikely that this would be large enough to be significant in such a small sample.

The number of Coloured and African applicants is proportional to the number of Coloured and African learners in the school population. There was however an under-representation of learners from Afrikaans households in the applicant pool. Conversely the proportion of applicants that were Xhosa speakers was higher than the proportion of Xhosa speakers in the non-applicant group. Even at the school with the highest proportion of learners with Afrikaans as their home language, Afrikaans speaking learners only made up 38 percent of the grade, in contrast, Xhosa learners were the majority group at three of the schools.

Of the learners that chose to apply to the programme, a smaller proportion report that they participate in a sports team compared to non-applicants. There were a couple of cases where sports practice clashed with Numeric classes which eventually resulted in two learners dropping out of the programme. This number is negligible. At endline the difference in rates of reported participation in sports is not significantly different between the two groups, and even at baseline there is no significant difference in participation in a cultural group by applicants versus non-applicants. This suggests that it is unlikely that learners were substituting afternoon sports for mathematics classes, which further suggests that the initial difference may have been due to activity preferences at primary school.

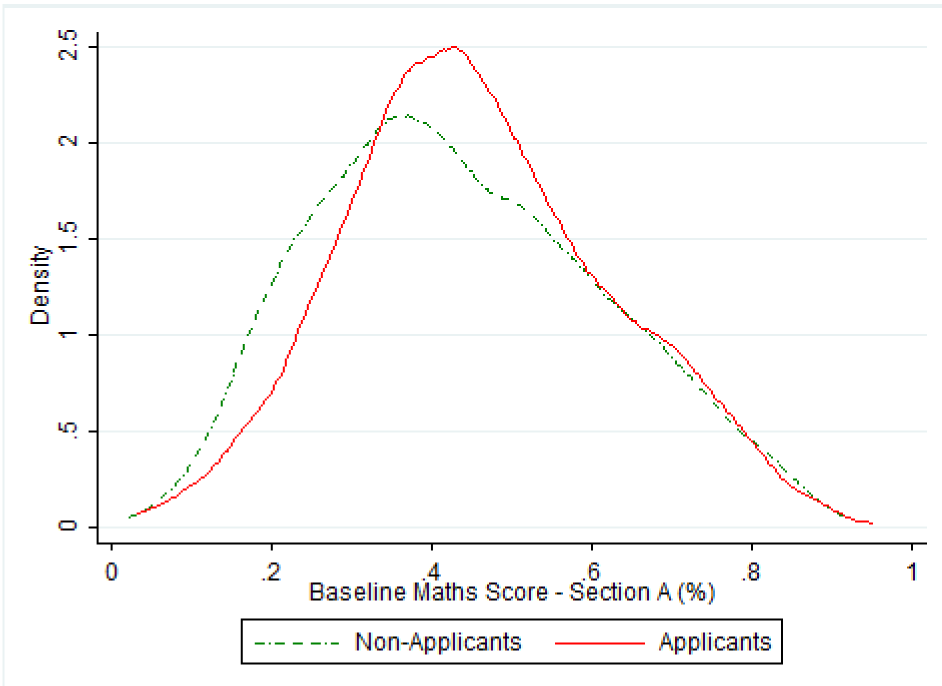
Table 4: Table showing baseline characteristics of applicants and non-applicants

| VARIABLES | Non-Applicant | Applicant | Difference *(Applicant-Non-Applicant)* |
|---|---|---|---|
| Baseline – Mathematics Section A Test Scores | 0.435<br>(0.179) | 0.457<br>(0.165) | 0.022**<br>(0.009) |
| Female | 0.485<br>(0.500) | 0.605<br>(0.489) | 0.120***<br>(0.024) |
| Age | 14.347<br>(0.992) | 14.148<br>(0.872) | -0.199***<br>(0.046) |
| African | 0.447<br>(0.497) | 0.452<br>(0.498) | 0.005<br>(0.025) |
| Asian, Indian or White | 0.033<br>(0.178) | 0.020<br>(0.139) | -0.013*<br>(0.008) |
| Coloured | 0.521<br>(0.500) | 0.529<br>(0.500) | 0.008<br>(0.025) |
| Home Language: Afrikaans | 0.137<br>(0.344) | 0.091<br>(0.288) | -0.047***<br>(0.015) |
| Home Language: English | 0.448<br>(0.497) | 0.473<br>(0.500) | 0.025<br>(0.025) |
| Home Language: IsiXhosa | 0.403<br>(0.491) | 0.425<br>(0.495) | 0.022<br>(0.025) |
| Home Language: Other Language | 0.012<br>(0.107) | 0.011<br>(0.103) | -0.001<br>(0.005) |
| Learner's Mother is Alive | 0.961<br>(0.193) | 0.959<br>(0.198) | -0.002<br>(0.01) |
| Learner's Mother has completed Matric | 0.571<br>(0.508) | 0.568<br>(0.504) | -0.003<br>(0.028) |
| Learner's Father has completed Matric | 0.601<br>(0.504) | 0.608<br>(0.499) | 0.007<br>(0.03) |
| Travel Time to School | 22.097<br>(16.514) | 21.839<br>(15.995) | -0.258<br>(0.822) |
| At least one computer at Home | 0.672<br>(0.470) | 0.607<br>(0.489) | -0.064**<br>(0.026) |
| Internet on a Computer at Home | 0.404<br>(0.491) | 0.358<br>(0.480) | -0.046*<br>(0.028) |
| Learner has repeated at least one Grade | 0.319<br>(0.466) | 0.277<br>(0.448) | -0.042*<br>(0.024) |
| Is a member of a sports team | 0.436<br>(0.496) | 0.376<br>(0.485) | -0.060**<br>(0.026) |
| In a member of a Cultural Group | 0.225<br>(0.418) | 0.215<br>(0.411) | -0.010<br>(0.022) |
| Like Mathematics A Lot | 0.389<br>(0.488) | 0.461<br>(0.499) | 0.072***<br>(0.026) |
| Satisfaction (Scale 1-7) | 5.333<br>(1.644) | 5.428<br>(1.560) | 0.095<br>(0.083) |
| Has a Phone | 0.813<br>(0.390) | 0.783<br>(0.412) | -0.030<br>(0.021) |
| Household members | 5.344<br>(2.187) | 5.297<br>(2.496) | -0.046<br>(0.123) |
| **Number of Observations** | **1543** | **569** | **2112** |

Notes: Column 1 and column 2 have standard deviations in parentheses. Standard errors are in parentheses in last column. Statistically significance differences are indicated by: *** p<0.01, ** p<0.05, * p<0.1

Whereas Table 4 shows the difference in means, Figure 4 shows the distributions of baseline Mathematics scores of applicants and non-applicants in their entirety. The graphs demonstrate that the higher average of applicants is not due to higher numbers of top end learners applying, but rather, that low scoring learners are much less likely to apply. The majority of applicants' baseline marks are between 30 and 60 percent showing that mid-range learners, in about the 3rd quintile, are relatively more likely to apply than top scoring learners.

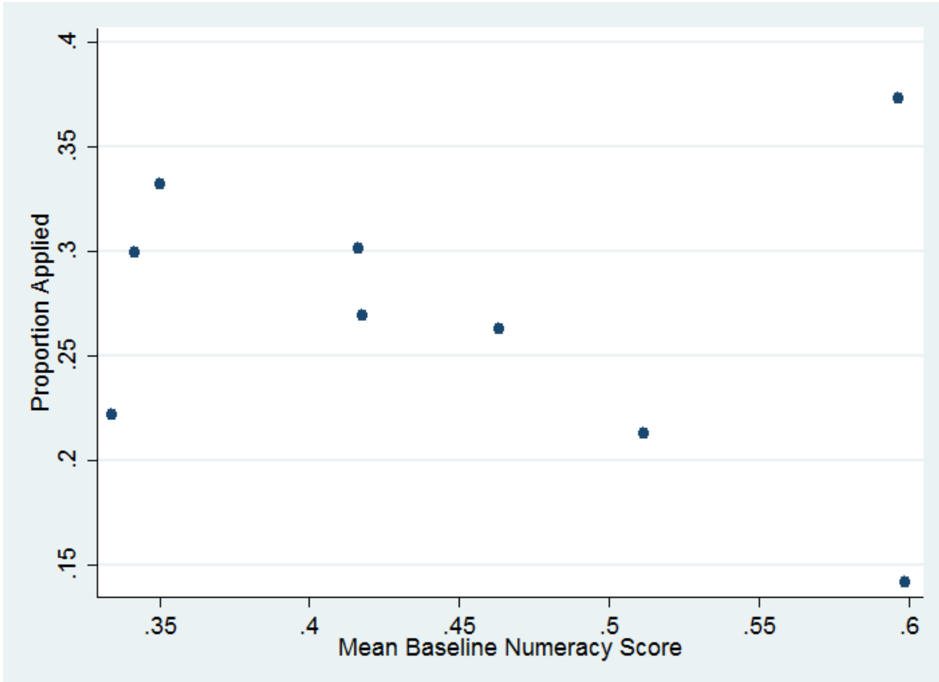
Figure 4: The distribution of baseline numeracy scores of applicants and non-applicants at baseline




Source: Own data collected in 2013

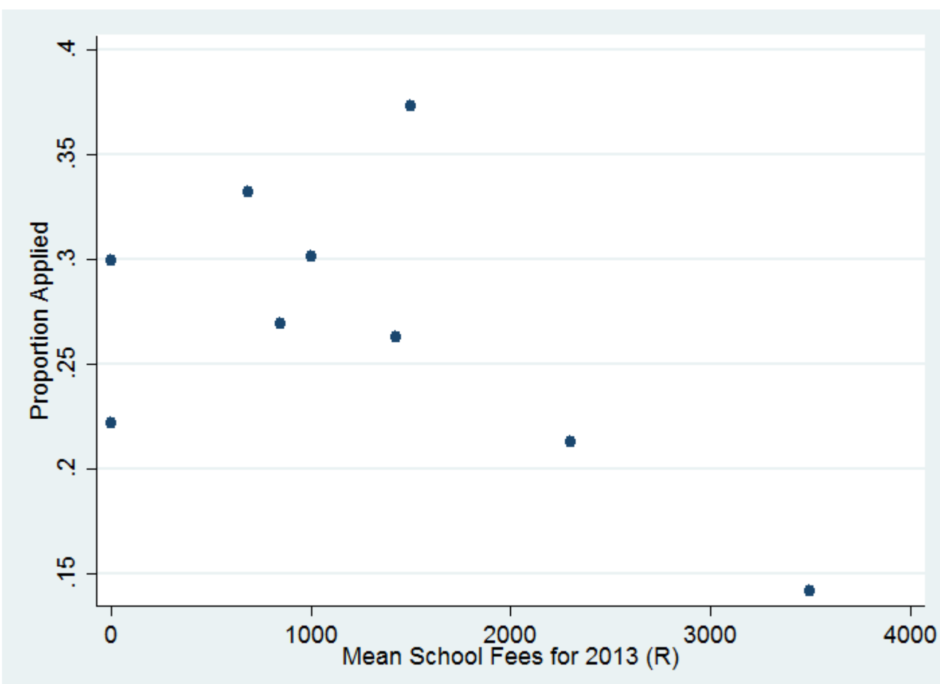
Figure 5 is a scatterplot which plots the applications received as a proportion of the total number of grade 8 learners, against the mean baseline mathematics score by school. A negative relationship is evident between the schools' mean baseline mathematics score and percentage of learners who applied to the programme. School 111 remains a notable exception to this trend, where despite high initial scores well over a third of the grade applied to the programme. Another less extreme outlier was school 107. The low number of applications received by this school was most likely due to upheaval at the school during the application period as the principal resigned during that same week at the beginning of the year. A very similar trend is observed in the scatter plot in Figure 6 where the average school fee is plotted against the proportion of learners that applied at each school. Here too the relationship is strictly negative. This suggests that there are two trends that are contributing to the distribution shown in Figure 4. Firstly, poorer learners and learners from underachieving schools are more likely to apply. Secondly, within a particular school, high achieving learners are more likely to apply. Thus the bulk of applicants is made up of high achievers from more disadvantaged schools. Table A1 in Appendix A presents the results from a probit selection model which confirms this trend.

Figure 5: Scatterplot showing the mean numeracy score against the proportion of applicants of each school




Source: Own data collected in 2013

Figure 6: Scatterplot showing the mean school fees against the proportion of applicants of each school




Source: Western Cape Education Department's (2013b) Find-A-School and own data collected in 2013

Another notable dissimilarity between applicants and non-applicants is the difference in response rates of the two groups highlighted in Table 5. The two main reasons for non-response include absenteeism on the day of the test or survey, and learners moving schools. Namely, if a learner did not start the year at their current school they would have missed the baseline testing period. There would also have been some sample attrition as some learners would have moved to a different school later in the year. In this case, they would not have been present for the endline tests and surveys. Any learner arriving at the school after the application period would automatically be counted as a non-applicant. Only learners who have either a baseline survey or a baseline mathematics test, and who would therefore have had the opportunity to apply, are included in Table 5. The difference in the number of missing observations between applicants and non-applicants for the midline mathematics test is 9.2 percentage points, whilst the difference is 9.8 percentage points for the endline mathematics test and 5.5 percentage points for the endline survey. All three differences are highly significant. At midline just more than 1 percent of applicants had moved schools or dropped out of school[29], whilst about 3.4 percent of non-applicants had done the same. This suggests that applicants come from homes or families that are less likely to move their children to a different school in the middle of a school year. However, even accounting for the proportion of learners that moved schools, absenteeism in non-applicants is also significantly higher at midline and endline that it is among applicants.

Table 5: Proportions of tests written and surveys completed by learners who were at the school in the first week of 2013

| VARIABLES | Non-Applicant | Applicant | Statistical Significance of Difference |
|---|---|---|---|
| Baseline Mathematics Tests Witten | 0.986 | 0.982 | - |
| Midline Mathematics Tests Witten | 0.862 | 0.955 | *** |
| Endline Mathematics Tests Witten | 0.862 | 0.955 | *** |
| Endline Surveys Completed | 0.885 | 0.943 | *** |
| **Number of Observations** | **1394** | **557** | |

Notes: Numbers in Columns 1 and 2 are means. Statistically significance differences are indicated by, *** p<0.01, ** p<0.05, * p<0.1, in the third column.

There is an underlying assumption that unobservable characteristics such as motivation and innate ability differ between applicants and non-applicants, as the act of applying reveals some information about learner characteristics such as initiative, energy and motivation. It has already been shown that on average applicants and non-applicants differed on a number of observable characteristics, including baseline mathematics scores, school, gender, age, race, home language, and participation in extra-mural activities. Thus it is almost certain that there would also be differences in unobservable characteristics between applicants and non-applicants. As a result differences between the applicant and non-applicant groups can only partially be accounted for using observable characteristics at baseline. For this reason, the programme effect that is calculated is an average effect and it is calculated using the difference in

[29] Given the very high enrolment at this age in the Western Cape, it is plausible though very unlikely that a large percentage of these dropouts can be attributed to learners who have stopped attending school entirely. It is far more probable that these are learners that have moved schools during the year.

outcomes at endline between the treatment and control groups. Here the average outcomes of the control group learners at endline provide a counterfactual (expected outcomes in the absence of the programme).

By comparing outcomes of treatment and control learners at endline the internal validity of the experiment is conserved, although it has some repercussions for the external validity of the results. Learners that did not apply to the programme in January may not have been as willing to try a new method, may have been less motivated to achieve academically or might just not have been as interested in school. Therefore, any programme effect is the average improvement expected, given that the learner had elected to participate in an after-school mathematics programme.

## 6. Data collection

### Test and Survey Schedule

Data collection took place during three distinct periods over the course of the year. Baseline tests and surveys were administered in January and February and, for one of the schools, a second round of tests was held in the first week of March[30]. All baseline testing was completed prior to randomisation. Midline testing was completed in June and endline surveys and tests were held in November.

A complete schedule of the tests and surveys is outlined in Table 6. Row 1 and 2 show that three standardised mathematics tests were conducted over the course of the year. At baseline these tests were written during a special Grade 8 testing session. However, in June and November the tests were incorporated into the school exam timetable and were written under exam conditions. These tests were administered to all Grade 8s, approximately 2 200 learners, in the nine participating schools. The demographics surveys were also administered to all Grade 8 learners at baseline and endline.

Additionally, all of the learners that applied to be on the programme (there were a total of 568 applicants) took part in another round of tests at baseline (Table 6, Row 4-6). This battery of tests included the Raven's Standard Progressive Matrices (SPM), an English literacy test, as well as a Computer Literacy and typing speed test. At endline this same set of tests was repeated, but was only written by learners in the evaluated sample.

The evaluated sample learners also completed a networking survey at midline, whilst at endline, learners in treatment and control also completed an additional incentivised test.

[30] The set of tests that were administered in March consisted of the Raven's Progressive Matrices, a test of English Literacy, and a Computer Literacy and typing test.

Table 6: Table indicating during which testing periods particular tests and surveys were conducted and which group of learners were tested or surveyed

| Test or Survey description | Baseline | Midpoint | Endline |
|---|---|---|---|
| 1. Standardised Mathematics Test – Section A (Numeracy) | A | A | A |
| 2. Standardised Mathematics Test – Section B (Core Grade 8 curriculum) | | A | A |
| 3. Demographics and Access to Technology Survey | A | | A |
| 4. Raven's Standard Progressive Matrices (SPM) | P | | E |
| 5. Computer Literacy and Typing Speed Test | P | | E |
| 6. English Literacy | P | | E |
| 7. Networks Survey | | E | |
| 8. Incentivised Mathematics Test | | | E |

Note: The letters indicate which group of learners wrote the test or completed the survey. A - All Grade 8's at participating schools, P – all applicants to the programme, E - Learner in the evaluated sample (control and treatment).

## Description of the Tests and Surveys

### Standardised Mathematics Tests

The most important measure is the standardised mathematics test which was written by all Grade 8 learners at participating schools. Three such tests were conducted over the course of the year, at baseline, midline and endline.

An Independent Assessment Committee was formed in October 2012. The sole responsibility of this committee was the creation of the three mathematics tests, taking particular care to ensure that they were set at the correct level of difficulty. The panel was made up of three members. Two were educators from public schools in the Western Cape, whilst the third was a consultant from IPA. The chief examiner was an experienced teacher who taught a Grade 8 class at a previously disadvantaged school. To get a more balanced perspective, the second teacher came from one of the top achieving schools in the area. The other two examiners moderated the test, and provided corrections and suggestions that were incorporated into the final draft.

The full test consisted of two sections: Section A and Section B. Section A covered primary school curriculum, whilst Section B tested concepts taught in Grade 7 and 8. Table 7 provides a rough breakdown of the grade levels that were tested in each section. It is clear from column 1 that 95 percent of questions in Section A tested competencies that the learners should have mastered by Grade 6. Section B was more difficult and two thirds of the questions in this section tested concepts that are only introduced in Grade 8. For this reason, the baseline test consisted of Section A only, whilst in June and November the learners wrote the full test. In order to simplify comparisons across time, the format of the test and question structures were identical in all three testing periods, although the numbers[31] were altered for each test[32].

Table 7: Table shows the level of difficulty of Section A and Section B by showing the percentage of questions that were asked in each section from various grade levels

| | Percentage of the Test (%) | |
|---|---|---|
| **Grade Level** | **Basic Numeracy Section A** | **Core Grade 8 Curriculum Section B** |
| 1-4 | 40 | - |
| 5 | 30 | - |
| 6 | 25 | 5 |
| 7 | 2.5 | 28 |
| 8 | 2.5 | 67 |
| **Total** | **100** | **100** |

### Learner Survey

A demographic survey was conducted with all learners at baseline and endline. This survey had four main sections: (1) demographic, household and parent information; (2) digital connectivity and computer fluency; (3) attitudes towards school; and (4) social cohesion. The first category asked questions relating to gender, age, race, area of residence, general information about their parents, parent's highest level of education as well as general household characteristics. The second section contained numerous questions which sought to ascertain what digital resources were available to the learner. A self-reported measure of computer fluency was also included. The computer fluency question listed eleven tasks, such as "Open a document" or "Send an e-mail". The respondent was then asked if they had ever used a computer to complete each of these tasks. This was included as an additional measure to try and ascertain whether the programme had an effect on computer use more generally. The section on school attitude asked questions about subject preferences as well as perception of own academic ability. The last section on social

[31] New numbers were used in each testing period, but the question was of the same type. For example Question 1.1 is 5×6 at baseline, 4×7 in the midline test and 6×7 at endline.

[32] This approach raised one concern. Teachers may have gotten a copy of the test at baseline and could have used it to provide the learners with a revision session just prior to the test. Although such a situation is plausible, it is unlikely that teachers would have chosen to use teaching time, right before their school exams, to prepare learners for a test which they viewed as less consequential than their school exams. Additionally, control and treatment learners would both have received the additional help so the comparison across these two groups would still be valid. This raises the more general question of the impact of teacher behaviour or attitude on the results. Fortunately, Mathematics teachers in most cases were not aware of which learners had applied to the programme as this had been organised by the class teacher and the liaison teacher at the school. Most teachers would therefore be unaware of which learners were in control and treatment. Teachers may however have become aware that certain learners were attending the programme later in the year. However it is not obvious how a teacher's actions might have changed if they were aware that a learner was attending the Numeric programme. It is possible that they may have spent more time with these learners to further enhance their performance as they were already performing well. Otherwise they may also have diverted attention to the rest of the class if they felt that as a result of the Numeric programme the treatment learners needed less of their help.

cohesion asked the learners about their general sense of wellbeing, whether they participate in non-academic activities and what support they felt was available to them at school.

### Raven's Standard Progressive Matrices (SPM)

The Raven's Standard Progressive Matrices aim to test a person's ability to make new insights and discern relationships (Raven, Raven & Court, 2003). The test gradually increases in difficulty. It is a multiple choice test that presents the learner with a panel which contains a pattern with a missing piece. An example of two such questions are presented in Appendix D. The test taker must then choose the tile that will correctly complete the pattern in the main panel from a set of possible tiles. The SPM were administered at baseline in order to control for general cognitive development among the learners, as it is possible that learners with higher overall cognitive development could have performed better on the programme.

The Raven's SPM were also included as an outcome measure. The Raven's SPM have been used repeatedly as a test of general cognitive development in technology in education impact evaluations. Two such impact assessments in Romania and Peru provided learners with home computers or individual notebooks. In both studies, there was no significant impact on reading and mathematics scores at endline, despite an increase in computer use and ownership. However there was a significant increase in cognitive development as measured by the learners' Raven's scores (Cristia et. al., 2012, Malamud & Pop-Eleches, 2010). On the Numeric programme it is hypothesized that not only the process of learning to use a computer but also the material itself may have had some impact on general cognitive skill, thus the test was also administered at endline.

### English Literacy

The English literacy test was a short 15 minute multiple choice test that consisted of three sections. The baseline test is shown in Appendix D. The first section consisted of a brief listening exercise. The second part tested vocabulary, whilst the third section contained a comprehension. The English literacy test was administered at baseline so that it could be used as a balancing variable in the random selection of learners onto the programme. On Khan Academy the videos and exercises are all in English. Therefore it was expected that a learner with a better grasp of the language may adapt to the new Khan Academy platform more quickly and in so doing may have gained more from the programme. However an improvement in English ability is also a possible outcome of the programme. Particularly, if a learner's home language was not English, then their English may have improved as a result of the programme. As such, English literacy was re-tested at endline.

### Computer Literacy and Typing Test

It is expected that consistent interaction with a computer over the course of the school year would have resulted in an improvement in computer literacy for learners on the programme. It is therefore an important secondary outcome of the programme and thus the test was administered at endline.

The test consisted of three sections: (1) General computer knowledge, (2) a web exposure questionnaire and (3) a typing test. The general computer literacy section included basic questions on hardware and software, as well as desktop navigation and web related questions. In the web exposure survey the leaners were asked if they had ever seen the logos or heard of 6 well know websites[33]. The full test that was used is included in Appendix D. For the last section the learners completed a one minute online typing test[34]. The typing test was run twice and both results recorded. The Computer Literacy test was also conducted at baseline as it is plausible that learners, who were already comfortable working on a computer, may have started to benefit from a computer based programme more quickly and may therefore have derived more value from the programme as a whole. For this reason typing speed was included as a balancing variable in the randomisation.

### Networking Survey

The Numeric classes took place after school, which meant that learners in control and treatment were still receiving the same mathematics instruction at school for the majority of the week. This suggests that a number of control and treatment learners would have frequently been in contact in situations where mathematics was taught and practiced. It is natural, therefore, to suppose that directly, through explanation, or indirectly, by introducing their friends to Khan Academy, learners in the control group may also have been affected by the programme.

In order to account for this possibility, a survey was conducted at midline with learners in the evaluated sample, to get a sense of which other learners from the control and treatment groups were within their peer group and family. The learners were asked to name their 5 best friends in Grade 8, and their neighbour(s) in their mathematics class at school. They were then asked to name any study partners they might have, whilst the last question asked if they had any siblings, cousins or if anyone else who lived in the same household was also in their grade at school.

### Incentivised Mathematics Test

Segal (2012) showed that in situations where the outcome of a test is immaterial to the test taker, the results often more accurately reflect motivation as opposed to cognitive ability. At baseline the desire to be accepted to the programme could have acted as motivating factor for most learners. However, at endline there are few external motivating factors. Learners at a number of the schools were aware that these tests were associated with the Khan programme, which may have resulted in differential motivation between control and treatment. In particular, learners in the control group, that were not selected to be on programme, may have been severely discouraged due to the initial rejection at the beginning of the year. Conversely, learners in treatment may have tried harder on the test as they knew and probably respected the staff members from the organization and would therefore have wanted to be seen to be performing well. This combination could translate into an artificially high treatment effect. In order to test for such a bias, due to a disparity in motivation between treatment and control, a second, shorter, incentivised

[33] The 6 websites were: Google, Wikipedia, Facebook, Skype, YouTube and Gumtree
[34] The test can be found at: http://10fastfingers.com/typing-test/english.

mathematics test was conducted in a closely controlled environment. This test was 15 minutes long and consisted of 22 questions. These questions were a subset of the 44 questions in Section A. Learners therefore had slightly less time to complete the questions than they would have had for the normal mathematics test. This test was conducted is tightly controlled environment with the examiner: learner ratio never exceeding 1:20.

## 7. Dataset

The complete dataset contains observations on 2112 Grade 8 learners, originating from the nine schools. Any leaner that wrote at least one of the tests or surveys was assigned a unique number and was added to the dataset. Therefore, this total includes all Grade 8 learners from all nine schools, even those that moved schools during the year. It comprises of the results of the three mathematics tests, as well as the responses on the baseline and endline learner surveys.

As the primary focus of this paper is to determine the magnitude of the treatment effect on mathematics outcomes, the main dataset that is utilised is a subset of the full dataset. This dataset contains only the 472 learners who were selected to be in the treatment or the control group. This subset is referred to as the evaluated sample, and consists of 236 individuals who comprise the control group and 236 learners in treatment. Table 8 shows a breakdown of the number of learners in the evaluated sample at each of the nine schools. Noticeably, three schools had enough applicants to be able to form two classes of 20 learners each[35], whilst the remainder only had enough applicants for a single class. At school 109 only 32 successful applications were received, however the programme managers decided to proceed with the intervention at the school with a reduced class size.

Table 8: Number of learners in the evaluated sample and Numeric classes by school

| School | Number of Learners in Evaluated Sample | Number of Classes |
|---|---|---|
| 101 | 40 | 1 |
| 102 | 40 | 1 |
| 104 | 40 | 1 |
| 105 | 80 | 2 |
| 106 | 80 | 2 |
| 107 | 40 | 1 |
| 109 | 32 | 1 |
| 110 | 40 | 1 |
| 111 | 80 | 2 |
| **Total** | **472** | **12** |

[35] These schools tended to be larger; hence the greater number of applicants. In schools with two classes, the classes were streamed.

## Missing Observations and Learner Attrition

Table 9 shows the proportion of observations obtained from the evaluated sample for each test and survey. Rows 1 to 3 show that there was a fairly high rate of test completion on the main measure, the standardised mathematics test. Of the sample of 472 learners, 97.5 percent of the sample wrote the baseline mathematics test, whilst 96.0 percent wrote the endline mathematics test. At endline there were 19 missing observations. Nine of these were from the control group and ten from the treatment group, a difference which is highly insignificant.

The proportion of completed endline surveys was slightly lower compared to the proportion of mathematics test written at endline, with a coverage of 94.9 percent. There was no difference in response rates between control and treatment on the endline survey. The number of missing observations for the baseline survey was very small at only 2.54 percent. There was also a significant difference in the number of baseline surveys collected from the control and treatment groups, with only 3 missing observations from the control group and 9 observations missing from treatment. Conversely at endline there were exactly 12 observations missing from treatment and 12 missing from control.

Table 9: Table to show the proportion of the evaluated sample that wrote the tests and completed surveys for each testing period by treatment and control

| | Control | Treatment | Total | Significant Difference *(Treatment-Control)* |
|---|---|---|---|---|
| **Baseline - Mathematics Test** | 0.979 | 0.975 | 0.977 | - |
| **Midline - Mathematics Test** | 0.919 | 0.941 | 0.930 | - |
| **Endline - Mathematics Test** | 0.962 | 0.958 | 0.960 | - |
| **Baseline - Survey** | 0.987 | 0.962 | 0.975 | * |
| **Endline - Survey** | 0.949 | 0.949 | 0.949 | - |
| **Baseline - English** | 0.758 | 0.792 | 0.775 | - |
| **Baseline – Raven's** | 0.792 | 0.847 | 0.820 | - |
| **Baseline - Computers** | 0.758 | 0.792 | 0.775 | - |
| **Midline - Networking** | 0.898 | 0.843 | 0.871 | * |
| **Endline - Incentivised** | 0.869 | 0.886 | 0.877 | - |

Notes: Numbers in Columns 1, 2 and 3 are means. Statistically significance differences are indicated by, *** p<0.01, ** p<0.05, * p<0.1, in the 4th column.

Information on only 82 percent of sample learners is available for the baseline Raven's SPM test. The proportions are even lower for the baseline English and Computer tests. For these two tests, there is information on only 77.5 percent of evaluated sample learners. The tests were written after-school, which meant that the number of absentees was higher than for tests held during school time. The numbers here are also lower because of the loss of a box of the test scripts for one of the schools[36] prior to the data being captured, resulting in the automatic loss of 8.5 percent of the observations.

[36] The school for which the baseline English and typing tests were lost was school 102.

A total of 87.1 percent of evaluated sample learners were interviewed for the networking survey. Overall, 6.5 percent more control than treatment learners were interviewed. The number of missing observations for the incentivised mathematics test is much higher than for the standardised mathematics test. Only 413 learners completed the incentivised test, of which 205 were control learners and 208 came from the treatment group. Here too, there is no significant difference in the missing observations between the control and treatment groups. Even at school level, missing observations were evenly split between control and treatment, and in none of the schools does this difference exceed two observations. There was however some correlation between the learners who did not write the incentivised tests and those who had dropped out of the Numeric programme by November. Looking only at learners in treatment, whilst 95.3 percent of learners who were still on the programme wrote the incentivised test, only 71.6 percent of learners who had dropped out wrote. The difference is highly significant. The 414 learners that wrote the incentivised test also scored 7.5 percentage points higher on the baseline test, suggesting that weak learners were either more likely to be absent or actively chose to avoid writing the incentivised test. Additionally, the average baseline score of treatment learners who did not write the test was lower than the average baseline score of untested control learners, although this difference is insignificant (p=0.217), it suggests the possibility that factors contributing to attrition were not consistent between control and treatment groups for the incentivised test.

## 8. Summary Statistics

The credibility of the results in a randomized controlled trial depends critically on the random selection into treatment and control at baseline. Thus, the most crucial role of the baseline summary statistics is to show that the random selection into treatment and control was conducted successfully, by confirming that there are no significant differences between the control and treatment group prior to the start of the programme.

Table 10 shows the means and the differences in the means of the control and treatment groups. Only one of the variables in question was found to be significantly different in control and treatment, using a basic t-test with unequal variance. The variable is a combination of all learners who are Asian, Indian or White. There are only 10 individuals or 2.15 percent of sample learners who fall into this category, thus the small size of this group is the probable cause of the imbalance between treatment and control.

The average baseline mathematics and English Literacy scores of the control and treatment groups are the same, although the standard deviation is somewhat higher in the treatment group. Both the Raven's and typing scores are marginally higher in the control group, but insignificantly so. Fifty-nine percent of the sample is female and the average age was 14 years in January 2013, with a standard deviation of under a year, which is the expected age of learners in Grade 8 given a school starting age of seven.

The majority of learners are either African or Coloured, with African learners comprising 45 percent of the sample whilst Coloured learners make up 52 percent. Asian, Indian and White learners account for the difference. IsiXhosa is the most commonly spoken home language at 46 percent, whilst 41 percent of the learners speak English at home and 11.7 percent are Afrikaans speaking.

It is interesting to note that a substantial proportion of the sample, 62.9 percent, has access to at least one computer at home. Internet access is less common. Only 37 percent of learners report having internet access on a computer at home. This shows that exposure to technology is already reasonably high prior to the start of the programme. There are two possible ways in which the effect size may be affected as a result of internet access. The first is that any effect of the programme may be greater than in areas with higher connectivity as leaners may opt to also use Khan Academy resources at home increasing the time spent on practicing mathematics. However, it also increases the probability that learners who are not on the programme, who may hear about Khan Academy from their friends, then proceed to make use of the website. Conversely, such spillover would result in an underestimate of the effect size.

Table 10: Baseline learner characteristics in the control and treatment groups

| VARIABLES | Control | Treatment | Difference *Treatment-Control* |
|---|---|---|---|
| Baseline-Mathematics Test Scores (Section A) | 0.464 | 0.463 | -0.001 |
| | (0.165) | (0.171) | (0.016) |
| Baseline -English Literacy Score | 0.558 | 0.563 | 0.243 |
| | (0.197) | (0.186) | (0.020) |
| Baseline – Raven's Score | 0.593 | 0.576 | 0.017 |
| | (0.169) | (0.175) | (0.017) |
| Baseline -Typing (Keys per minute) | 0.608 | 0.602 | -0.006 |
| | (0.330) | (0.336) | (3.808) |
| Female | 0.555 | 0.623 | 0.068 |
| | (0.498) | (0.486) | (0.045) |
| Age | 14.106 | 14.177 | 0.07 |
| | (0.753) | (0.946) | (0.079) |
| African | 0.391 | 0.438 | 0.047 |
| | (0.489) | (0.497) | (0.046) |
| Asian, Indian or White | 0.034 | 0.009 | -0.026* |
| | (0.182) | (0.092) | (0.013) |
| Coloured | 0.575 | 0.554 | -0.021 |
| | (0.495) | (0.498) | (0.046) |
| Home Language: Afrikaans | 0.086 | 0.103 | 0.017 |
| | (0.281) | (0.304) | (0.027) |
| Home Language: English | 0.532 | 0.479 | -0.053 |
| | (0.500) | (0.501) | (0.046) |
| Home Language: IsiXhosa | 0.373 | 0.402 | 0.028 |
| | (0.485) | (0.491) | (0.045) |
| Home Language: Other Language | 0.009 | 0.017 | 0.009 |
| | (0.092) | (0.130) | (0.010) |
| Learner's Mother is Alive | 0.953 | 0.965 | 0.012 |
| | (0.213) | (0.185) | (0.019) |
| Learner's Mother has completed Matric | 0.585 | 0.545 | -0.040 |
| | (0.504) | (0.509) | (0.051) |
| Learner's Father is Alive | 0.860 | 0.859 | -0.001 |
| | (0.347) | (0.349) | (0.033) |
| Learner's Father has completed Matric | 0.590 | 0.653 | 0.063 |
| | (0.505) | (0.490) | (0.054) |
| Number of Household Members | 5.350 | 5.413 | 0.063 |
| | (2.164) | (3.056) | (0.248) |
| Travel Time to School | 19.382 | 21.231 | 1.849 |
| | (12.735) | (15.523) | (1.125) |
| At least one computer at Home | 0.614 | 0.640 | 0.026 |
| | (0.488) | (0.481) | (0.047) |
| Internet on a Computer at Home | 0.335 | 0.410 | 0.075 |
| | (0.473) | (0.493) | (0.051) |
| Learner has repeated at least one Grade | 0.257 | 0.308 | 0.051 |
| | (0.438) | (0.463) | (0.042) |
| Learner Likes Mathematics a Lot | 0.453 | 0.467 | 0.014 |
| | (0.499) | (0.500) | (0.047) |
| Is a member of a sports team | 0.412 | 0.355 | -0.057 |
| | (0.493) | (0.479) | (0.046) |
| In a member of a Cultural Group | 0.187 | 0.231 | 0.044 |
| | (0.391) | (0.422) | (0.039) |
| **Number of Observations** | **236** | **236** | **472** |

Notes: Column 1 and column 2 have standard deviation in parentheses. Standard errors are in parentheses in last column. Statistically significance differences are indicated by, *** p<0.01, ** p<0.05, * p<0.1, in the third column

## 9. Analysis of Endline Outcomes: The treatment effect

The main objective of the Numeric after-school programme was to improve the mathematics outcomes of participants. As such, the primary focus of the analysis was on determining the overall effect of the programme on the mathematics scores of learners, and then checking the consistency and robustness of the results.

First, a basic intention-to-treat (ITT) approach was used to calculate the effect of treatment. Under ITT learners that were selected to be in treatment, regardless of actual treatment status, counted as treated learners. The same applies for the control group. Selection into treatment or control, as opposed to attendance, is used as the criterion as it is assumed that drop-out from the programme is non-random. If this non-random group of treatment learners that dropped out is excluded from the endline average, it may result in an inflated treatment effect, as the results from the corresponding group of potential drop-outs in the control group are nevertheless included in the average score of the control group. In this experiment, there was a high correlation between selection into treatment and actual treatment status. None of the control learners were ever treated. In comparison, 97 percent of learners in treatment attended at least one class, whilst 71 percent of learners received a full year of treatment.

The effect size was determined using baseline and endline mathematics test scores and can be calculated using Glass's delta ($\delta$). Glass's delta is defined as the difference in the means of the treatment and control groups divided by the standard deviation of the control group (Glass, 1977).

$$\delta = \frac{\mu_{treatment} - \mu_{control}}{\sigma_{control}}$$

The measure $\delta$ presents the effect size in units of standard deviations (s.d.). This allows for greater comparability across different tests and studies, by providing a measure of the magnitude of the difference in outcomes between treatment and control.

In this thesis, only baseline and endline scores are used in the calculation of effect size. The midline tests are used to determine learning trajectories, but are not included in the main analysis. An approach using only baseline and endline measures is likely to produce results with a large error if the outcomes are noisy. Here, however, the endline test results on Section A and B are highly correlated with other measures of mathematics aptitude[37]. This suggests that they provide a fairly accurate measure of overall mathematics ability at endline, allowing for slight differences in performance of the learner on the days of the test, different testing conditions and differences between markers[38]. Due to the relatively small error in the outcome measure and high correlation between baseline and endline mathematics scores, this approach is justified (McKenzie, 2012).

[37] The correlation between the incentivised test marks and the endline test Section A and Section B marks are high at 0.878 and 0.819 respectively. Likewise the correlation between second term school marks and endline Section A results is also high at 0.606, whilst it is 0.678 for endline Section B scores. This last result is unexpectedly high, considering that this does not account for differences in reporting between schools and teachers.

[38] The error rate due to marking and capturing was estimated to be 1.0 percent.

This section comprises of four parts and presents thirteen results. The first subsection examines, in detail, the trends and distributions in average mathematics outcomes. The effect size of the Numeric programme is also calculated. Six major results are discussed in this subsection. Namely, (1) that there is an observable programme effect at midline which becomes highly significant at endline. The results on the endline mathematics test are also found to be significantly higher for treatment learners in (2) basic numeracy and for (3) core Grade 8 curriculum. Additionally, these (4) improvements are evident for learners throughout the distribution and (5) treatment learners score significantly higher on all of the basic numeracy areas that were tested. Furthermore, (6) learners who were selected to be in treatment are significantly more likely to state that they like mathematics than control group learners at endline.

The second part takes a closer look at attendance and drop-out rates and how these are related to the programme effect. This subsection contains two further results. Firstly, (7) Numeric drop-out rates are negatively related to baseline mathematics scores. Similarly, (8) there is a strong positive relationship between time spent on Khan Academy and an improvement in learners' scores on both basic numeracy and core Grade 8 curriculum.

The third part examines whether the treatment effect varies between different subgroups of the evaluated sample. Four main findings are discussed, starting with the result that (9) neither mathematics ability nor (10) attitude towards mathematics at baseline had a differential effect on treatment. This is augmented by the finding that (11) gender, race and home language all had no differential effects on treatment. Similarly, (12) neither general cognitive ability, English ability nor typing speed at baseline had a significant heterogeneous treatment effect on basic numeracy results. However, (13) treatment learners who are in the middle and top third of the baseline English distribution improved significantly more on core Grade 8 curriculum compared to treatment learners in the bottom third

The last subsection contains a number of robustness checks. Spillover as a result of the programme through peer groups and in-class interactions is found to be insignificant, whilst differential motivation in test taking between treatment and control learners is shown to be positive however its effects on the treatment effect is highly insignificant. Finally, the last table demonstrates that there are no large outliers, at the school level, that are inflating the treatment effect.

## Mathematics Test Results

### Trends in the Mathematics Test Scores

***Result 1: A continuous improvement in average mathematics scores of treatment learners, relative to control learners, is observed over the course of the year.***

Table 11 reports the average scores on the mathematics tests over the three testing periods. Columns 1 and 2 show the means of the control and treatment groups, whilst Column 3 reports the difference between treatment and control. Row 1 shows that at baseline, there is no significant difference in the means, with mean scores of 46.4 and 46.3 percent respectively for the control and treatment groups. By midline, differences start to emerge; the treatment group outperforms the control group by about 3.5 percentage points on the basic numeracy test (Section A), significant at a 5 percent level, whereas the treatment group scores an insignificant 1.3 percentage points higher on core Grade 8 curriculum material (Section B). The treatment group scores significantly higher in both sections at endline, with a difference of 6.0 percentage points (0.321 s.d.) in Section A and an average of 4.1 percentage points (0.246 s.d.) in Section B, significant at a 1and 5 percent level of significance.

Since the three basic numeracy tests were of a comparable standard, the Section A results can be compared across time. Column 1 of Table 11, shows that the average score of the control group increased by 4.8 percentage points from 46.4 percent at baseline to 51.2 percent at midline. There is then a decrease of 2.4 percentage points from midline to endline. However, there is still an improvement of 2.4 percentage points from baseline to endline[39]. Even in the absence of any outside intervention, secondary schools appear to be teaching and consolidating primary school competencies. This would be consistent with the hypothesis that schools revise primary school material, tested in Section A, early in the Grade 8 year to allow some of the learners to catch up, thereby moderating for inevitable differences in instruction between primary schools. Between midline and endline, the mean Section B scores of the control group increased by 2.3 percentage points. This suggests that the core grade 8 curriculum, tested in Section B, is the main material covered in school mathematics classes is the third and fourth terms.

Comparatively, in the treatment group, there is an even larger increase of 8.5 percentage points in basic numeracy scores from baseline to midline, but in the treatment group there is no corresponding drop in Section A scores between midline and endline. Instead, the mean remains at 55 percent. As was the case in the control group, Section B scores of treatment learners increase from midline to endline. Here, again, the difference in the treatment group is greater with the mean score increasing by 5.2 percentage points.

[39] Looking only at the control group it is evident that the Section A results increase significantly from baseline to midline and then drop by about 2.5 percent between midline and endline. There are a number of reasons that may have contributed to this pattern. Section A does not test Grade 8 material, so whilst the curriculum provides for revision at the beginning of the year, learners would not normally encounter these types of problems in their school Maths classes in the third and fourth terms, as a result they may forget some of the methods. The marking at endline was also stricter for the longer questions, which potentially reduced marks by a further percent or two. Also, learners were only given feedback on their performance on these tests at the very end of the year and this lack of feedback may have resulted in reduced effort on the endline test.

Table 11: Mathematics test scores at baseline, midline and endline for treatment and control

| VARIABLES | Control | Treatment | Difference *Treatment-Control* |
|---|---|---|---|
| Baseline - Mathematics Section A | 0.464 | 0.463 | -0.001 |
| | (0.165) | (0.171) | (0.016) |
| Midline - Mathematics Section A | 0.512 | 0.548 | 0.035** |
| | (0.172) | (0.191) | (0.017) |
| Midline - Mathematics Section B | 0.241 | 0.254 | 0.013 |
| | (0.143) | (0.163) | (0.015) |
| Endline - Mathematics Section A | 0.488 | 0.548 | 0.060*** |
| | (0.187) | (0.196) | (0.018) |
| Endline - Mathematics Section B | 0.264 | 0.305 | 0.041** |
| | (0.167) | (0.194) | (0.017) |

Notes: Column 1 and column 2 have standard deviations in parentheses. Standard errors are in parentheses in last column. Statistical significance is indicated by *** p<0.01, ** p<0.05, * p<0.1, in the third column.

### Multivariate Analysis

The main analysis investigates the effect of selection into the programme on the mathematics scores of learners at endline. Here the equation of interest is:

$$Y_{isEL} = \propto + \beta T_{is} + \gamma Y_{isBL} + \pi Y_{isBL}^2 + \theta \boldsymbol{S'}_s + \rho \boldsymbol{Z'}_{isBL} + \delta \boldsymbol{X'}_{is} + \varepsilon_{is} \qquad (1)$$

The dependent variable, $Y_{isEL}$, is the mathematics test score of learner $i$ at school $s$ at endline ($EL$), whist $Y_{isBL}$ is their baseline ($BL$) mathematics score. A second order term, $\pi Y_{isBL}^2$, was included to allow for a non-linear relationship between baseline and endline mathematics scores. $T_{is}$ is a dummy variable which indicates whether the learner was selected to be in treatment and $\boldsymbol{S'}_s$ is a vector of school dummies.

$\boldsymbol{Z'}$ is a matix of tests scores at baseline which include English literacy, the Raven's Standard Progressive Matrices results as well as a typing speed test. As the number of missing variables at baseline was reasonably high, and the missing values were not missing completely at random, a missing indicator method approach was adopted (White & Thompson, 2005)[40]. Namely, the missing values on all covariates were set to zero. Furthermore, a dummy variable was included for each covariate which was equal to 1 if the covariate had missing values and zero for all non-missing observations.

$\boldsymbol{X'}$ is a matrix of learner characteristics. These include demographic details such as age, gender and home language. Variables measuring attitudes towards school, access to technology, parent's education and some household related statistics were also included in $\boldsymbol{X'}$. Lastly, $\varepsilon_{is}$ is the error term for individual $i$.

[40] Both missing observations and item non-response were not random. More saliently, they were not uncorrelated with the outcome variable, namely endline Mathematics results. This claim is discussed in detail in Appendix B.

***Result 2: Treatment learners score significantly higher on basic numeracy at endline than control learners.***

Table 12 presents the main regression results. In the first four columns the dependent variable is the Section A mathematics test score at endline, whilst Columns 5 to 8 have the endline Section B results as the regressand. The first column shows a treatment effect of 6.0 percentage points with no controls. In Column 2 a linear and second order term of baseline mathematics results were included along with a set of school dummies to control for school level differences. This improves the fit of the model, but does not change the significance of the treatment effect whereas the size decreases marginally to 5.9 percentage points. It is also evident that baseline Section A scores are a good predictor of endline Section A scores. Although provision is made for a non-linear relationship between baseline and endline Section A mathematics results the relationship is found to be linear, with a highly significant estimate on the linear term, whilst the estimate on the second order term is practically small and statistically insignificant.

In Column 3 there is no material change to the size or significance of the treatment effect when learner demographic controls and the baseline test results on the Raven's, English Literacy and typing tests are included, nor is there a change to the treatment effect when the full set of controls are included in Column 4. As there appears to be no significant change due to the addition of the controls in Columns 4 and 8, the specification which is used repeatedly for the remainder of the thesis is the simpler functional form presented in Columns 3 and 7, with minor alterations where necessary.

The endline basic numeracy results consistently show a difference of between 5.7 and 6 percentage points of treatment learners relative to control group learners. Given a standard deviation of 18.7 percentage points in the basic numeracy scores at endline, this equates to an improvement of 0.321 of a standard deviation in basic numeracy. Alternatively, an average learner who was selected to be in treatment is able to answer 12 percent more of the paper correctly than a learner from the control group, holding individual, school and household characteristics as well as baseline test scores constant. This result is large relative to the effect sizes that are typically observed in education interventions[41]. However, it is of a similar magnitude to the highly successful primary school computer assisted learning programme run in Vadodara, India. This programme also had a mathematics focus and the treated learners improved by 0.35 standard deviations more than control learners in the first year of the programme (Banerjee et. al., 2007).

[41] Kremer, Brannen & Glennerster (2013) present a summary of the results of 28 different randomised controlled trails on primary school education interventions. The interventions differ widely both in locality and type of intervention, but out of the 28, only 7 resulted in an improvement in tests scores of more than 0.2 standard deviations.

***Result 3: Treatment learners score significantly higher than control learners on core Grade 8 curriculum at endline.***

Section B tests Grade 8 curriculum material, which is never explicitly taught on the Numeric programme. However, it is assumed that a better understanding of underlying mathematical principles as well as an increase in the time spent practicing mathematics would contribute to an improvement in school Mathematics grades. The estimates of treatment on Section B scores are reported in Columns 5 to 8. They mirror the results in Columns 1 to 4. The improvement due to treatment on core grade 8 curriculum is smaller than the improvement on basic numeracy, but still highly significant.

The average treatment effect is 4.1 percentage points in Column 5. This is equivalent to an effect size of 0.234 standard deviations. In Column 6, when baseline mathematics results and school effects are held constant, the treatment effect falls slightly to 3.7 percentage points. In column 7 individual characteristics and baseline tests scores on the Raven's, English Literacy and typing test are added as controls, whilst in Column 8 the full set of controls is added. The coefficient on treatment is significant at the one percent level in all regressions. This shows that learners in treatment are achieving results that are between 3.7 to 3.9 percentage points higher than control group learners, holding individual demographic and household variables, school and baseline test results constant. The mean score of the control group was 26.4 percent on core Grade 8 curriculum at endline. This means that holding baseline mathematics results, school, individual and household characteristics, baseline scores on the English, Raven's, and typing tests constant, the average treatment learner was able to answer 15 percent more questions correctly than a comparable control group learner.

Table 12: OLS Estimates of treatment on mathematics scores on Section A and Section B at endline

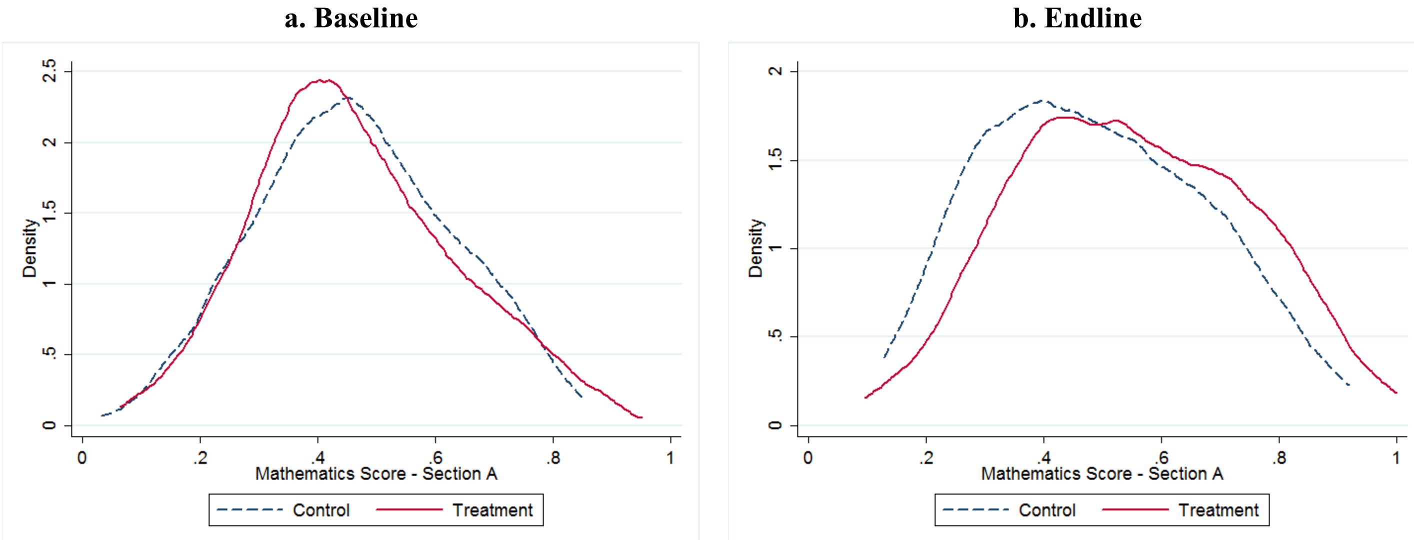
| | Endline Mathematics Test Scores | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Basic Numeracy - Section A | | | | Grade 8 - Section B | | | |
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment | 0.060*** | 0.057*** | 0.058*** | 0.058*** | 0.041** | 0.037*** | 0.039*** | 0.037*** |
| | (0.018) | (0.010) | (0.010) | (0.010) | (0.017) | (0.011) | (0.010) | (0.011) |
| Baseline-Section A | | 0.845*** | 0.716*** | 0.725*** | | 0.175 | 0.025 | 0.061 |
| | | (0.133) | (0.140) | (0.151) | | (0.132) | (0.130) | (0.139) |
| (Baseline-Section A)$^2$ | | -0.000 | 0.020 | -0.012 | | 0.619*** | 0.641*** | 0.591*** |
| | | (0.126) | (0.129) | (0.138) | | (0.131) | (0.124) | (0.133) |
| Female | | | -0.022** | -0.008 | | | -0.003 | 0.005 |
| | | | (0.010) | (0.012) | | | (0.012) | (0.013) |
| Age | | | 0.334*** | 0.291** | | | 0.306*** | 0.377*** |
| | | | (0.105) | (0.118) | | | (0.104) | (0.116) |
| Baseline – Raven's Score | | | 0.107*** | 0.093** | | | 0.155*** | 0.131*** |
| | | | (0.035) | (0.039) | | | (0.035) | (0.037) |
| Baseline -English Literacy Score | | | 0.076** | 0.095** | | | 0.112*** | 0.129*** |
| | | | (0.038) | (0.039) | | | (0.037) | (0.039) |
| Baseline -Typing (Keys per minute/100) | | | 0.029 | 0.026 | | | 0.020 | 0.026 |
| | | | (0.021) | (0.021) | | | (0.023) | (0.023) |
| At least one computer at Home | | | | -0.011 | | | | -0.006 |
| | | | | (0.015) | | | | (0.014) |
| Internet on a Computer at Home | | | | 0.008 | | | | -0.007 |
| | | | | (0.013) | | | | (0.014) |
| Travel Time to School | | | | 0.001 | | | | -0.000 |
| | | | | (0.028) | | | | (0.033) |
| Learner Likes Mathematics a Lot | | | | 0.024** | | | | 0.020 |
| | | | | (0.011) | | | | (0.012) |
| Constant | 0.488*** | 0.063* | 0.438*** | 0.384*** | 0.264*** | -0.041 | 0.163 | 0.209 |
| | (0.012) | (0.037) | (0.129) | (0.137) | (0.011) | (0.037) | (0.126) | (0.140) |
| School Controls | N | Y | Y | Y | N | Y | Y | Y |
| Learner Demographic Controls | N | N | Y | Y | N | N | Y | Y |
| RCE Controls | N | N | Y | Y | N | N | Y | Y |
| Household, education and attitude Controls | N | N | N | Y | N | N | N | Y |
| Observations | 453 | 453 | 453 | 453 | 453 | 453 | 453 | 453 |
| R-squared | 0.024 | 0.707 | 0.741 | 0.754 | 0.013 | 0.627 | 0.668 | 0.687 |

Notes: Robust standard errors are in parentheses. Statistical significance is given by *** p<0.01, ** p<0.05, * p<0.1. School controls consist of a dummy variable for each school. Learner demographic controls include race and home language. RCE controls also include a dummy variable indicating if there was a missing observation for the English literacy, Raven's or typing test at baseline. Household, education and attitude controls consist of mother alive, father alive, mother has matric, father has matric, learner has repeated at least one grade, participant in a sports team or cultural group as well as dummy variables for each covariate that is equal to 1 if the item for that individual is missing. The full table with all controls is reported in Appendix A, Table A2.

***Result 4: A significant improvement in basic numeracy and core Grade 8 curriculum is observed for treatment learners across the full distribution.***

Even with the sparse information provided in Table 11, it is clear that the standard deviation (17.2 percent) of the treatment group at baseline is higher than the standard deviation (16.5 percent) in the control group. This difference is evident in the distribution observed in Figure 7a below. At baseline, there are marginally more learners in the treatment group that are obtaining marks below 20 percent. However, the two distributions are reasonably similar for learners scoring below 25 percent. The difference is evident in the middle and upper ends of the distributions. The control plot lies to the right of the treatment plot for all values between 30 and 80 percent. This shows that the bulk of the control group is receiving higher scores compared to equivalently ranked individuals in treatment. However, of the learners scoring above 80 percent, the majority are in treatment. As a number of the top achieving learners in the sample are in treatment, the analysis of the distribution as a whole becomes particularly important.

Figure 7: Distribution of mathematics scores – Basic numeracy (Section A) at baseline and endline



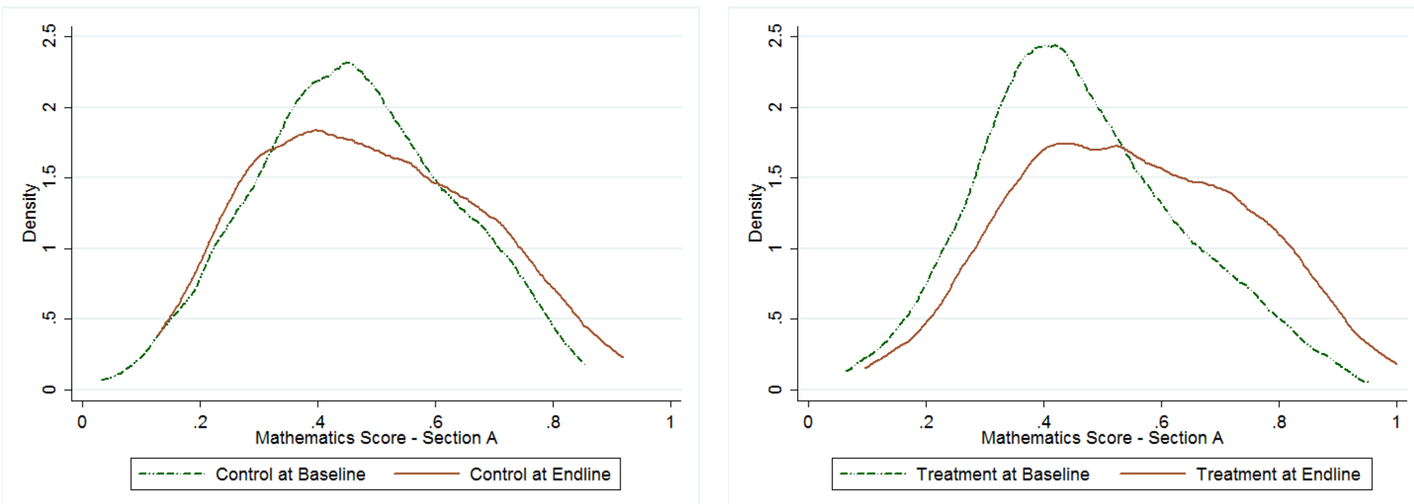

Source: Own data collected in 2013

Figure 7b shows the endline distribution of basic numeracy scores of the treatment group relative to the control group. It is evident that with the exception of a few of learners at the very bottom of the distribution, the treatment function lies to the right of the control function at every point along the distribution.

Figure 8 compares the baseline and endline distributions within the control and treatment groups. The overall increase in the mean of the combined treatment and control groups is 5.16 percentage points. There is, however, also an increase in the standard deviation in both the control and treatment group's endline numeracy scores. The first graph of Figure 8 shows that from baseline to endline the distribution of numeracy scores in the control group flattens and moves marginally to the right. At endline, there is an increase in the number of learners scoring above 60 percent. However, whilst it is clear that many of the learners are performing better, the proportion of learners achieving less than 30 percent has also increased. In comparison in Figure 8b it is evident that in the treatment group the entire distribution has shifted to the right by endline. Here it can be observed that even the very lowest performing students in treatment have improved over the course of the year. Results from prior studies suggest that learners at the bottom of the distribution benefit the most from computer assisted learning (Banerjee et. al, 2007, Barrow, Markman & Rouse, 2009). Here, however, it appears that learners throughout the distribution are improving, potentially, by similar amounts[42]. Estimates reported in Table 19 serve to confirm this hypothesis.

Figure 8: Distribution of mathematics scores – Basic numeracy of treatment and control groups comparing baseline and endline

**a. Control** **b. Treatment**




Source: Own data collected in 2013

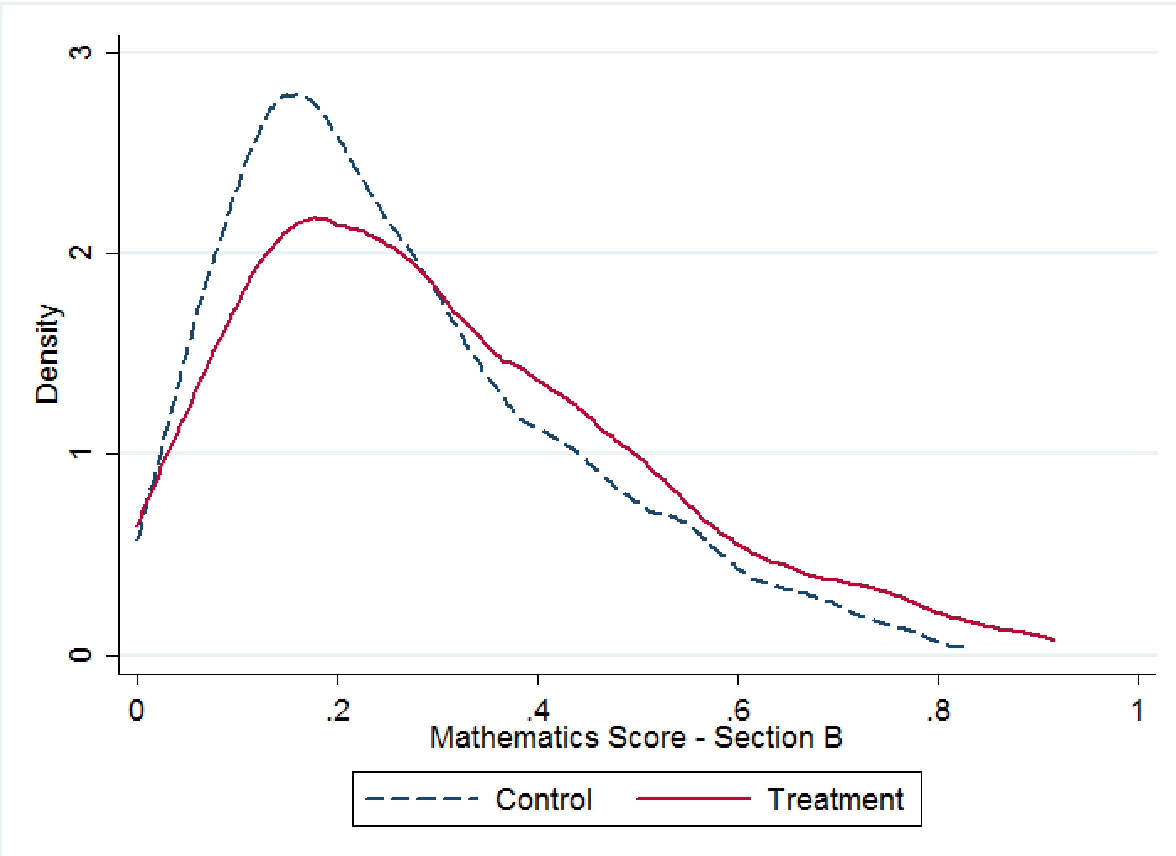
Figure 9 shows the distribution of the endline Section B mathematics scores of the control and the treatment groups. The average score obtained for Section B was considerably lower than the average score achieved in Section A. The scores are generally very poor with 75 percent of learners achieving below 40 percent at endline on core Grade 8 material. Also, whilst the distribution of the Section A scores approximated those of a normal distribution, the distribution of the Section B scores more closely resembles a log-normal distribution with a median at around 20 percent and a long tail to the right,

The general pattern, when comparing the control and treatment groups, is similar to the one observed in the endline Section A distributions presented in Figure 7b. This same pattern is observed in Figure 9 where it is evident that the treatment scores are higher than the scores obtained by control group learners at all points along the distribution, with the exception of a small proportion of learners at the very bottom. In contrast to the Section A results where the standard deviations of the treatment and control group at

[42] As Banerjee and Duflo (2009) point out, when heterogonous treatment effects are considered, the distribution of the outcome variable at endline is not the same as the distribution of the treatment effect. As a cross reference a distribution of the difference in Section A outcomes between baseline and endline are included in Appendix A, Figure A1.

endline were similar[43], the treatment distribution of Section B results is much flatter at endline, corresponding to a higher standard deviation in the treatment group of 19.4 percent compared to a standard deviation of only 16.7 percent in the control group.

Figure 9: Distribution of mathematics scores – Grade 8 curriculum at endline




Source: Own data collected in 2013

The average effect of treatment is positive and highly significant with an increase in 5.8 percentage points for Section A and 3.9 percentage points on Section B in the treatment group when compared to the control group, holding all else constant.

### Results on Subsections of the Mathematics Test

This next section investigates which mathematical competencies are being improved as a result of the programme. Section A tested five main topics: Basic Arithmetic, Fractions, Decimals, Number Patterns and Word Problems, whereas Section B covers seven subsections: Number Systems, Introduction to Variables, Solving Basic Equations, Geometry, Statistics, Word Problems and Area and Volume.

[43] Specifically, the difference in standard deviations at baseline was 0.6 for Section A and this difference was 0.9 at endline. It is expected that the standard deviation of Section B scores, had Section B been written at baseline, would have been greater in the treatment group. So because the distributions are so different, it is difficult to speculate how much of this difference in variance is due to baseline differences between control and treatment distributions in ability and how much is a result of the programme.

***Result 5: Treatment learners perform significantly better on all basic numeracy subsections and achieve higher average scores on all core Grade 8 curriculum subsections.***

Table 13 presents a breakdown of regression results showing the treatment effect on various subsections of the test. The first panel shows the five subsections of Section A, whilst Panel 2 shows the seven subsections of Section B. In Column 1, the coefficients on treatment are presented, with the standard errors in Column 2. The last column presents the relative weight of each subsection within Section A or Section B.

It is clearly evident from the results in Column 1 of Table 13 that treatment learners are outperforming control learners in every subsection of Section A and Section B, as the coefficient on treatment is large and positive for all subsections. In the top panel, the effect of treatment is also highly significant for all five subsections of Section A. The subsection with the smallest coefficient on treatment is Fractions. Here treated learners are scoring only 4.5 percentage points higher than equivalent control group learners, holding baseline mathematics, English, Raven's SPM's and typing speed test results, school and demographic variables constant. The Decimals section was the subsection which was most improved by the programme with learners in treatment scoring 7.2 percentage points higher, on average, than control learners, holding all else constant. Neither the coefficient on Fractions nor the coefficient on Decimals is significantly different from the total effect of 5.8 percent[44].

In panel 2, whilst treatment learners are scoring higher on average in all seven subsections than control learners, the treatment effect is only significant at the 1 percent level for two subsections, Introduction to Variables and Solving Equations, and at the 5 percent level for Geometry.

This shows that the programme is not targeting one specific skill or set of skills. Rather it is addressing a wide range of numeracy skills. Similarly, the programme is improving outcomes in all subsections in Section B. However, the biggest and most significant gains are observed on questions relating to Introduction to Variables, Equations and Geometry. Although a number of the concepts from Geometry would have been covered in the primary school curriculum, variables are only introduced in Grade 8, thus the learners on treatment achieved higher results on material testing new concepts.

[44] Working with a null hypothesis of $\beta_{treat} = 0.059$. The p-value for the fraction subsection is 0.303 whilst the p-value in the decimal subsection is 0.439. So neither is significantly different from 0.059 at a 30 percent level of significance.

Table 13: Shows the effect of treatment on endline scores of various subsections of Section A and Section B of the mathematics tests

| **Top Panel: Basic Numeracy Competencies** | | | |
|---|---|---|---|
| | Endline – Section A Results | | |
| | Treatment | Standard Error | Proportion of Section A Marks |
| Basic Arithmetic Score | 0.056*** | (0.012) | 0.35 |
| Fractions Score | 0.044*** | (0.013) | 0.29 |
| Decimals Score | 0.071*** | (0.016) | 0.24 |
| Number Patterns Score | 0.072*** | (0.020) | 0.11 |
| Word Problems Score | 0.064*** | (0.017) | 0.19 |
| **Bottom Panel: Core Grade 8 Curriculum Competencies** | | | |
| | Endline – Section B Results | | |
| | Treatment | Standard Error | Proportion of Section B Marks |
| Number Systems Score | 0.014 | (0.021) | 0.07 |
| Introduction to Variables Score | 0.043*** | (0.013) | 0.28 |
| Solving Equations Score | 0.051*** | (0.013) | 0.13 |
| Geometry Score | 0.042** | (0.019) | 0.17 |
| Statistics Score | 0.039 | (0.024) | 0.10 |
| Word Problems (B) Score | 0.031 | (0.019) | 0.15 |
| Area and Volume Score | 0.031 | (0.021) | 0.10 |

Notes: Number of observations for all 12 regressions is 453. Column 1 shows the coefficient on treatment, with the robust standard errors in parentheses in column 2. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include a linear and a squared term of baseline mathematics Section A test scores along with a dummy if the baseline mathematics score was missing. Dummy variables for the schools, gender, age, race and home language and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for the English literacy, Raven's or typing test at baseline. Column three shows what proportion of the marks for Section A or Section B were allocated to each subsection.

### Attitude Towards Mathematics

Besides the direct impact on test and school grades, participating in mathematics related activities in a more informal setting with constant feedback and positive reinforcement in the form of energy points, badges and certificates, should not only inspire a sense of achievement in the learners, but also build confidence and improve the overall attitude of the learners towards mathematics.

A large body of research on mathematics anxiety has found a correlation between low performance on mathematics tests and heightened mathematics anxiety as well as a negative attitude towards the subject (Hembree, 1990). There is no final consensus on the direction of the causality and it is likely that there is some feedback where low performance raises anxiety and any anxiety also hampers performance. However, regardless of the direction of the causality, a corresponding decrease in negative attitudes towards mathematics should be observed alongside the increase in test scores. Learners were asked three questions relating to their attitude towards mathematics in the survey at endline. The first simply asked whether they liked the subject a lot, a little, or not at all. The second asked them how much more they liked mathematics now compared to how they felt in January, ranging from a lot less (-3) to a lot more (3) on a 7 point scale. The third question asked what they felt was the highest Level[45] they could achieve if their tried their hardest. The results are reported in Table 15.

***Result 6: Learners in treatment are significantly more likely to report having a positive attitude towards mathematics at endline.***

Only 42.7 percent of learners in the control group stated that they liked mathematics a lot at endline, compared to 45.5 percent at baseline (Table 6, Row 1). Conversely the number of treatment learners who said that they liked mathematics a lot increased from 47.1 percent at baseline to 54.7 at endline. This is a full 12 percentage points higher than the control group. Correspondingly, when learners were asked how their feelings towards mathematics had changed over the year, the responses from learners in the treatment group were also more positive. Whilst the median learner from the control group responded that they liked mathematics a little more (+1) at endline compared to baseline, the median leaner from the treatment group stated that they liked mathematics more (+2) at endline than they had at baseline.

These results are supported by evidence from a feedback form, filled out by 194 learners who were still on the programme in September. Figure C4 in Appendix C shows that more than three quarters of the respondents said that they liked Mathematics or that they liked it a lot. Additionally, 71.65 percent of learners (Appendix C, Figure C6) stated that the Numeric sessions helped them in their regular classes often or all the time as opposed to helping them sometimes or not at all. Many learners also mentioned that they were pleased that their school grades had improved and attributed this to the Numeric intervention.

Table 15: Table showing the means of three variables measuring attitude towards mathematics at endline for learners in the evaluated sample.

| VARIABLES | Control | Treatment | Difference (Treatment-Control) |
|---|---|---|---|
| Endline - Like Mathematics A Lot | 0.430 | 0.549 | 0.120** |
| | (0.496) | (0.499) | (0.049) |
| Endline – Math Attitude Comparison Scale | 1.035 | 1.410 | 0.375** |
| (Scores range from -3 to 3) | (1.697) | (1.590) | (0.163) |
| Endline – Perception of Highest Possible Level Achievable in School Mathematics | 5.703 | 5.763 | 0.060 |
| (Levels range from 1 to 7) | (1.280) | (1.237) | (0.120) |
| **Number of Observations** | **222** | **219** | **441** |

Notes: Column 1 and column 2 have standard deviations in parentheses. Standard errors are in parentheses in last column. Statistical significance is indicated by *** p<0.01, ** p<0.05, * p<0.1, in the third column.

[45] In South African secondary schools, school grades are divided into 7 levels. Level 1 is classified as “not achieved” which means the learner obtained lowest level with a score between 0 and 29 percent, whilst level 7 represents an ”outstanding achievement” which is equivalent to a grade between 80 and 100 percent.

## Drop-Out and Attendance Rates

### Programme Drop-Out Rates

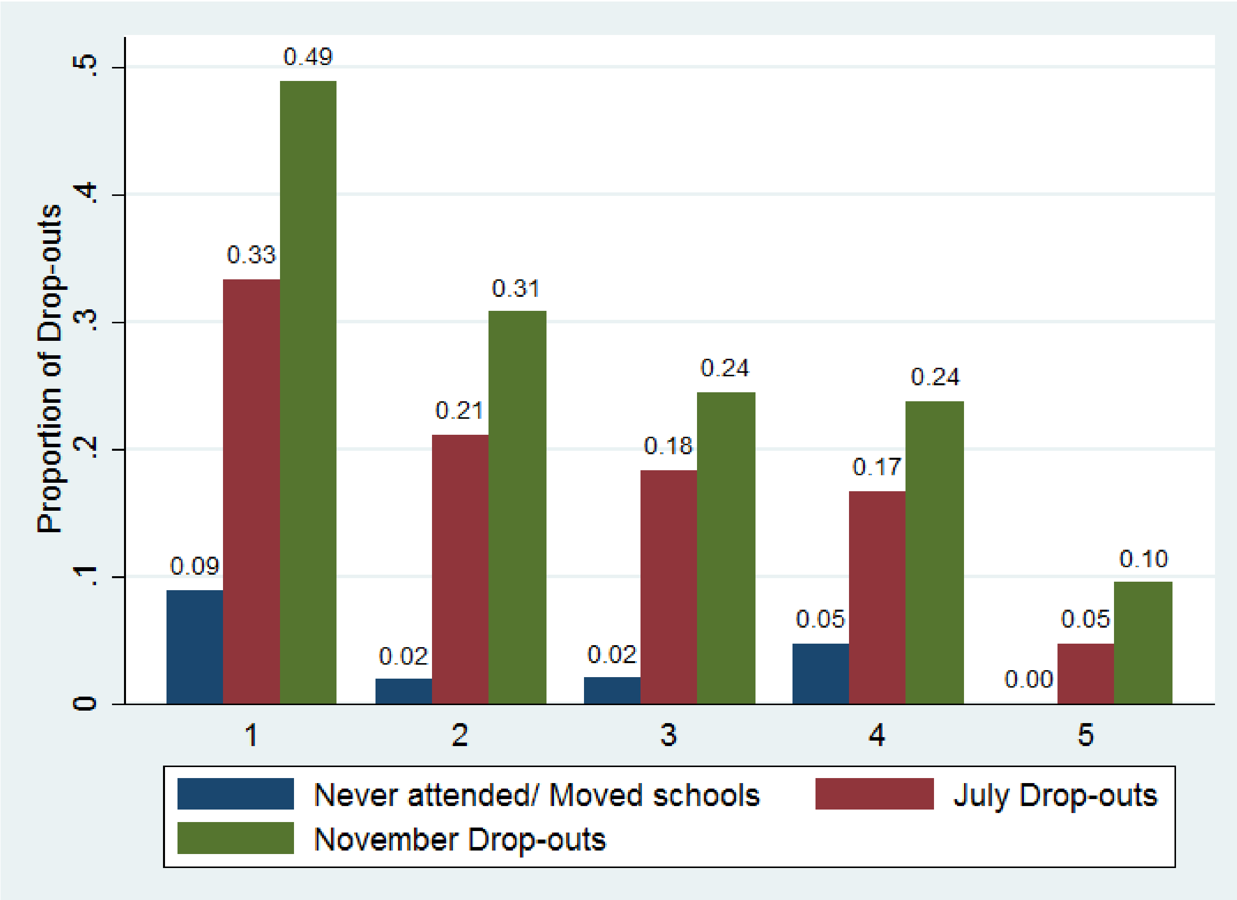
The success of the programme is closely linked to the real take-up rates among selected learners. Of the 236 learners that had originally been selected, there were 8 learners in total who did not attend a single class. Overall the drop-out rate by the end of the second term was moderate. A total of 191 (81 percent) learners were still actively attending classes. Over the next two terms, a further 23 learners dropped out of the programme. This equates to a drop-out rate of 19 percent over the first two terms and a total drop-out rate of 29 percent over the course of the year. Therefore, only 7 out of every 10 learners in the sample completed a full year of treatment.

> ***Result 7: Treatment learners that perform poorly on the baseline mathematics test are significantly more likely to drop out of the programme over the course of the year.***

In the results of an intention-to-treat analysis all learners selected into treatment are regarded as treated learners. However, in order for the treatment to have had any impact, learners need to have attended Numeric classes. Dropping out or ceasing to attend is a decision on the side of the learner that is decidedly non-random. Figure 10 shows the proportion of learners that dropped out of the programme by baseline mathematics quintiles[46]. It is evident that there is a strong negative correlation between achievement on the baseline mathematics test and drop-out rates. By the end of July, 33 percent of learners from the bottom quintile had stopped attending Numeric classes, whereas only 5 percent of learners from the fifth quintile had dropped out. In November, half of the learners in the bottom quintile were no longer attending classes compared to only 10 percent of top quintile learners.

[46] Quintile 1 comprises of learners from the lowest 20 percent of the baseline mathematics distribution, whilst quintile 5 consists of the top performers at baseline.

Figure 10: Proportion of learners that did not attend any classes, dropped out by the end of the second term and those that had dropped out at the end of the 4th term by baseline numeracy quintile




### Attendance Rates and Active Learning

These learners that dropped out cannot be excluded from the analysis, as this new subsample of the treatment group would no longer be comparable on average to the control group. It would nevertheless be useful to confirm that learners who attended the programme or attended the programme for longer performed better at endline compared to learners that dropped out over the course of the year.

As the majority of the Numeric programme takes place online, various learner statistics are automatically recorded by the Khan Academy website. In order to get a sense of the intensity of treatment, two of these statistics are considered. The first is the number of skills Mastered. This can be viewed as a proxy for the amount of work done by the learner on the programme. Unfortunately, Khan Academy changed their system in June/July of 2013, which meant that this variable only counts the modules completed by learners in the third and fourth terms. The second is the total amount of time spent on Khan Academy. This variable measures the total amount of time each learner spent doing exercises and watching videos online on Khan Academy from February to November 2013[47]. This is an alternate measure of attendance. Learners were generally very careful to log on to their accounts during class so that any work done was recorded and could count towards certificates. It also has the added benefit of being able to track the time learners spent on the website outside of Numeric classes[48].

Table 16 shows summary statistics for the number of skills mastered and the total time spent on the website. The mean number of skills Mastered was 57[49]. There was a heaping at 0 with 48 learners[50] who did not master any skills, whilst the maximum number of skills mastered was 281 out of a maximum of 466. The mean amount of time spent on the website was 29 hours and 20 minutes, with a minimum of 0 and a maximum of just under 160 hours. Histograms of distributions of the variables, skills Mastered and time spent on Khan Academy, of treatment learners are included as Figure A4 and Figure A5 in Appendix A. The correlation between the hours spent on the Khan Academy website and the number of skills Mastered is 0.849, showing that learners who spent more time on Khan Academy also generally mastered a greater number of skills.

Table 16: Table to show the mean number of skills Mastered and time spent on Khan Academy

| | **Skills Mastered** | **Total Time Spent on Khan Academy (hours)** |
|---|---|---|
| Mean | 57.09 | 29.34 |
| Standard Deviation | 46.91 | 21.85 |
| Minimum | 0 | 0.00 |
| Maximum | 281 | 158.40 |
| **Number of Observations** | **223** | **223** |

Source: This data was downloaded from the Khan Academy website on the 16 January 2014.

[47] The measure of total time spent on Khan Academy is the sum of the time spent doing exercises and watching videos. However as the learners did not watch videos online it is essentially just a measure of how long they spent answering questions. It does not just measure the time spent logged onto your account. However if a learner started a question and got sidetracked whilst they were busy answering the question, then the website would count this period toward their total time. Therefore total time spent on Khan Academy is a good gauge of attendance, but not necessarily of the amount of work done.

[48] Most learners spent very little time on Khan Academy outside of Numeric classes. However there was a group of five learners (2.1 percent of treatment learners) from one school who were clear outliers in this regards. The least active member of this group spent three times more time on the website than the median learner, whilst the most active learners spent 158 hours on Khan Academy, which about 6 times the median amount of 26.64 hours. The programme introduces the learners to a free resource, and provided they are able to gain access to the internet outside of Numeric class it is unsurprising that a small number of learners started to use the website more frequently.

[49] The recommended curriculum on Khan Academy lists 68 skills that a learner should typically have Mastered following the typical Grade 3 to Grade 5 United States curriculum. However, the learners were not required to complete problems in any particular order. As a result some learners may have completed very easy exercises whilst other may have challenged themselves with more difficult problems.

[50] This number is roughly equivalent to the 45 learners who had dropped out by July.

The following specification was employed to investigate the relationship between the number of skills Mastered and amount of time spent on the website and the results obtained on Section A and Section B of the endline mathematics test:

$$Y_{isEL} = \propto + \beta T_{is} + \beta_2 \boldsymbol{H'}_{\mathrm{is}} + \gamma Y_{isBL} + \pi Y_{isBL}^2 + \theta \boldsymbol{S'}_s + \rho \boldsymbol{Z'}_{isBL} + \delta \boldsymbol{X'}_{is} + \varepsilon_{is}. \qquad (2)$$

Equation 2 is based on the model presented in Equation 1. However the variable $\boldsymbol{H'}_{\mathrm{is}}$ has been added. The variable $\boldsymbol{H'}_{\mathrm{is}}$ is a matrix consisting of the number of skills Mastered and the total time each treatment learner has spent on the Khan Academy website. It is assumed that the majority of control learners will have spent very little time on the Khan Academy website. Additionally, as there was no way to track the control learner's activities on Khan Academy, all control learners were automatically assigned a zero for number of skills Mastered and hours spent on the website[51]. In contrast to Equation 1, the treatment variable, $T_{is}$ , in this specification only represents *selection* into treatment. The estimate $\beta$ represents the average improvement to endline scores for a treatment learner who dropped out at the beginning of the year and has therefore spent zero hours on Khan Academy and has Mastered no skills. The expectation is that there will be no significant impact on mathematics outcomes simply as a result of *selection* into treatment. Rather, a learner's mathematics outcomes will only have improved if they have actively spent time on the Khan Academy website and have Mastered a number of skills. Here, $\beta_2$, is the estimate of interest. This coefficient represents the average improvement on mathematics test scores for a treatment learner that has Mastered an additional 100 skills on Khan Academy or has spent 100 more hours on the website, holding all other factors constant.

***Result 8: Learners with higher attendance at Numeric classes and learners that completed more exercises on Khan Academy benefited the most from the programme.***

In Table 17 the estimates on skills Mastered are shown in Row 2, whilst Row 3 shows the estimates on the total number of hours spent on Khan Academy. Columns 1 to 3 have Section A of the endline mathematics test as the independent variable, whilst in Column 4 to Column 6, Section B is the independent variable.

In the first three columns the coefficient on treatment is effectively zero, showing that there is zero detectable effect on endline basic numeracy scores as a result of *selection* into treatment. Thus for a learner who was selected into treatment, but spent zero hours on the website and Mastered no skills, there is no discernable impact on basic numeracy outcomes.

However, the time spent on Khan Academy and number of skills Mastered are positively related to endline mathematics outcomes. The estimate on skills Mastered by learners in treatment in Column 1 has a magnitude of 0.102 and is highly significant. This shows that, ceteris paribus, mastering 60 additional skills on Khan Academy resulted in a mean improvement of 6 percent on the endline basic numeracy test. Similarly the estimate on the interaction term of time spent on Khan Academy by learners in treatment in column 2 is also highly significant and shows that spending 30 additional hours on Khan Academy raised

---

51

Section A test scores by an average of 5.9 percentage points, holding all else constant. The regression in column three looks at both the effect of skills Mastered and time spent on Khan Academy on endline Section A results. The estimate on skills Mastered and total hours spent on Khan Academy are both positive, but only the coefficient on skills Mastered is significant. Thus, whilst the average increases due to an additional hour on Khan Academy or Mastering an additional skill are of a similar magnitude, on average, it is clear that mastering additional skills is more likely to result in an improvement in endline outcomes.

In contrast to the result on Section A, the estimate of selection into treatment on endline core Grade 8 curriculum test results is negative and significant in columns 4 to 6. This suggests that treatment learners that elected to drop out of the Numeric programme are, on average, performing worse on core Grade 8 curriculum material than a control group learner with the same observable characteristics at baseline. Section B tests new material. Therefore, learners who drop out are performing relatively poorly in Grade 8, holding all factors constant. An illustration of this can be observed in Figure A3 of Appendix A.

Table 17: Table to show OLS estimates of selection into treatment, number of skills Mastered and time spent on Khan Academy on endline mathematics results.

| | Endline Mathematics Test Score - Section A | | | Endline Mathematics Test Score - Section B | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.000 | 0.001 | -0.003 | -0.035** | -0.031** | -0.037*** |
| | (0.013) | (0.012) | (0.013) | (0.014) | (0.014) | (0.014) |
| Skills Mastered on Khan Academy ÷ 100 | 0.102*** | | 0.077*** | 0.131*** | | 0.116*** |
| | (0.014) | | (0.028) | (0.019) | | (0.031) |
| Total Hours on Khan Academy ÷ 100 | | 0.199*** | 0.061 | | 0.244*** | 0.035 |
| | | (0.027) | (0.058) | | (0.040) | (0.063) |
| Baseline Mathematics Score -Section A | 0.683*** | 0.673*** | 0.678*** | -0.017 | -0.027 | -0.020 |
| | (0.132) | (0.135) | (0.133) | (0.126) | (0.125) | (0.127) |
| (Baseline Mathematics Score -Section A)$^2$ | -0.006 | 0.036 | 0.006 | 0.609*** | 0.661*** | 0.615*** |
| | (0.126) | (0.125) | (0.126) | (0.127) | (0.122) | (0.129) |
| Constant | 0.400*** | 0.402*** | 0.398*** | 0.113 | 0.119 | 0.113 |
| | (0.120) | (0.127) | (0.122) | (0.114) | (0.123) | (0.115) |
| Number of Observations | 453 | 453 | 453 | 453 | 453 | 453 |
| R-squared | 0.764 | 0.761 | 0.765 | 0.713 | 0.704 | 0.713 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include a dummy if the baseline mathematics score was missing. Dummy variables for the schools, gender, age, race and home language and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for the English literacy, Raven's or typing test at baseline.

A very similar pattern is observed in Columns 4 to 6 for the endline Section B mathematics results. When considered separately the estimates of number of skills Mastered as well as time spent on Khan Academy were positive and highly significant. Column 4 shows that, holding all else constant, a treatment learner that Mastered 10 additional skills scored 1.3 percentage points higher on core Grade 8 curriculum.

Column 5 shows that a learner scored 2.4 percentage points higher on Section B of the endline mathematics test if they had spent 10 more hours on the website, ceteris paribus. In Column 6, where both variables are included in the function the estimate of number of skills mastered was highly significant, whilst the estimate on total hours spent on the website was highly insignificant. This shows that an improvement in endline outcomes is more likely to be observed if learners are actively practicing mathematics in their Numeric classes.

## Heterogeneous Treatment Effects

The following section investigates whether the improvement observed in test scores varies across different subgroups of the evaluated sample. The equation used to analyse the differential effect of treatment on particular subgroups is an extension of Equation 1, shown below:

$$Y_{isEL} = \propto + \beta T_{is} + \beta_2 \boldsymbol{R'}_{is} + \beta_3 (T_{is} \times \boldsymbol{R'}_{is}) + \gamma Y_{isBL} + \pi Y_{isBL}^2 + \theta \boldsymbol{S'}_s + \rho \boldsymbol{Z'}_{isBL} + \delta \boldsymbol{X'}_{is} + \varepsilon_{is} \quad (3)$$

Two terms have been included $\boldsymbol{R'}_{is}$ and $(T_{is} \times \boldsymbol{R'}_{is})$. The variable $\boldsymbol{R'}_{is}$ is a vector or matrix of dummies for the subgroup being investigated for learner $i$ at school $s$. The coefficient on the interaction term, $\beta_3$, is the estimate of interest. The interaction term, $(T_{is} \times \boldsymbol{R'}_{is})$, is equal to one if an individual is both in treatment and in the subgroup being analysed. As such, a significant estimate on such an interaction term implies that the effect of treatment is different for this particular subgroup. As demonstrated in Table 10, treatment is uncorrelated with baseline measures of all subgroups examined, which suggests that the treatment effect should be unaffected by the inclusion of these additional terms, unless there is an interaction between treatment and the subgroup of interest. Thus if there is a significant change in the estimate on treatment, it is expected that at least one of the interaction terms would then show up as significant.

### Baseline Mathematics Ability and Attitude towards Mathematics

This section provides a more detailed examination of the relationship between baseline mathematics ability as well as attitude towards mathematics, and treatment. Table 18 shows the three different subgroups that were considered. The first two specifications divided learners up according to their Mathematics ability at baseline. In the top panel of Table 18, learners were separated into quintiles by baseline mathematics results, whilst in the middle panel learners were spilt into a two groups. The group of learners with above median baseline mathematics results is labeled as Top Students, whereas those with below median scores are labelled as Bottom Students.

Figure 7b and Figure 9 suggested that the improvement from baseline to endline due to treatment was similar for learners across the distribution. This finding differs from recent findings on CAL programmes for Grade 3 learners in India and Grade 9 learners in the United States, where struggling learners at the bottom of the distribution benefitted more from CAL (Banerjee et. al, 2007, Barrow, Markman & Rouse, 2009). A result showing that baseline ability does not differentially effect treatment would, however, be unsurprising as CAL allows learners to move through the material at their own pace. On the Numeric programme learners also have some autonomy over what they learn. Thus, learners who have many gaps in their knowledge can work through the basics, whilst more advanced learners can quickly move onto more difficult material, allowing learners at any level to make progress.

In the bottom panel, the evaluated sample learners are divided into two groups. The first consisted of learners that stated, at baseline, that they liked mathematics a lot, whilst the second group comprised of learners that only liked mathematics sometimes or not at all. A significant number of learners in the evaluated sample did not state that they liked mathematics a lot, yet they still chose to apply to an extra-curricular mathematics programme. It is hypothesized that a learner would be more engaged in class and more likely to persevere at a task in a subject that they liked. Thus a regression is included to assess whether initial attitude towards mathematics has a differential effect on treatment, or whether learners that initially did not like mathematics a lot did just as well as a result of the programme.

Table 18: Table comparing the mean baseline and endline Section A mathematics scores of control and treatment for various subgroups based on baseline mathematics scores and attitude towards mathematics.

| | **Control** | | **Treatment** | | **Difference** *(Endline - Baseline)* | |
|---|---|---|---|---|---|---|
| | Baseline | Endline | **Baseline** | **Endline** | Control | Treatment |
| **Top Panel: Baseline Mathematics Section A - Quintiles** | | | | | | |
| 1 (Bottom Quintile) | 0.24 | 0.28 | **0.24** | **0.34** | 0.04 | 0.09 |
| | (0.07) | (0.10) | **(0.06)** | **(0.13)** | (0.10) | (0.11) |
| 2 | 0.37 | 0.42 | **0.37** | **0.46** | 0.05 | 0.09 |
| | (0.03) | (0.11) | **(0.02)** | **(0.14)** | (0.12) | (0.13) |
| 3 | 0.46 | 0.45 | **0.45** | **0.53** | -0.01 | 0.07 |
| | (0.03) | (0.12) | **(0.03)** | **(0.13)** | (0.12) | (0.13) |
| 4 | 0.56 | 0.57 | **0.56** | **0.64** | 0.01 | 0.08 |
| | (0.03) | (0.12) | **(0.03)** | **(0.12)** | (0.11) | (0.11) |
| 5 (Top Quintile) | 0.70 | 0.73 | **0.73** | **0.80** | 0.02 | 0.07 |
| | (0.06) | (0.10) | **(0.08)** | **(0.10)** | (0.09) | (0.09) |
| **Middle Panel: Baseline Mathematics Section A – Above and below Median** | | | | | | |
| Bottom Students | 0.33 | 0.37 | **0.34** | **0.43** | 0.03 | 0.09 |
| | (0.09) | (0.13) | **(0.09)** | **(0.15)** | (0.11) | (0.12) |
| Top Students | 0.60 | 0.61 | **0.62** | **0.69** | 0.01 | 0.07 |
| | (0.10) | (0.15) | **(0.11)** | **(0.15)** | (0.11) | (0.1) |
| **Bottom Panel: Attitude towards Mathematics at baseline** | | | | | | |
| Don't like it or like it sometimes | 0.41 | 0.44 | **0.43** | **0.50** | 0.02 | 0.07 |
| | (0.15) | (0.16) | **(0.16)** | **(0.18)** | (0.13) | (0.11) |
| Like it a lot | 0.53 | 0.55 | **0.51** | **0.61** | 0.02 | 0.09 |
| | (0.16) | (0.19) | **(0.18)** | **(0.20)** | (0.09) | (0.12) |
| **Total** | **0.46** | **0.49** | **0.46** | **0.55** | **0.02** | **0.08** |
| | **(0.17)** | **(0.19)** | **(0.17)** | **(0.20)** | **(0.11)** | **(0.12)** |
| **Number of Observations** | **231** | **223** | **230** | **221** | **461** | **444** |

In the top panel of Table 18 it is clear that the average improvement of learners in the treatment group from baseline to endline in the five quintiles was very similar, in the range of 7 to 9 percentage points. However, the differences between average endline mathematics scores of the treatment compared to the control group are more variable.

In the middle panel the average improvement of treatment learners from baseline to endline compared to the average improvement of the control group do not appear to be significantly different between the two subgroups. However, in the bottom panel, treatment learners that liked mathematics a lot at baseline had improved 7 percentage points more than the control group, whilst learners that liked mathematics less at baseline only improved 5 percentage points more than the control group.

Table 19 reports the OLS estimates of the extended model presented in Equation 3, which includes the baseline mathematics ability and attitude dummies along with the interaction terms. The results of the first specification which includes dummies from the first set of subgroups and their interaction terms are reported in Column 1 and Column 4. In this subgroup learners were divided into quintiles by baseline mathematics results in order to test if learners in any of the 5 quintiles were performing significantly better or worse than their treatment counterparts. Similarly, regression 2 and regression 5 report the treatment and heterogeneous treatment effects of top and bottom learners. Finally in regression 3 and regression 6 an interaction term and dummy for a positive attitude towards mathematics at baseline are included.

***Result 9: There is no differential treatment effect for students in different quintiles of the baseline distribution.***

The results in Column 1 show that the estimate on treatment increases to 6.9 percentage points when the interaction terms are included; the coefficients on all five interaction terms are small and statistically insignificant. This shows that the effect of treatment is not significantly different on average for any of the five quintiles. This is confirmed by the results in Column 2, which show that, holding all other factors constant, the treatment effect for a top student is 0.7 percentage points lower than the treatment effect for a comparable student in treatment who had a below median score at baseline. This difference is, however, not statistically significant.

Results presented in column 4 show that there is also no significant difference in the efficacy of treatment across baseline mathematics quintiles when the outcome variable is core Grade 8 curriculum instead of basic numeracy. Similarly, column 5 shows that there is no differential treatment effect for top students on endline Section B test scores.

***Result 10: A positive attitude towards mathematics at baseline does not result in significantly higher mathematics outcomes for treatment learners***

Column 3 and column 6 compare the success of learners with different initial attitudes towards mathematics at baseline. In column 3, the estimate on Like Mathematics a Lot is positive, at 2.2 percentage points, but insignificant. Surprisingly, even in the control group, learners who stated that they liked mathematics a lot did not improve significantly more than learners with a worse initial attitude towards mathematics. The estimate on the interaction term between treatment and Like Mathematics a Lot is 0.5 percentage points. This is highly insignificant and thus learners who initially liked mathematics did not improve relatively more on the programme than learners who did not like mathematics as much at baseline. The same result is observed for Section B scores, shown in column 6.

The treatment effect in the first 3 columns in Table 19 is not significantly different from the 5.8 percentage points reported in Table 12. Likewise in column 4, 5 and 6 none of the treatment effects are significantly different from the original 3.9 percentage points. This supports the claim that there are no heterogeneous treatment effects by baseline mathematics score or baseline attitude towards mathematics.

Table 19: OLS estimates showing the interaction between treatment and baseline mathematics scores and attitude towards mathematics.

| | Endline Mathematics Test Scores – Section A | | | Endline Mathematics Test Scores – Section B | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.069*** | 0.071*** | 0.056*** | 0.035* | 0.041*** | 0.045*** |
| | (0.025) | (0.016) | (0.013) | (0.020) | (0.014) | (0.012) |
| Baseline - Section A | | | 0.703*** | | | -0.009 |
| | | | (0.142) | | | (0.132) |
| (Baseline - Section A)2 | | | 0.011 | | | 0.664*** |
| | | | (0.131) | | | (0.126) |
| Maths – Quintile 2 | 0.128*** | | | 0.080*** | | |
| | (0.021) | | | (0.018) | | |
| Maths – Quintile 3 | 0.143*** | | | 0.093*** | | |
| | (0.022) | | | (0.017) | | |
| Maths – Quintile 4 | 0.226*** | | | 0.124*** | | |
| | (0.023) | | | (0.025) | | |
| Maths – Quintile 5 | 0.345*** | | | 0.295*** | | |
| | (0.024) | | | (0.026) | | |
| Maths – Quintile 2 × Treatment | -0.043 | | | -0.029 | | |
| | (0.034) | | | (0.029) | | |
| Maths – Quintile 3 × Treatment | 0.004 | | | 0.011 | | |
| | (0.035) | | | (0.029) | | |
| Maths – Quintile 4 × Treatment | -0.002 | | | 0.034 | | |
| | (0.034) | | | (0.034) | | |
| Maths – Quintile 5 × Treatment | -0.002 | | | 0.021 | | |
| | (0.032) | | | (0.035) | | |
| Top Student | | 0.157*** | | | 0.120*** | |
| | | (0.019) | | | (0.018) | |
| Top Student × Treatment | | -0.007 | | | 0.012 | |
| | | (0.024) | | | (0.025) | |
| Like Mathematics A Lot | | | 0.022 | | | 0.024 |
| | | | (0.015) | | | (0.015) |
| Like Mathematics A Lot × Treatment | | | 0.005 | | | -0.010 |
| | | | (0.019) | | | (0.021) |
| Constant | 0.627*** | 0.721*** | 0.454*** | 0.168 | 0.234 | 0.186 |
| | (0.132) | (0.146) | (0.130) | (0.129) | (0.143) | (0.125) |
| Number of Observations | 444 | 444 | 443 | 444 | 444 | 443 |
| R-squared | 0.731 | 0.637 | 0.749 | 0.660 | 0.551 | 0.677 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include: dummy variables for the schools, gender, age, race and home language and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for gender, age, race, home language or the English literacy, Raven's or typing test at baseline.

### Gender, Race and Home Language

It is possible that the structure of the Numeric programme or the learning style best suited to Khan Academy is particular to a gender, race or culture. For example in the case of gender, the social nature of the programme may appeal more to girls. However, boys may enjoy the focus on individual achievement and other game-like features of Khan Academy. A priori, it is not possible to know where there will be differential effects by gender. The questions on the Khan Academy website, as well as the videos, are all in English. As a result, learners with a better command of the language may be benefitting more. Additionally, as the founder is from the United States of America, many of the anecdotes and situations in the videos are decidedly American, which could potentially have differing effects on learners who come from different cultures.

Table 20 presents an overview of the average baseline and endline results of learners in the treatment and control groups, for the subgroups under consideration. The top panel separates out the boys and the girls. It is evident that, even at baseline, the average score of the girls in the evaluated sample is 3 percentage points lower than that of the boys[52]. Within the treatment group at endline, the difference in average scores between boys and girls in treatment increased to 5 percentage points. The middle panel showed the results of the two major race groups. At baseline, African learners were scoring 8 percentage points lower on Section A of the mathematics test. However between baseline and endline the improvement in average test scores of African and Coloured learners is comparable, although Coloured learners in treatment appear to be performing marginally worse on average when compared to the control group. Lastly, the bottom panel shows the average scores for the last set of subgroups, which divided learners up by home language. Afrikaans and isiXhosa speaking leaners have very similar baseline mathematics scores, whilst English speaking learners scores about 10 percentage points higher, on average, at baseline. There does not appear to be a significant treatment differential by race as the average increases in mathematics score of each race group between baseline and endline are similar (about 5 to 7 percentage points), on average, although English speaking treatment learners appear to be improving the least as a result of the programme.

[52]This difference in baseline mathematics score by gender is only observed in the applicant group. There is no significant gender difference observed in the basic numeracy outcomes of all grade 8's at baseline. See Table A3 in Appendix A.

Table 20: Table comparing the mean baseline and endline Section A mathematics scores of control and treatment for gender, race and home language subgroups.

| | Control | | Treatment | | Difference *(Endline - Baseline)* | |
|---|---|---|---|---|---|---|
| | Baseline | Endline | **Baseline** | **Endline** | Control | Treatment |
| **Top Panel: Gender** | | | | | | |
| Male | 0.48 | 0.50 | **0.48** | **0.58** | 0.02 | 0.10 |
| | (0.16) | (0.20) | **(0.18)** | **(0.21)** | (0.12) | (0.12) |
| Female | 0.45 | 0.48 | **0.45** | **0.53** | 0.03 | 0.07 |
| | (0.17) | (0.17) | **(0.17)** | **(0.19)** | (0.11) | (0.11) |
| **Middle Panel: Race** | | | | | | |
| African | 0.42 | 0.43 | **0.41** | **0.50** | 0.02 | 0.09 |
| | (0.14) | (0.16) | **(0.15)** | **(0.19)** | (0.11) | (0.13) |
| Coloured | 0.49 | 0.52 | **0.50** | **0.58** | 0.03 | 0.08 |
| | (0.17) | (0.20) | **(0.17)** | **(0.20)** | (0.11) | (0.10) |
| **Bottom Panel: Home Language** | | | | | | |
| Afrikaans | 0.41 | 0.45 | **0.38** | **0.49** | 0.03 | 0.10 |
| | (0.18) | (0.19) | **(0.17)** | **(0.20)** | (0.10) | (0.08) |
| English | 0.51 | 0.54 | **0.53** | **0.61** | 0.03 | 0.08 |
| | (0.17) | (0.20) | **(0.17)** | **(0.20)** | (0.12) | (0.10) |
| isiXhosa | 0.41 | 0.42 | **0.41** | **0.50** | 0.01 | 0.08 |
| | (0.13) | (0.15) | **(0.15)** | **(0.18)** | (0.11) | (0.13) |
| **Number of Observations** | 231 | 227 | **230** | **226** | 461 | 453 |

***Result 11: The demographic characteristics of gender, race and home language have no differential effects on treatment***

Table 21 presents the multivariate regression results of the interaction effects of treatment with gender, home language and race. The models in Column 1 and Column 4 contain an interaction term between gender and treatment. The treatment for girls was 1 percent point lower on average than the treatment effect for boys on Section A and 0.4 percentage points lower on Section B of the endline mathematics test, holding all other factors constant. However neither is significantly different from zero. Similarly, learners with English or isiXhosa as their home language do not improve significantly more or less than learners on treatment who speak Afrikaans at home, holding all else constant. Neither does the race of the learner appear to have any effect on the efficacy of treatment. Coloured learners in the treatment group improved an average of 0.9 percentage points less on Section A and 0.3 percent more on Section B than an African learner in treatment, holding all other factors constant. These differences are practically small and highly statistically insignificant, suggesting that the magnitude of the treatment effect on basic numeracy and core Grade 8 curriculum is not significantly affected by gender, race or home language.

Table 21: OLS estimates showing the interaction effects of treatment with gender, home language and race.

| | Mathematics Test Scores – Section A | | | Endline Mathematics Test Scores – Section B | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.064*** | 0.058** | 0.064*** | 0.041** | 0.040 | 0.036** |
| | (0.015) | (0.028) | (0.016) | (0.017) | (0.031) | (0.017) |
| Female | -0.017 | -0.022** | -0.022** | -0.001 | -0.003 | -0.004 |
| | (0.014) | (0.011) | (0.011) | (0.014) | (0.012) | (0.012) |
| Female × Treatment | -0.010 | | | -0.004 | | |
| | (0.020) | | | (0.021) | | |
| Home Language : English | -0.017 | -0.015 | -0.018 | 0.004 | 0.004 | 0.004 |
| | (0.016) | (0.025) | (0.016) | (0.018) | (0.025) | (0.018) |
| Home Language : English × Treatment | | -0.006 | | | -0.000 | |
| | | (0.031) | | | (0.034) | |
| Home Language : Xhosa | -0.119** | -0.127** | -0.122** | -0.105 | -0.106 | -0.115* |
| | (0.050) | (0.053) | (0.051) | (0.065) | (0.068) | (0.066) |
| Home Language : Xhosa × Treatment | | 0.010 | | | -0.000 | |
| | | (0.032) | | | (0.035) | |
| Coloured | -0.108** | -0.108** | -0.104** | -0.113* | -0.112* | -0.123** |
| | (0.046) | (0.047) | (0.049) | (0.062) | (0.062) | (0.062) |
| Coloured × Treatment | | | -0.009 | | | 0.003 |
| | | | (0.020) | | | (0.021) |
| Constant | 0.440*** | 0.440*** | 0.441*** | 0.164 | 0.164 | 0.172 |
| | (0.128) | (0.130) | (0.129) | (0.126) | (0.128) | (0.126) |
| Number of Observations | 453 | 451 | 449 | 453 | 451 | 449 |
| R-squared | 0.741 | 0.741 | 0.739 | 0.668 | 0.670 | 0.667 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include: a dummy if the baseline mathematics score was missing, dummy variables for the schools, age and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for the English literacy, Raven's or typing test at baseline.

### Baseline Raven's SPM, English and Typing Ability

A learner's baseline cognitive skills, English fluency and level of comfort working with a computer all have the potential to significantly affect, either the take-up of the programme, or may hamper or augment a learner's progress on the programme. A higher level of baseline computer literacy and cognitive skill, as measured by the Raven's SPM, might enable the learners to adapt to working on a computer more quickly and they might thereby start to benefit from the programme earlier. Similarly, a better understanding of English would be helpful to a learner on the programme as they would better understand the videos and exercises and, as such, may derive more value from Numeric. It is also possible that learners who initially struggle with the language or find it difficult to navigate on a computer would be more likely to drop out of the programme.

A set of subgroup dummies and interaction terms needed to be created for the analysis. Firstly, for the Raven's results, learners were divided into terciles based on their baseline Raven's scores. There were two reasons for the decision to use terciles as opposed to only two categories, above and below the median. Firstly, because of the high drop-out rate amongst learners at the bottom of the baseline mathematics distribution, if learners at the bottom or the middle of the distribution gained proportionally more from treatment this might not be observed if there were only two subgroups. Additionally, it is possible, that there may be differences by school[53] which might not be picked up if there was only a top and bottom dummy[54]. Three interaction terms of the Raven's terciles with treatment were then created. Similarly, tercile and interaction dummies were then created for the baseline English literacy test, as well as the typing test results.

Table 22 presents the average results of the basic numeracy section of learners in the three English, Raven's and typing terciles, in control and treatment. The top panel shows that treatment learners in the top and middle English terciles are improving significantly more from baseline to endline than treatment learners in the bottom tercile for English. Similarly, the results of learners in the top tercile of baseline Raven's results improve significantly more from baseline to endline than learners in the bottom Raven's tercile. The bottom panel shows that there are no obvious differences in the average improvements between typing terciles. It is however worth noting that the sample groups here are smaller than for mathematics and demographic subgroups, as 19 to 21 percent of learners have missing baseline information for the English, Raven's and typing tests.

The Raven's, English Literacy and typing test scores were collected in an after-school setting and as a result the proportion of missing observations was much higher than for the in-school tests. The full set of baseline Raven's, English and typing results was only available for 366 learners in the evaluated sample. A further 11 of these 366 learners had missing baseline or endline mathematics test scores. Missing observations on various baseline test result were not missing completely at random[55]. As such, changes in sample size are likely to result in changes in the magnitude of the treatment effect, simply because the two groups are not comparable. Therefore the sample in Table 23 and Table 24 is limited to only look at the 355 learners with full information on all baseline tests as well as the endline mathematics test. Column 1 in both tables has no subgroups or interaction terms, and therefore presents the magnitude of the treatment effect when the sample is restricted to only the 355 learners that are being studied in this section.

[53] See the discussion about inter-school and intra-school trends looking at the probability of a learner applying in: Section 5: Experimental Design, Learner Application and Selection, Applicant Characteristics.

[54] Neither were a greater number of subdivisions such as quartiles or quintiles used. Due to the smaller sample size, it would have been less likely that a significant effect would have been observed, as the subgroup sizes would have been relatively small. For example if the sample had been divided into quintiles, the average number of learners within each subgroup would have been about 35.

[55] Learners who had not written the baseline Raven's, English and typing tests scores 5 percentage points lower on the baseline mathematics test, significant at the 5 percent level. Coloured learners and older learners were also more likely to have missed the tests. See Figure B5 and B6 in Appendix B for further detail.

Table 22: Table comparing the mean baseline and endline Section A mathematics scores of control and treatment for various subgroups.

| | Control | | Treatment | | Difference *(Endline - Baseline)* | |
|---|---|---|---|---|---|---|
| | Baseline | Endline | **Baseline** | **Endline** | Control | Treatment |
| **Top Panel: Baseline English Literacy - Terciles** | | | | | | |
| 1 (Bottom Tercile) | 0.36 | 0.38 | **0.38** | **0.45** | 0.02 | 0.07 |
| | (0.14) | (0.16) | **(0.13)** | **(0.17)** | (0.12) | (0.13) |
| 2 | 0.49 | 0.53 | **0.47** | **0.58** | 0.04 | 0.10 |
| | (0.14) | (0.18) | **(0.16)** | **(0.18)** | (0.12) | (0.11) |
| 3 (Top Tercile) | 0.60 | 0.63 | **0.59** | **0.70** | 0.02 | 0.11 |
| | (0.13) | (0.14) | **(0.16)** | **(0.16)** | (0.09) | (0.1) |
| **Middle Panel: Baseline Raven's Score - Terciles** | | | | | | |
| 1 (Bottom Tercile) | 0.37 | 0.38 | **0.36** | **0.43** | 0.01 | 0.06 |
| | (0.15) | (0.16) | **(0.14)** | **(0.16)** | (0.11) | (0.12) |
| 2 | 0.48 | 0.51 | **0.49** | **0.58** | 0.04 | 0.09 |
| | (0.15) | (0.17) | **(0.15)** | **(0.17)** | (0.11) | (0.12) |
| 3 (Top Tercile) | 0.57 | 0.60 | **0.59** | **0.72** | 0.03 | 0.12 |
| | (0.15) | (0.19) | **(0.16)** | **(0.14)** | (0.11) | (0.11) |
| **Bottom Panel: Baseline Typing (100 Correct keystrokes per minute) - Terciles** | | | | | | |
| 1 (Bottom Tercile) | 0.39 | 0.41 | **0.40** | **0.48** | 0.01 | 0.07 |
| | (0.14) | (0.16) | **(0.16)** | **(0.19)** | (0.11) | (0.13) |
| 2 | 0.47 | 0.51 | **0.47** | **0.59** | 0.05 | 0.11 |
| | (0.18) | (0.20) | **(0.17)** | **(0.18)** | (0.13) | (0.13) |
| 3 (Top Tercile) | 0.56 | 0.59 | **0.55** | **0.65** | 0.03 | 0.09 |
| | (0.17) | (0.18) | **(0.17)** | **(0.18)** | (0.09) | (0.09) |
| **Number of observations** | 184 | 182 | **197** | **193** | 381 | 375 |

Table 23 looks at the effects of Raven's, English and typing tests and the interaction of treatment and Raven's, English and Typing tests on Section A endline mathematics scores, whilst Table 24 does the same for core Grade 8 curriculum, tested in Section B at endline

In Rows 2 the sets of English subgroups and their interaction terms are included in the regression. Row 3 looks at the effects of the two Raven's subgroups on the treatment effect, whilst the two sets of subgroups of baseline typing ability and their interaction terms are added to the basic regression and the results are reported in column 4.

> ***Result 12: Neither general cognitive ability, English ability nor typing speed at baseline have a significant differential treatment effect on basic numeracy results.***

Row 1 of Table 23 shows that the estimate on treatment is not altered significantly when any of the subgroup dummies and interaction terms are included in the regressions. Additionally the coefficients on the interaction terms are all small and statistically insignificant. There is therefore no evidence to suggest that baseline cognitive skills, English ability or typing speed alter the average improvement expected on basic numeracy due to treatment.

Table 23: OLS Estimates showing the effect of treatment and the interaction of treatment with baseline measures of English Literacy, general Cognitive ability and typing speed on endline Section A mathematics test scores

| | Endline Mathematics Test Scores – Section A | | | |
|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) |
| Treatment | 0.068*** | 0.062*** | 0.051** | 0.066*** |
| | (0.011) | (0.020) | (0.021) | (0.020) |
| English - Tercile 2 | | 0.043* | | |
| | | (0.023) | | |
| English - Tercile 3 | | 0.026 | | |
| | | (0.021) | | |
| English - Tercile 2 × Treatment | | -0.017 | | |
| | | (0.031) | | |
| English - Tercile 3 × Treatment | | 0.027 | | |
| | | (0.025) | | |
| Raven's - Tercile 2 | | | 0.014 | |
| | | | (0.020) | |
| Raven's - Tercile 3 | | | 0.029 | |
| | | | (0.021) | |
| Raven's - Tercile 2 × Treatment | | | 0.022 | |
| | | | (0.027) | |
| Raven's - Tercile 3 × Treatment | | | 0.032 | |
| | | | (0.029) | |
| Typing - Tercile 2 | | | | 0.030 |
| | | | | (0.022) |
| Typing - Tercile 3 | | | | 0.034 |
| | | | | (0.023) |
| Typing - Tercile 2 × Treatment | | | | 0.009 |
| | | | | (0.030) |
| Typing - Tercile 3 × Treatment | | | | -0.005 |
| | | | | (0.025) |
| Constant | 0.349** | 0.387*** | 0.389*** | 0.325** |
| | (0.140) | (0.141) | (0.135) | (0.139) |
| Number of Observations | 355 | 355 | 355 | 355 |
| R-squared | 0.753 | 0.755 | 0.753 | 0.754 |

Robust standard errors in parentheses. Controls include: a dummy variable for each school. Learner biographic details: gender, age, race and home language; baseline English, Raven's and typing scores and a dummy for missing observation on the English literacy, Raven's or typing test at baseline. *** p<0.01, ** p<0.05, * p<0.1

***Result 13: Treatment learners with a medium to high level of baseline English literacy improve significantly more on core Grade 8 curriculum compared to treatment learners with poor baseline English scores.***

Table 24 reports on the same set of regressions as Table 23, but with Section B endline mathematics results as the outcome variable as opposed to Section A. The conclusions here are slightly different. In Column 3, upon the inclusion of the two dummies for middle and top baseline English terciles and the interaction terms of middle and top English terciles with treatment, the treatment effect decreases to an insignificant 0.9 percentage points, which is also significantly different to the estimate on treatment reported in Column 1[56]. Additionally, the estimate on the interaction terms between the middle tercile and top terciles and treatment are large, at 5.9 and 4.6 percentage points respectively. They are both statistically significant at the 10 percent level. Thus, holding other baseline characteristics constant, on average, treatment learners in the middle third of the baseline English distribution are scoring 5.9 percentage points higher, on average, than a learner on treatment that scored in the bottom third for English at baseline, whilst a top English tercile learner is scoring 4.6 percentage points higher, on average, holding all else constant.

It is plausible that there is a similar underlying pattern for baseline cognitive skills which cannot be detected due to a small sample size. In column 3 of Table 24, the treatment effect is also lower, at 2.2 percentages points, and insignificant. However, neither one of the interaction terms of the Raven's baseline terciles with treatment are significant. Furthermore, the estimate of 2.2 percentage points is not significantly different to 4.2 percentage points (p-value = 0.223) reported in Column 1. Therefore, there is insufficient evidence to conclude that the treatment effect differs for learners at different levels of cognitive development at baseline.

There is no evidence of any differential effect of the programme on core Grade 8 curriculum outcomes for learners with different typing speeds at baseline. This can be seen in Column 4 as both of the estimates on the interaction terms are small and highly insignificant. Additionally, there is no change to the size of the treatment effect.

[56] Assuming a null hypothesis that there has been no change in the treatment effect from column 1 to column 2 ( $H_0$: $\beta_{treatment} = 0.042$) produces a p-value of 0.0673. Namely the treatment effect in column 2 of Table 24, is significantly different to the original treatment effect at the 7 percent level.

Table 24: OLS estimates showing the effect of treatment and the interaction of treatment with baseline measures of English Literacy, general cognitive ability and typing speed on endline Section B mathematics scores

| | Endline Mathematics Test Scores – Section B | | | |
|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) |
| Treatment | 0.042*** | 0.009 | 0.022 | 0.043** |
| | (0.012) | (0.018) | (0.017) | (0.021) |
| English - Tercile 2 | | -0.006 | | |
| | | (0.020) | | |
| English - Tercile 3 | | 0.026 | | |
| | | (0.020) | | |
| English - Tercile 2 × Treatment | | 0.059* | | |
| | | (0.031) | | |
| English - Tercile 3 × Treatment | | 0.046* | | |
| | | (0.027) | | |
| Raven's - Tercile 2 | | | 0.005 | |
| | | | (0.017) | |
| Raven's - Tercile 3 | | | 0.059*** | |
| | | | (0.020) | |
| Raven's - Tercile 2 × Treatment | | | 0.038 | |
| | | | (0.024) | |
| Raven's - Tercile 3 × Treatment | | | 0.031 | |
| | | | (0.031) | |
| Typing - Tercile 2 | | | | 0.024 |
| | | | | (0.022) |
| Typing - Tercile 3 | | | | 0.029 |
| | | | | (0.025) |
| Typing - Tercile 2 × Treatment | | | | -0.006 |
| | | | | (0.030) |
| Typing - Tercile 3 × Treatment | | | | 0.002 |
| | | | | (0.028) |
| Constant | 0.067 | 0.104 | 0.154 | 0.053 |
| | (0.144) | (0.143) | (0.139) | (0.143) |
| Number of Observations | 355 | 355 | 355 | 355 |
| R-squared | 0.697 | 0.699 | 0.702 | 0.697 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include: linear and a squared term of baseline mathematics Section A test scores along with a dummy if the baseline mathematics score was missing, dummy variables for the schools, gender, age, race and home language as well as a dummy variable indicating if there was a missing item for gender, age or race, home language.

## Robustness

There are a number of general concerns regarding the robustness of the results. Three such concerns are addressed in this section. The first is that spillover might have occurred. The second is that treatment and control learners may have been motivated differently during the endline mathematics test. The third deals with the possibility that one or two schools are outliers that are solely responsible for the positive treatment effect observed.

### Spillover

Spillover could occur through three different channels. Firstly, if teachers learned about this freely available resource through the intervention and as a result had decided to use the material in class, either directly or indirectly, this may have resulted in some spillover. Whilst it is possible that this might have occurred, past research has shown that changing teacher behaviour and their pedagogical approach is difficult. Even with substantial support and training, teachers often prefer to continue using methods with which they are familiar (Barrara-Osorio and Linden, 2009). It is also highly probable that some contamination of the control group took place due to daily social interactions of control learners with treatment learners in class, or after-school, either because treatments learners tutored their friends or because a change in approach or attitude in the treatment learners impacted the outlook or behavior of control learners. The third possibility is that control learners were introduced to the website by Numeric or by learners on the programme and as a result started to use this free online resource. In all three cases, spillover is most likely to cause an underestimate of the treatment effect, suggesting that in the presence of spillover, a 0.32 standard deviations improvement would then be the lower bound of the true effect size.

Knowledge gained in an after-school setting or a change in attitude may have been transferred to learners who are not on the programme during mathematics class or within a social circle. Such interactions were accounted for by establishing which other learners from the treatment and control groups were in the peer group of learners in the evaluated sample. The learners were asked to name five best friends in their grade at school as well as the person or people they sat next to in their school mathematics class. If the learners had any study partners, homework buddies or family members in their grade at school these were recorded as well. Connections between learners in the evaluated sample were then identified.

A summary is provided in Table 25. The numbers reported in the table are the mean number of learners. For example in Column 1, the first cell reports that a control learner has an average of 0.83 best friends in the control group. Whist Row 2 shows that control learners have a mean number of 0.93 friends who are in the treatment group[57]. Row 1 and 2 of Table 25 show that both treatment and control learners are more likely to report having a treatment learner as a friend than they are to report having a control learner as a friend.

[57] Each learner was asked to name 5 best friends. Thus, for every 10 control learners that were asked to name 5 best friends, 50 Grade 8 learners would have been mentioned. Of those 50, a mean of 8.3 learners would have been control group learners, whereas 9.3 would have been in treatment.

There are only two significant differences between treatment and control in the table. Treatment learners reported an average number of 0.95 mathematics neighbours from treatment, whilst control learners only reported an average of 0.627 mathematics neighbours from the treatment group. It is possible that much of this difference is due to learners in treatment misunderstanding the question and taking it to mean neighbours in Numeric mathematics classes[58].

The second significant difference in the table between treatment and control is the mean number of study partners who are in the control group. Ten treatment learners, on average report having one more control group study partner than ten control group learners.

Table 25: Summary of the mean number of best friends, neighbours in mathematics class and study partners in the evaluated sample for the control and treatment groups.

| | Control | Treatment | Total | Significant Difference *(Treatment-Control)* |
|---|---|---|---|---|
| Mean Number of Best Friends in Control | 0.835<br>(0.789) | 0.849<br>(0.802) | 0.842<br>(0.794) | - |
| Mean Number of Best Friends in Treatment | 0.934<br>(0.937) | 1.095<br>(1.192) | 1.012<br>(1.069) | - |
| Mean Number of Maths Neighbours in Control | 0.533<br>(0.698) | 0.603<br>(0.73) | 0.567<br>(0.714) | - |
| Mean Number of Maths Neighbours in Treatment | 0.627<br>(0.771) | 0.950<br>(1.154) | 0.783<br>(0.987) | *** |
| Mean Number of Study Partners in Control | 0.283<br>(0.538) | 0.382<br>(0.67) | 0.331<br>(0.607) | * |
| Mean Number of Study Partners in Treatment | 0.434<br>(0.735) | 0.407<br>(0.823) | 0.421<br>(0.778) | - |
| Mean Number of Family Members in Control | 0.009<br>(0.097) | 0.025<br>(0.157) | 0.017<br>(0.13) | - |
| Mean Number of Family Members in Treatment | 0.019<br>(0.136) | 0.02<br>(0.141) | 0.019<br>(0.138) | - |

Notes: Number of observations is 410. Column 1 and column 2 have standard deviations in parentheses. Statistical significance of the difference between treatment and control is indicated by *** p<0.01, ** p<0.05, * p<0.1 in the third column.

The second sphere of influence considered was a learner's school class. This was a likely source of spillover, both because of the large amount of time learners in the same class spend together, and also because it meant they would attend school mathematics classes together. There were 48 classes in total across the nine schools. The maximum number of Grade 8 classes per school was seven and the minimum number was four. Table 26 shows that the average class size was 42.4 learners, and the mean number of evaluated sample learners in each class was just below 10. This meant that there were an average of about

[58] The question asked: "Who do you sit next to in your mathematics class at school". The meaning of this question is clear to control learners. However treatment learners have both school and Numeric mathematics classes at school. Although the interviewers did specifically mention that this did not include Numeric classes, it is possible that a non-negligible portion of learners still included neighbors from their after school classes. As the subsequent analysis is restricted to control learners only, the possible difference in interpretation of the questions by treatment learners does not affect the analysis.

5 treatment and 5 control learners in each class, which is equivalent to 23.4 percent of the class. The class with the highest proportion of evaluated sample learners had 50 percent of the class in treatment or control, whilst there was also one class where not a single learner had applied, and as such, none of the learners were in the evaluated sample.

Table 26: Table to show the total number of learners as well as treatment and control learners in the 48 classes at participating schools

| | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| Number of Treatment Learners in Class at School | 4.917 | 2.751 | 0 | 13 |
| Number of Control Learners in Class at School | 4.917 | 2.664 | 0 | 11 |
| Number of Evaluated Sample Learners (Treatment + Control) in Class at School | 9.833 | 4.844 | 0 | 20 |
| Class Size | 42.354 | 5.905 | 34 | 59 |
| Proportion of Class in Treatment | 0.117 | 0.067 | 0.000 | 0.325 |
| Proportion of Class in Control | 0.117 | 0.065 | 0.000 | 0.256 |
| Proportion of Class in Evaluated Sample (Treatment+ Control) | 0.235 | 0.118 | 0.000 | 0.500 |

In Table 27 four specifications are used to try and pick up possible spillover to the control group. As such, the sample in these eight regressions consists of control group learners only. The first specification looks at the number of best friends from the control and the treatment group. Leaners that applied to the programme were normally among the higher achieving learners within a school. As such, a learner with more applicant friends (regardless of whether they are in treatment or control), may have performed better as they had more high achieving, diligent or motivated friends who may have been more likely to endorse academic activities. Thus, the number of control friends along with the number of treatment friends was included, to account for this possibility. The second specification does the same for neighbours in mathematics class. The next two specifications use class level variables. In the third Column, class size, the number of treatment learners and the number of learners in the evaluated sample in the class are included. Class size itself could be affecting mathematics outcomes of the learners in the class and therefore it was included. The last specification only looks at the proportion of treatment and the proportion of evaluated learners in a class. This last measure was included because in classes with a large number of treatment learners, this could lead to a change in general attitude of the learners or the teacher. The variables quantifying the number and proportion of total evaluated learners within a class are also included to check that this proportion is not just correlated to a greater number of high performing students in a class as a result of streaming within the school[59].

[59] The approach used in columns 3, 4, 7 and 8 is based on the method used by Miguel and Kremer (2004).

Table 27: OLS Regression estimates showing the effects of having a best friend, or a neighbour in mathematics class, from treatment or control

| | Endline Mathematics Test Scores - Section A | | | | Endline Mathematics Test Scores - Section B | | | |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Mean Number of Best Friends in Control | -0.001 | | | | -0.008 | | | |
| | (0.010) | | | | (0.010) | | | |
| Mean Number of Best Friends in Treatment | 0.001 | | | | -0.002 | | | |
| | (0.008) | | | | (0.009) | | | |
| Mean Number of Mathematics Neighbours in Control | | 0.017* | | | | 0.014 | | |
| | | (0.010) | | | | (0.014) | | |
| Mean Number of Mathematics Neighbours in Treatment | | 0.010 | | | | -0.003 | | |
| | | (0.014) | | | | (0.010) | | |
| Number of Treatment Learners in Class at School | | | 0.002 | | | | 0.001 | |
| | | | (0.007) | | | | (0.006) | |
| Number of Evaluated Sample (Treatment + Control) Learners in Class at School | | | 0.002 | | | | 0.004 | |
| | | | (0.005) | | | | (0.004) | |
| Class Size | | | 0.003 | | | | 0.004 | |
| | | | (0.006) | | | | (0.004) | |
| Proportion of Class in Treatment | | | | 0.090 | | | | 0.081 |
| | | | | (0.285) | | | | (0.250) |
| Proportion of Class in Evaluated Sample (Treatment + Control) | | | | 0.070 | | | | 0.154 |
| | | | | (0.193) | | | | (0.154) |
| Constant | 0.656*** | 0.617*** | 0.461 | 0.582*** | 0.392** | 0.337* | 0.090 | 0.257 |
| | (0.211) | (0.211) | (0.352) | (0.192) | (0.194) | (0.192) | (0.273) | (0.167) |
| Number of Observations | 199 | 199 | 221 | 221 | 199 | 199 | 221 | 221 |
| R-squared | 0.720 | 0.724 | 0.728 | 0.727 | 0.643 | 0.645 | 0.672 | 0.670 |

Notes: Sample restricted to control learners only. Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include: linear and a squared term of baseline mathematics Section A test scores along with a dummy if the baseline mathematics score was missing, dummy variables for the schools, gender, age, race and home language , and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for gender, age or race, home language, English literacy, Raven's or typing test at baseline.

Column 1 and Column 5 of Table 27 show that having a best friend in treatment or control had no significant impact on the endline mathematics test scores of control learners. Oddly, the estimate on number of control mathematics neighbours is positive and significant at the 5 percent level, whilst the estimate on number of mathematics neighbours in treatment is smaller and insignificant. This phenomenon can be seen in Column 2. Here an additional mathematics neighbour from the control group, all else constant, would raise endline numeracy scores by 1.7 percentage point, whilst having a neighbor in treatment raised the average numeracy score by 1.0 percentage points. However, the difference between the estimate on the number of control neighbours and the estimate on the number of treatment neighbours is highly insignificant. This same effect is observed in column 6 on Section B endline results. Here, Section B results are an average of 1.4 percentage points higher on average for every additional control group mathematics neighbor, although the effect is highly insignificant. This results support the claim that there is negligible spillover taking place in mathematics class as a result of the programme. In fact, learners with one additional control group neighbor perform marginally better on basic numeracy than control learners with on additional treatment neighbour. However, this suggests that being in a class with a greater number of evaluated sample learners might be correlated with higher mathematics outcomes.

Column 3 and Column 7 show that the coefficient on the number of treatment learners as well as the number of evaluated sample learners is large and positive for both basic numeracy as well as core grade 8 curriculum, when class size is included as a variable. However, neither the number of learners in treatment nor the number of evaluated sample learners in a class has any significant effect on outcomes for control learners. Additionally, although it is insignificant, the positive coefficient on class size is an unexpected result.

The estimates in Column 4 and Column 8 are practically large. A control learner in class in which 10 percent of learners are in treatment scores 0.9 percentage points higher on Section A and 0.81 percentage points higher on Section B than an equivalent learners in a class with zero treatment learners. Additionally, a control learners in a class where 25 percent of learners are in the evaluated sample scores 1.4 percentage points higher on basic numeracy and 3.1 percentage points higher than a control group learner in a class where only 5 percent of learners are in the evaluated sample. It is evident that for control learners, the proportion of treatment and evaluated sample learners in their class is positive, but has no statistically significant effect on their endline outcomes. Altogether the results in Table 27 demonstrates that with this small sample of between 199 and 221 learners there is no concrete evidence to show that there is much spillover taking place via treatment friends. It does however not exclude the possibility that there could be some social, in-class interactions between control and treatment learners that are resulting in positive spillover at the classroom level.

Table 28 investigates a second possible route through which spillover can occur, the Khan Academy website. At endline learners were asked whether they had had any contact with Khan Academy. The results are shown in Table 28. It is clear that a very high proportion of treatment learners, 94.1 percent, have used the Khan Academy website at some point during the year. An e-mail account linked to their Khan Academy accounts was created for every treatment learner in January, so correspondingly 87.8 percent also state that they have a Khan Academy account. However, only half of the learners seem to know that this account is also an e-mail address, underlining that despite greater access to computers and the internet, learners had not necessarily picked up a wider range of computer skills. Additionally, almost 74 percent report that they had watched a video or done at least one exercise on the website in the fourth term, which roughly corresponds to the proportion of learners that remained on the programme for the entire year.

The proportion of control group learners that reported having used the Khan Academy website is high. Just fewer than 53 percent of control learners replied saying that they had watched a video or done an exercise on the Khan Academy website at some point during the year, whilst 27 percent say that they have a Khan Academy account. Whilst it is likely that these learners are infrequent users of the site, it nevertheless supports the notion that that there may have been some spillover to the control group through this medium.

Table 28: Khan Academy website use and existence of online accounts for learners in treatment and control at endline

| VARIABLES | Control | Treatment | Significant Difference *(Treatment-Control)* |
|---|---|---|---|
| Used Khan Academy Website | 0.527 | 0.941 | *** |
| | (0.500) | (0.237) | |
| Used Khan Academy Website in the 4th Term of 2013 | 0.270 | 0.738 | *** |
| | (0.445) | (0.441) | |
| Khan Academy Account | 0.120 | 0.878 | *** |
| | (0.326) | (0.328) | |
| Own e-mail Address | 0.359 | 0.479 | ** |
| | (0.554) | (0.565) | |
| Facebook Account | 0.466 | 0.426 | - |
| | (0.525) | (0.512) | |
| **Number of observations** | **238** | **236** | |

Notes: Column 1 and column 2 have standard deviations in parentheses. Statistical significance of the difference between treatment and control is indicated by *** p<0.01, ** p<0.05, * p<0.1 in the third column.

### Motivation

The mathematics test was designed to test mathematics proficiency. A learner's performance on the test however, is determined by both their aptitude and effort. If the level of effort the learner exerts in taking the test is correlated with treatment, this may result in bias. This is a concern as a number of the learners may have been aware that the tests were linked with the Numeric programme. This may have induced higher effort in the treatment group or might have lead to lower motivation in the control group as they were rejected at the beginning of the year. Here the omitted variable, motivation, was hypothesized to be positively correlated with treatment. A differential response of this nature would have resulted in an artificially high treatment effect.

The approach taken to eliminate some of the bias in the measure due to differential effort of the learners, was the introduction of incentivised testing. Here the learners in the evaluated sample were also given a separate test at endline in which they were incentivised. In the incentivised test, learners received a small monetary payment in the form of airtime for correct answers, thereby reducing the role of intrinsic motivation in test taking as a determinant of a learner's final mathematics score. The incentivised test was constructed by selecting 22 questions from Section A of the endline mathematics test and changing the numbers in the examples, keeping the question structure the same[60]. Twenty-one[61] of these questions were

[60] This same technique of keeping the structure of the question the same and altering the numbers, which was used to maintain consistent standards between baseline, midline and endline, was applied here. For example Question 1.6 on the incentivised test asked the learner to calculate 0.6×4, whilst the corresponding question in the non-incentivised test asked the learners to find 0.3×0.8.

[61] One of the questions that was included in the incentivised test was altered slightly, when a denominator was changed on a fraction question. This meant that the question on the incentivised test was much more difficult and

set at the same level of difficulty and tested the exact same competencies that were tested in Section A. Thus, a subset of endline Section A questions can be matched exactly with questions asked on the incentivised test. These two tests were both held at the end of the year with the non-incentivised test preceding the incentivised test by no more than 10 days[62].

Table 29 presents a comparison of the treatment effect on the twenty-one matched questions from Section A in Columns 1 to 3 and the treatment effect measured on the incentivised test in Columns 4 to 6. There were a total of 413 learners (210 treatment and 203 control learners) who wrote both the incentivised and the non-incentivised test. Only data from these 413 observations is used in the regressions shown in Table 29. The treatment effect is large and significant under all three specifications using either the incentivised or non-incentivised test. A closer examination of the effect sizes of treatment shows that the treatment effect in a non-incentivised setting appears to be about 1 percentage point higher, on average, than the treatment effect measured on the incentivised test. For example, the treatment effect in Column 5 is 6.7 percentage points, whilst the corresponding treatment effect on matched questions from Section A suggests a higher treatment effect of 7.7 percentage points. Whist the difference between these two estimates is highly statistically insignificant; it nevertheless suggests the possibility that there might have been a small overestimation of the treatment effect due to differential motivation in test taking in the treatment and control groups[63].

therefore no longer directly comparable to the original question. As a result this question is excluded and the comparison is made using the remaining 21 questions that were asked in the incentivised test.

[62] All of the incentivised tests were intentionally scheduled to take place after the non-incentivised tests. This was done as the non-incentivised tests were written during school time, whilst the incentivised tests often took place after school. Thus the number of missing observations for the non-incentivised test was likely to be higher and test taking conditions were likely to vary more. Additionally, at some schools, if learners had already taken one test that they associated with Numeric, they might have been more likely to be absent for the endline exam. The non-incentivised test was viewed as the main measure, whilst the incentivised test was used mainly as a robustness check. As such the main measure was prioritized and this test was run first to ensure that the number of missing observation were kept to the absolute minimum.

It is possible that some learning could have taken place at endline in the non-incentivised test that could have boosted learner's marks on the incentivised test. However, when the treatment learners that wrote both tests are compared, this group of learners scored 50 percent on average for both tests. It is also not obvious why such a learning effect would have been different between treatment and control learners. If learning took place, then a direct comparison of the two tests might not be viable, however the comparison between control and treatment is still valid.

Another potential issue as a result of the ordering of the tests is that learners may have experienced some exam fatigue. They were rewriting a mathematics test which was similar to one they had already written. However because the second test was incentivised and also shorter, this is unlikely to have been as much of a problem. Again, there is no reason to expect that such a test fatigue effect would have affected control and treatment learners differently.

[63] Any difference in motivation could, of course, also be side-effect of the programme. In such a case it should be arguably be included in the treatment effect.

Table 29: OLS estimates comparing the treatment effect observed on the incentivsed test and on the non-incentivised Section A endline mathematics test questions

| | Endline - 21 Section A Questions Non-Incentivised Test | | | Endline Incentivised Mathematics Test Scores | | |
|---|---|---|---|---|---|---|
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.082*** | 0.077*** | 0.080*** | 0.073*** | 0.067*** | 0.069*** |
| | (0.024) | (0.013) | (0.013) | (0.024) | (0.013) | (0.013) |
| Baseline - Section A | | 0.817*** | 0.657*** | | 1.071*** | 0.953*** |
| | | (0.171) | (0.185) | | (0.160) | (0.170) |
| Constant | 0.082*** | 0.077*** | 0.080*** | 0.073*** | 0.067*** | 0.069*** |
| | (0.024) | (0.013) | (0.013) | (0.024) | (0.013) | (0.013) |
| Observations | 413 | 413 | 413 | 413 | 413 | 413 |
| R-squared | 0.028 | 0.705 | 0.748 | 0.023 | 0.694 | 0.728 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls in column 2 and column 3 include: A squared term of baseline mathematics Section A test scores along with a dummy for a missing baseline mathematics score, and dummy variables for the schools. Column 3 and column 6 also introduce the controls: gender, age, race and home language, and RCE controls which have the baseline results on the English, Raven's and typing tests as well as a dummy variable indicating if there was a missing observation for gender, age or race, home language, English literacy, Raven's or typing test at baseline.

### School Level Effect

The Numeric programme encompassed nine schools in 2013. This section checks to see if the overall treatment effect was dependent on the performance of individual schools. Table 30 shows that the difference in the average endline Section A results of control and treatment learners in different schools was large. Contrast school 104, where both control and treatment learners obtained an average of 58 percent at endline, with school 106 and 107 where the average scores of treatment learners is 10 percent higher than the control group average at endline.

Table 30: Table comparing the mean baseline and endline Section A mathematics scores of control and treatment in each of the nine schools.

| School | Control | | Treatment | | Difference (Endline - Baseline) | | Number of Observations |
|---|---|---|---|---|---|---|---|
| | Baseline | Endline | **Baseline** | **Endline** | Control | Treatment | |
| 101 | 0.54 | 0.52 | **0.51** | **0.55** | -0.02 | 0.04 | **37** |
| | (0.15) | (0.18) | **(0.14)** | **(0.15)** | (0.09) | (0.08) | |
| 102 | 0.44 | 0.45 | **0.45** | **0.50** | 0.00 | 0.05 | **37** |
| | (0.11) | (0.14) | **(0.16)** | **(0.17)** | (0.10) | (0.10) | |
| 104 | 0.49 | 0.58 | **0.47** | **0.58** | 0.09 | 0.09 | **38** |
| | (0.18) | (0.19) | **(0.14)** | **(0.14)** | (0.15) | (0.08) | |
| 105 | 0.42 | 0.43 | **0.39** | **0.47** | 0.01 | 0.07 | **77** |
| | (0.16) | (0.18) | **(0.13)** | **(0.17)** | (0.10) | (0.11) | |
| 106 | 0.39 | 0.37 | **0.40** | **0.47** | -0.02 | 0.07 | **75** |
| | (0.12) | (0.12) | **(0.13)** | **(0.17)** | (0.09) | (0.15) | |
| 107 | 0.34 | 0.40 | **0.34** | **0.50** | 0.06 | 0.16 | **36** |
| | (0.10) | (0.14) | **(0.13)** | **(0.24)** | (0.14) | (0.14) | |
| 109 | 0.56 | 0.58 | **0.56** | **0.62** | 0.03 | 0.05 | **29** |
| | (0.14) | (0.17) | **(0.15)** | **(0.16)** | (0.11) | (0.09) | |
| 110 | 0.38 | 0.38 | **0.40** | **0.43** | 0.00 | 0.01 | **36** |
| | (0.18) | (0.16) | **(0.18)** | **(0.16)** | (0.11) | (0.07) | |
| 111 | 0.60 | 0.67 | **0.63** | **0.76** | 0.07 | 0.13 | **80** |
| | (0.14) | (0.14) | **(0.16)** | **(0.14)** | (0.08) | (0.09) | |
| **Number of Observations** | 231 | 227 | **231** | **226** | 223 | 222 | **445** |

Thus, in order to verify that a significant treatment effect is not dependent upon the inclusion of a particular school, the key treatment specification from Table 12 was re-run 9 times. Each time one of the schools was excluded from the evaluated sample. Table 31 presents a summary of the results of these 18 regressions. In Row 1, for example, all learners from school 101 were dropped and the treatment effect on basic numeracy and core Grade 8 curriculum was calculated for the remaining 8 schools. The coefficient on treatment remains highly significant in all cases, even after the reduction in sample size. Neither does the magnitude of the estimate on treatment change significantly. The treatment estimate for Section A, ranges between 5.1 and 6.5 percentage points, whilst for Section B the lowest treatment effect obtained was 3.5 percent whilst the highest was 4.4 percent. This shows that the treatment effect was not the product of a huge success in only one or two of the schools, but rather that the average effect of treatment was relatively consistent across schools.

Table 31: Table showing the regression estimates of treatment on endline mathematics test scores for Section A and Section B when each of the nine schools are excluded from the sample.

| | **Endline Mathematics Test Scores – Section A** | | | **Endline Mathematics Test Scores – Section B** | | | **Number of Observations** |
|---|---|---|---|---|---|---|---|
| | Treatment | Standard Errors | R-squared | Treatment | Standard Errors | R-squared | |
| Without School 101 | 0.062*** | (0.010) | 0.743 | 0.044*** | (0.011) | 0.670 | 415 |
| Without School 102 | 0.060*** | (0.010) | 0.748 | 0.042*** | (0.011) | 0.684 | 415 |
| Without School 104 | 0.065*** | (0.010) | 0.756 | 0.042*** | (0.011) | 0.673 | 415 |
| Without School 105 | 0.059*** | (0.010) | 0.753 | 0.037*** | (0.012) | 0.681 | 376 |
| Without School 106 | 0.051*** | (0.010) | 0.757 | 0.037*** | (0.011) | 0.693 | 375 |
| Without School 107 | 0.057*** | (0.010) | 0.754 | 0.035*** | (0.011) | 0.683 | 417 |
| Without School 109 | 0.060*** | (0.010) | 0.743 | 0.036*** | (0.011) | 0.668 | 423 |
| Without School 110 | 0.063*** | (0.010) | 0.738 | 0.040*** | (0.011) | 0.668 | 415 |
| Without School 111 | 0.055*** | (0.011) | 0.669 | 0.036*** | (0.011) | 0.549 | 373 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1. Controls include a linear and a squared term of baseline mathematics Section A test scores along with a for a missing baseline mathematics score, dummy variables for the schools, gender, age, race and home language and RCE controls which have the baseline results on the English, Raven's and typing tests as well as dummy variables indicating if there was a missing observation for the English literacy, Raven's or typing test at baseline.

## 10. Discussion

Introducing computer assisted learning into schools is one of the proposed means of raising the quality of instruction in developing countries, by simultaneously addressing any specific gaps in knowledge that a learner might have as a result of absenteeism and disruptions during teaching time. The Numeric programme is one such programme that used CAL in the form of Khan Academy to revise and teach basic numeracy. This thesis analysed the effects of the year-long pilot programme in 9 schools in the Cape Town Metro on standardised mathematics test results.

### Summary of Results

The endline mathematics results showed that the Numeric programme was successful at improving the basic numeracy (Section A) of learners selected into the treatment group by an average of 6.0 percentage points (0.32 s.d.) above the mean scores obtained by learners in the control group. Thus, the chief aim of increased numeracy was achieved. Furthermore, learners in treatment also outperformed learners in the control group by a mean of 4.1 percentage points (0.25 s.d) in Section B, which tested Grade 7 and 8 curriculum material. In this manner CAL was being used to increase the rate at which later learning could occur. Moreover, treatment learners were achieving higher average scores in every subsection or topic area tested at endline, showing that the programme did not focus on teaching a particular topic or set of concepts, but rather improved numeracy in general and allowed the learners to revise fundamental concepts and master a broad range of new skills.

This marked improvement as a result of selection into treatment suggests that learners' mathematics ability has objectively improved, whilst learner feedback shows that learners are aware of this improvement in their mathematical ability. Furthermore, learner feedback shows that learners on the programme are also enjoying mathematics more than learners in the control group. CAL has the added benefit in that gamefication and instant feedback enhance the learning experience which could have resulted in a more enjoyable experience (Schonfield, Eurich-Fulcer & Britt, 1994). This increase in enjoyment could also be because the learners were conscious of the fact that their competency had improved and were proud of their accomplishments. The learners reported that Numeric classes had been useful in mathematics classes at school and so it could simply be that mathematics is no longer perceived to be as difficult.

The effect of treatment was also found to be reasonably consistent across learners, regardless of their baseline mathematics ability, gender, race, home language, school, attitude towards mathematics at baseline and baseline typing speed. It is encouraging that the programme does not seem to target a particular group, but rather improves outcomes for a diverse set of learners. However, there appeared to be a differential treatment effect by baseline English ability, but only in the Section B results. Treatment learners in the middle and top English tercile improved an average of about 5 percentage points more than treatment learners in the bottom English tercile, significant at the 10 percent level. It is possible that learners who were already proficient at English were able to benefit the most from the programme. Furthermore, learners in the middle tercile may have benefitted from the programme both in terms of an improvement in English competence, as well as through enhanced mathematical ability. Learners on the programme whose home language was not English would have had three additional hours of hearing and reading English a week, and the type of vocabulary they would have come across would have been particularly useful in mathematics. This improvement in English ability may then have benefited these learners in their Mathematics classes at school, where the medium of instruction is English, as well as aiding them in understanding textbooks and questions in examinations.

There was little evidence to suggest that there was spillover from the programme to learners in the control group. However, a high number of control learners, 53 percent, said that they had at some point made use of Khan Academy. Although it is unlikely that these leaners will have used the website frequently or consistently enough to improve mathematics outcomes, as only 34 percent of control learners have internet access on a computer at home and Khan Academy is not yet used at schools. The high number of students who have accessed the website is in itself a positive outcome, as it suggests that the existence of such the programme is a non-threatening way of raising awareness about Khan Academy and online academic resources more generally with both learners and teachers. If the school community is able to observe the benefits of computer assisted learning, this could serve as a platform to expand to other grades or might result in higher rates of adoption if teachers at participating schools were trained to use the Khan Academy website.

### Lessons Learned

The Numeric programme takes place after school and is not linked to the South African school curriculum. This differentiates Numeric from most computer assisted learning interventions. Essentially, the programme offered mathematics classes as an extra-curricular activity as opposed to incorporating it into the learner's school day as a school subject.

Education interventions, particularly computer based interventions, by and large take the form of in-school programmes. These then act as a supplement for traditional teaching and would be held during school hours. On the other hand, an out-of-school or after-school programme, such as Numeric, would take place outside of school hours. An Indian study contrasted the results of an in-school and an out-of-school programme using CAL for learners in Grade 1 to Grade 3 in NGO-run schools in poor urban and rural areas. When an in-school programme was implemented, resulting mathematics scores were 0.57 standard deviations lower at treatment schools compared to the results of learners at schools where CAL was not utilised. The out-of-school programme complemented the existing arrangement and had the overall effect of an increase in 0.28 standard deviations in mathematics scores (Linden, 2008). The study fails to disentangle whether this increase is simply due to an increase in teaching time, or whether the use of technology is particularly effective as a remedial teaching tool when used in an after-school setting as a supplement to traditional learning and teaching. However, it clearly shows that there are certain settings in which an after-school programme using CAL may be an effective option. In South Africa the proportion of learners who progress to secondary school is reasonably high[64]. However, fundamental

64 The Department of Basic Education (2011b) in a Dropout and Learner Retention Strategy report states that 2.7 percent of students enrolled in Grade 7 in 2007 dropped out between 2007 and 2008. The same report claims that 5.1 percent of Grade 7 students were repeating Grade 7 in 2009. Noting that the original sample of Grade 7's would

skills and knowledge of basic numeracy are often lacking. It would be implausible to expect Grade 8 mathematics teachers to re-teach concepts covered in the Grade 3 or Grade 5 curriculum, although many of the learners in their classes may not have fully comprehended the concepts. Thus, at a Grade 8 level the introduction of an after-school programme solves this problem. Since it takes place outside of school, the learners and coaches are not constrained by the school curriculum. Simultaneously, learners are snot at risk of falling further behind in mathematics as they are still attending their regular school classes.

The fact that Numeric is an after-school programme also ensured that it was an optional activity. Thus the application process, particularly seeing which learners chose to opt into the programme, is very interesting. Certain non-cognitive skills such as motivation and capacity for learning would have played a part in a learner's decision to apply to the programme. These skills are developed throughout a child's life, and as such the home environment would have played an instrumental role in the development of these skills (Heckman, 2006). Some supplementary research is still being undertaken by UCT, in which parents are being interviewed telephonically. Here the chief purpose is to establish whether there is an observable difference in the parental productivity between applicants and non-applicants, looking specifically at parental involvement in the education of their child.

### Learner Application and Selection

There is, however, already a host of information available about both applicants and non-applicants. Section 5 provided an overview of some of the differences. Firstly, two trends were observed correlating likelihood of application with baseline mathematics results. The first noted that there was a negative correlation at the school level between the mean mathematics score at baseline and the proportion of learners that applied to the programme, whereas the second observation noted that within a school, higher performing learners were more likely to apply to the programme than low achieving learners.

It is possible, that the first trend is due to the parents of learners at low achieving schools realizing that their children are learning less than they have the potential to learn and taking corrective action in any way they can. However, as the correlation between income and average mathematics achievement is very high, it is also likely that at least some of this trend is explained by income (see Figure 5) and not just poor mathematics outcomes. Another likely hypothesis is that learners coming from higher income families are more likely to have safe and fun alternate activities available to them in the afternoons. These could take the form of extra-mural activities such as sports, religious or cultural activities, or could comprise of other forms of entertainment at home. But ultimately, this would have lowered the probability that such a learner would have voluntarily signed up for additional mathematics classes on their remaining free afternoons.

The second trend simply shows that within a school top learners are more likely to apply. Learners who achieved higher test results at baseline are also less likely to drop out if they are selected to participate on

have included both repeating and non-repeating students, this puts the lower bound of students who progress to high school at 94.6 percent (under the assumption that the only learners that drop out of Grade 7 are repeating learners). Realistically, this number is unlikely to be much higher than about 3 or 4 percent. Alternatively, between 81.6 and 90.9 percent of learners that enroll in Grade 1 complete Grade 7 and start Grade 8.

the programme. Regression results from Table 16 and Figure A3 show that learners who dropped out of the programme by the middle of the year did significantly worse than the average learner in the control group. Persistence on the programme could be viewed as measurement of the motivation and perseverance of each learner and thus learner that dropped out were initially less motivated than those who completed the programme. However, the learners who dropped out did no worse than the control group learners on Section A at endline. Section A contains material that was taught in primary school, whilst Section B tests new material that is introduced in Grade 7 and 8. It suggests that there are probably underlying factors that are affecting performance at school as well as increasing the likelihood of dropping out of the programme. Drop-out rates are negatively correlated with baseline mathematics ability. This suggests that initially worse performing students, and therefore also poorer students, are more likely to experience an event in their personal or family life[65], which results in them dropping out of the programme and also has a negative impact on their school work.

### Impacts of Gender and Language

Boys and girls appear to approach the programme differently, which is first evident in the probability of application. Namely, girls are 35 percent more likely to apply than boys. Applicant boys also performed 4.6 percentage points higher, on average, than non-applicant boys. This difference is significant at the 1 percent level, whereas the difference in average scores between applicant and non-applicant girls is a negligible 0.2 percentage points. Either this indicates that girls are more interested in applying to such a mathematics programme, or it is possible that the chosen method of recruitment appealed more to teenage girls than teenage boys, and as a result only boys who would already have been interested applied. It is also not the case that girls are just applying without any real intent of participating. Proportionately fewer girls than boys dropped out of the programme over the course of the year as can be seen in Table A4 in Appendix A. However, the girls appeared to benefit slightly less from the programme on average, although none of the differences are significant. The mean number of hours girls in treatment spent on Khan Academy was about the same as the time spent on the programme by boys. However the outcomes of girls were not as positive, particularly considering the lower drop-out rate. Girls mastered about 6 fewer skills than boys on average and although the difference is not significant, girls in treatment improved by about 3 percentage points less on average than boys between baseline and endline on Section A of the mathematics test. The differences in characteristics of girl and boy applicants suggest that it is self-selection into the programme that is responsible for at least a proportion of this difference in outcomes.

There was a disproportionately small number of Afrikaans speaking applicants who applied to the programme. It is possible that Afrikaans learners might have been dissuaded from applying because they perceived it as an exclusively English programme, as the intervention used English material and was officially offered only in English. Whilst any school with two major languages would have had a

[65] Example of reasons cited for learners dropping out that would have constituted such an event: learner is suddenly needed at home to do chores or look after siblings in the afternoon, learner got a boyfriend or girlfriend, transport issues such as a bus strike, where the parents were unable to afford to pay for alternate transport to school, learner started hanging out with the "wrong" crowd.

bilingual coach assigned to them, this is not something that the learners would have been aware of at the outset, and in future years it may be worthwhile to take the time to be clear during the application procedure that language does not necessarily need to be a barrier to the use of Khan Academy. This is supported by the finding that there is no differential treatment effect by home language and improvements on basic numeracy scores are not significantly different based on knowledge of English at baseline. However, for core Grade 8 curriculum material, initial English competency was shown to have a differential effect on treatment, significant at the 10 percent level. One possible explanation is that learners who struggle with English find it difficult to follow in school mathematics classes, so despite an improvement on basic numeracy they are unable to improve their school mathematics grades as quickly. This relationship between language and success in grade 8 outcomes should be investigated more closely in future studies[66].

### Limitations on the external validity of the results

There are limitations intrinsic to the design of the experiment and the nature of the programme. Two major concerns are the external validity of the results within South Africa and the potential change in the treatment effect as a result of the scaling up of the programme.

The external validity of the results is limited due to a number of factors. The Cape Metro Area is not representative of the country as a whole, and whilst it is well suited as a testing ground, it is not immediately obvious that similar effect sizes would be expected in other parts of the country, or even in more rural parts of the province. The schools selected for the pilot were wealthier than the average South African school, however, the results from this sample of 9 schools are encouraging. The difference between treatment and control at endline on Section A of both quintile 3 schools was 10 percentage points, which is 4 percentage points greater than the average treatment effect, whilst the difference in Section B scores between treatment and control were also higher than the average of 4.1 percentage points (see Table 17 and Table 30).

It is not clear whether treatment would be more or less effective in a rural as opposed to an urban setting. A number of factors may play a part in increasing or decreasing the effectiveness of the programme. Further research would need to be conducted in order to establish how effective the Numeric programme would be in a poor, rural setting. Additionally, increasing the sample size to test the intervention in multiple different locations would produce valuable information.

The pilot was only run in schools which already had functional, but underutilised, computer laboratories. The cost and time implications of setting up the appropriate infrastructure are huge, and it is likely that the setup cost could significantly decrease the cost effectiveness of the programme in schools or communities where the computer facilities and internet connectivity do not yet exist. However, the setup of such a facility would most likely provide to school with benefits besides improved numeracy. Future

[66] A reasonable share (about a quarter) of the Khan Academy videos have been translated into isiXhosa and as such there are a number of interesting questions around language and learning which could be examined in future.

studies would need to examine what the cost implications and feasibility of such an extension to the programme would be.

The Numeric programme in its current form is a small, privately run, non-profit organization. As a result of the relatively small size, the level of oversight, motivation and quality of staff, follow up with learners, teachers and parents and general management is of a high standard. Thus, there are some concerns with regards to scalability. An example of such a case is the Perry Preschool Project. This randomized trial was conducted in the 1960's and positive effects on learner achievement were still evident up until the age of 27 (Barnett, 2011). However, large scale public programmes utilising the same methods failed to achieve any gain in cognitive outcomes. One explanation is that the efficacy of the programme depends on the quality of implementation, which may be affected when scaling takes place, especially when the scaling up comes about as a result of a large-scale government intervention. The programme is currently in the form of an after-school programme that is in all probability complementing what teachers are teaching in the classroom. Additionally, learners needed to apply to participate as it was an optional after-school activity. There is no evidence to suggest similar improvements could be expected of an in-school programme targeting all learners which adopted a similar strategy to the Numeric after-school programme.

A last caveat in the interpretation of the main result is that the average impact on standardised mathematics scores was as a result of the Numeric programme as a whole. No attempt was made to determine which aspects of the programme were responsible for the improvement on mathematics scores. Further, evaluation would be required to determine which elements or combinations of features are most essential for improving the mathematics outcomes of learners.

### Ongoing research

There are a number of areas in which data collection is still ongoing. The telephonic parent interviews have already been mentioned. This thesis has examined only a single outcome measure in detail, namely, endline mathematics test scores. However, it is likely that the programme may have improved learners' outcomes in other areas. Notably, as this is a computer based intervention, computer fluency and computer literacy is likely to have increased. Similarly, spending time on a computer or solving Mathematics Olympiad problems may have increased the cognitive function of the learners. Some of these alternate outcomes can be measured using the Raven's, English Literacy and Computer Literacy tests that were conducted at baseline. These three tests were scheduled to be repeated at endline and at the time of writing they had not yet been concluded. Two schools still needed to be completed. Likewise, school marks and attendance data for 2013 is also still being collected and captured and will form the basis for future analysis.

## 11. Conclusion

There is an urgent need in South Africa to discover ways in which to improve the quality of education which is on offer for learners who are currently at school, as well as unemployed and unenrolled youths. The combination of an after-school programme using computer assisted learning proved a powerful way to significantly improve the mathematics outcomes of secondary school learners in their first year of secondary school. The after-school nature of the programme allowed the implementing organization to effectively utilise the infrastructure and resources of the schools, whilst simultaneously maintaining the freedom to operate outside of the nationally determined curriculum. The use of CAL, furthermore, allowed for individualized learning and focused remediation that significantly improved the basic numeracy of the learners on the programme, whilst simultaneously also improving their outcomes on core Grade 8 curriculum. Overall, the Numeric programme is a positive step towards addressing one of the problems facing learners in South African secondary schools, namely, a poor foundation in mathematics.

## 12.References


Angrist, J. & Lavy, V. 2002. New Evidence on Classroom Computers and Pupil Learning. *The Economic Journal*. 112: 735-765.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. 2007. Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*. 122(3): 1235-1264.

Banerjee, A. V. & Duflo, E. 2009. The Experimental Approach to Development Economics. *Annual Review of Economics*. 1:151–78

Barnett, W. S. 2011. Effectiveness of Early Educational Intervention. *Science*. 333: 975-978.

Barrara-Osorio, F. & Linden, L. L. 2009. The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia. World Bank Policy Research Working Paper 4836.

Barro, R. J. & Lee, J. J. 2001. International data on educational attainment: updates and implications. *Oxford Economic Papers*. 53(3): 541-563.

Barro, R. J. 2001. Human Capital and Growth. *The American Economic Review*. 91(2): 12-17.

Barrow, L., Markman, L. & Rouse, C. E. 2009. Technology's Edge: The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*. 1(1): 52-74.

Burde, D. & Linden, L. L. 2012. The Effect of Village Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan. National Bureau of Economic Research Working Paper 18039.

Cristia, J. P., Ibarraran, P., Cueto, S., Santiago, A. & Severin, E. 2012. Technology and Child Development: Evidence from the One Laptop per Child Programme. Inter-American Development Bank Working Paper No. IDB-WP-304.

Cross, M., Mungadi, R., & Rouhani, S. 2002. From policy to practice: Curriculum reform in South African education. *Comparative Education*. 38(2):171-187.

Department Of Basic Education. 2004. White Paper on e-Education. Governmente Gazette, No. 26762 , 2 September 2004. Available: http://www.education.gov.za/DocumentsLibrary/Legislation/WhitePapers/tabid/191/Default.aspx [2014, 21 January].

Department Of Basic Education. 2010. The status of the language of learning and teaching (LOLT) in South African public schools: A quantitative overview. Available: http://www.education.gov.za/LinkClick.aspx?fileticket=LlfRGMZxPRg%3D&tabid=422&mid=1261 [2014, 14 February].

Department Of Basic Education. 2011a. Curriculum News May 2011. Available: http://www.education.gov.za/LinkClick.aspx?fileticket=RlQ3WgihTOA%3D&tabid=348&mid=1018 [2014, 28 January].

Department Of Basic Education. 2011b. Report on Dropout and Learner Retention Strategy. Report to the Portfolio Committee on Education in June 2011, Available: http://www.education.gov.za/LinkClick.aspx?fileticket=jcSsY0rHcME%3D&tabid=358&mid=1261 [2014, 11 February].

Department Of Basic Education. 2013a. Annual Report 2012/2013. Pretoria: Department of Basic Education.

Department Of Basic Education. 2013b. Education Statistics in South Africa 2011. [Online data file]. Available: http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx [2013, 21 January].

Department Of Basic Education. 2013c. Masterlist data - March 2013 Ordinary Schools: Western Cape Province [Online data file]. Available: http://www.education.gov.za/EMIS/EMISDownloads/tabid/466/Default.aspx [2013, 23 October].

Fairlie, R. W. & Robinson, J. 2012. Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. Unpublished.

Finn, A., Leibbrandt, M., & Woolard, I. 2009. Income and expenditure inequality: Analysis of the NIDS Wave 1 dataset. National Income Dynamics Study. NiDS Discussion Paper 5.

Galston, W. A. 2001. Political knowledge, political engagement, and civic education. *Annual Review of Political Science.* 4(1): 217-234.

Glass, G. V. 1977. Integrating Findings: The Meta-Analysis of Research. *Review of Research in Education*. 5: 351-379.

Glewwe, P, Kremer, M & Moulin, S. 2002. Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya. Poverty Action Lab Paper No. 8.

Glewwe, P. 2002. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature*. 40(2): 436-482.

Glewwe, P. & Kremer, M. 2002. Schools, Teachers, and Education Outcomes in Developing Countries. *Handbook of the Economics of Education*. 2: 945-1017

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). Highlights from TIMSS 2007: Mathematics and Science Achievement of US Fourth-and Eighth-Grade Students in an International Context. (Research report; NCES 2009-001). Washington,DC, U.S: US National Center for Education Statistics.

Hembree, R. 1990. The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*. 21(1): 33-46.

John, V. 2014. Matric pass rate highest yet at 78.2%. Mail and Guardian Online. Available: http://mg.co.za/article/2014-01-06-matric-pass-rate-highest-yet-at-782 [Accessed on 21 January 2014].

Khan Academy. 2014. Khan Academy FAQ. Available: http://khanacademy.desk.com/ [2014, 21 January].

Kremer, M. & Holla, A. 2009. Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?. *Annual Review of Economics.* 1(1): 513-542.

Kremer, M., Brannen, C., & Glennerster, R. 2013. The challenge of education and learning in the developing world. *Science*, 340(6130): 297-300.

Linden, L. L. 2008. Complement or Substitute? The Effect of Technology on Student Achievement in India. Available: http://www.poverty-action.org/node/2241 [2013, 18 August].

Louw, J., Muller, J. & Tredoux, C. 2008. Time-on-task, technology and mathematics achievement. *Evaluation and Program Planning*. 31: 41-50.

Malamud, O. & Pop-Eleches, C. 2010. Home Computer Use and the Development of Human Capital, National Bureau of Economic Research Working Paper 15814.

Mankiw, N. G., Romer, D. & Weil, D. N. 1992. A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics.* 107(2): 407-437.

Mckenzie, D. 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economic*. 99(2): 210-221.

Miguel, E. & Kremer, M. 2004. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*. 72: 159-217.

Mybroadband (2014). Broadband price per Mbps shocker for South Africa (11 January 2014), Available at: http://mybroadband.co.za/news/broadband/93895-broadband-price-per-mbps-shocker-for-south-africa.html [Accessed 14 February 2014].

Ramatapana, K. & Makonye, J. P. 2012. From too much freedom to too much restriction: The case of teacher autonomy from National Curriculum Statement (NCS) to Curriculum and Assessment Statement (CAPS). *Africa Education Review*. 9(1): S7-S25

Raven, J., Raven, J. C. & Court, J. H. 2003. Raven Manual: Section 3. San Antonio: Pearson

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N. & Means, B. M. 2000. Changing How and What Children Learn in School with Computer-Based Technologies. The Future of Children, Children and Computer Technology. 10(2): 76-101.

Schonfield, J. W., Eurich-Fulcer, R. & Britt, C. L. 1994. Teachers, Computer Tutors, and Teaching: The Artificially Intelligent Tutor as an Agent for Classroom Change. *American Educational Research Journal*. 31(3): 579-607.

Spaull, N. 2011. Primary School Performance in Botswana, Mozambique, Namibia, and South Africa. SAQMEQ Working Paper 8, September 2011.

Spaull, N. 2013. South Africa's Education Crisis: The quality of education in South Africa 1994-2011. (CDE report October 2013). Johannesburg, South Africa: Center for Development and Enterprise

Speakman, R. & Welch, F. 2006. Chapter 13: Using Wages to Infer School Quality. In Kenneth J. Arrow and Michael D. Intriligator (eds). Handbook of the Economics of Education. 2: 818-864.

Statistics South Africa. 2010a. Census At School Results 2009, Available http://www.statssa.gov.za/censusatschool/docs/Census_At_School_2009_Report.pdf [2014, 21 January].

Statistics South Africa. 2010b. Republic of South Africa Millenium Development Goals Country Report 2010. Available: http://www.info.gov.za/view/DownloadFileAction?id=132022 [2013, 17 August].

Taylor, S. 2011. Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study. Ph.D. Thesis. University of Stellenbosch.

Van der Berg, S. 2006. The targeting of public spending on school education, 1995 and 2000. *Perspectives in Education: Education Finance*. 24(2): 49.

Van der Berg, S. 2007. Apartheid's enduring legacy: Inequalities in education. *Journal of African Economies*. 16(5): 849-880.

Western Cape Education Department. 2009. WCED celebrates 1000 Khanya schools, Media Release. Available: http://wced.pgwc.gov.za/comms/press/2009/41_khanya.html [2014, 21 January].

Western Cape Education Department. 2013a. Background to the national quintile system. 14 October 2013. Available: http://wced.pgwc.gov.za/comms/press/2013/74_14oct.html [2014, 11 February].

Western Cape Education Department. 2013b. Education management information systems: Find-a-school. Available: http://wcedemis.pgwc.gov.za/wced/findaschool.html [2013, 23 October].

White, I. R. & Thompson, S. G. 2005. Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine*. 24(7): 993-1007.

World Bank. 2014. World Development Indicators [Online data file]. Available: http://data.worldbank.org/country/south-africa [2014, 21 January].

# 13. Appendix A

Table A1: Logit estimates showing the probability of applying to the Numeric programme

| | Applied | | |
|---|---|---|---|
| **VARIABLES** | **(1)** | **(2)** | **(3)** |
| Within School – Mathematics Tercile 2 | 0.416*** | 0.416*** | 0.338** |
| | (0.127) | (0.127) | (0.132) |
| Within School – Mathematics Tercile 3 | 0.564*** | 0.564*** | 0.440*** |
| | (0.128) | (0.128) | (0.135) |
| School 102 | 0.368 | | |
| | (0.257) | | |
| School 104 | 0.344 | 0.122 | 0.137 |
| | (0.251) | (0.219) | (0.225) |
| School 105 | 0.771*** | 0.362 | 0.622** |
| | (0.220) | (0.241) | (0.257) |
| School 106 | 0.438* | -0.146 | -0.117 |
| | (0.224) | (0.337) | (0.350) |
| School 107 | 0.133 | -0.451 | -0.449 |
| | (0.230) | (0.341) | (0.353) |
| School 109 | -0.519** | -0.215 | -0.279 |
| | (0.263) | (0.409) | (0.427) |
| School 110 | 0.479* | 0.149 | 0.310 |
| | (0.247) | (0.237) | (0.248) |
| School 111 | 0.853*** | 0.650*** | 0.599*** |
| | (0.225) | (0.188) | (0.198) |
| Mean School Fees (in thousands) | | -0.254 | -0.248 |
| | | (0.177) | (0.192) |
| Female | | | 0.409*** |
| | | | (0.109) |
| Age | | | -0.147** |
| | | | (0.066) |
| Coloured | | | 0.543* |
| | | | (0.318) |
| Asian, Indian or White | | | 0.735 |
| | | | (0.469) |
| Home Language: English | | | 0.580*** |
| | | | (0.197) |
| Home Language: IsiXhosa | | | 1.133*** |
| | | | (0.365) |
| Home Language: Other Language | | | 1.667*** |
| | | | (0.637) |
| Constant | -1.608*** | -1.024*** | -0.152 |
| | (0.196) | (0.317) | (1.065) |
| **Number of Observations** | **1,922** | **1,922** | **1,809** |

Notes: Logit regression. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: OLS estimates of treatment on mathematics Scores on Section A and Section B at endline

| | Endline Mathematics Test Scores | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Basic Numeracy - Section A | | | | Grade 8 - Section B | | | |
| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treatment | 0.060*** | 0.057*** | 0.058*** | 0.058*** | 0.041** | 0.037*** | 0.039*** | 0.037*** |
| | (0.018) | (0.010) | (0.010) | (0.010) | (0.017) | (0.011) | (0.010) | (0.011) |
| Baseline-Section A | | 0.845*** | 0.716*** | 0.725*** | | 0.175 | 0.025 | 0.061 |
| | | (0.133) | (0.140) | (0.151) | | (0.132) | (0.130) | (0.139) |
| (Baseline-Section A)$^2$ | | -0.000 | 0.020 | -0.012 | | 0.619*** | 0.641*** | 0.591*** |
| | | (0.126) | (0.129) | (0.138) | | (0.131) | (0.124) | (0.133) |
| School 102 | | 0.002 | 0.027 | 0.014 | | 0.036 | 0.064* | 0.051 |
| | | (0.022) | (0.031) | (0.031) | | (0.025) | (0.034) | (0.033) |
| School 104 | | 0.076*** | 0.064** | 0.060** | | 0.076*** | 0.066*** | 0.062** |
| | | (0.024) | (0.026) | (0.028) | | (0.023) | (0.023) | (0.024) |
| School 105 | | 0.013 | -0.004 | -0.010 | | 0.077*** | 0.064*** | 0.059*** |
| | | (0.019) | (0.019) | (0.021) | | (0.020) | (0.020) | (0.021) |
| School 106 | | -0.007 | 0.006 | -0.003 | | 0.050** | 0.070*** | 0.068** |
| | | (0.020) | (0.026) | (0.027) | | (0.020) | (0.026) | (0.027) |
| School 107 | | 0.075*** | 0.088*** | 0.076** | | 0.137*** | 0.148*** | 0.141*** |
| | | (0.029) | (0.032) | (0.033) | | (0.030) | (0.034) | (0.035) |
| School 109 | | 0.038* | 0.040 | 0.043* | | 0.052** | 0.047** | 0.044* |
| | | (0.023) | (0.024) | (0.026) | | (0.022) | (0.022) | (0.024) |
| School 110 | | -0.018 | -0.022 | -0.032 | | 0.063*** | 0.071*** | 0.058** |
| | | (0.021) | (0.023) | (0.024) | | (0.023) | (0.025) | (0.025) |
| School 111 | | 0.102*** | 0.080*** | 0.072*** | | 0.131*** | 0.112*** | 0.106*** |
| | | (0.017) | (0.017) | (0.018) | | (0.021) | (0.021) | (0.021) |
| Female | | | -0.022** | -0.008 | | | -0.003 | 0.005 |
| | | | (0.010) | (0.012) | | | (0.012) | (0.013) |
| Age | | | 0.334*** | 0.291** | | | 0.306*** | 0.377*** |
| | | | (0.105) | (0.118) | | | (0.104) | (0.116) |
| Coloured | | | -0.108** | -0.100* | | | -0.113* | -0.120* |
| | | | (0.046) | (0.052) | | | (0.062) | (0.065) |
| Asian, Indian and White | | | -0.126** | -0.122** | | | -0.120* | -0.140** |
| | | | (0.050) | (0.056) | | | (0.068) | (0.070) |
| Home Language: English | | | -0.017 | -0.022 | | | 0.004 | -0.002 |
| | | | (0.016) | (0.017) | | | (0.018) | (0.021) |
| Home Language: IsiXhosa | | | -0.119** | -0.124** | | | -0.105 | -0.123* |
| | | | (0.050) | (0.056) | | | (0.065) | (0.070) |
| Home Language: Other_language | | | -0.053 | -0.047 | | | -0.073 | -0.094 |
| | | | (0.075) | (0.084) | | | (0.073) | (0.068) |
| Baseline – Raven's Score | | | 0.107*** | 0.093** | | | 0.155*** | 0.131*** |
| | | | (0.035) | (0.039) | | | (0.035) | (0.037) |
| Baseline -English Literacy Score | | | 0.076** | 0.095** | | | 0.112*** | 0.129*** |
| | | | (0.038) | (0.039) | | | (0.037) | (0.039) |
| Baseline -Typing (Keys | | | 0.029 | 0.026 | | | 0.020 | 0.026 |

| | | | | | | |
|---|---|---|---|---|---|---|
| per minute/100) | | (0.021) | (0.021) | | (0.023) | (0.023) |
| Learner's Mother is Alive | | | 0.033 | | | 0.003 |
| | | | (0.028) | | | (0.027) |
| Learner's Mother has completed Matric | | | -0.014 | | | 0.000 |
| | | | (0.013) | | | (0.013) |
| Learner's Father is Alive | | | -0.006 | | | -0.007 |
| | | | (0.017) | | | (0.017) |
| Learner's Father has completed Matric | | | 0.004 | | | 0.019 |
| | | | (0.013) | | | (0.013) |
| At least one computer at Home | | | -0.011 | | | -0.006 |
| | | | (0.015) | | | (0.014) |
| Internet on a Computer at Home | | | 0.008 | | | -0.007 |
| | | | (0.013) | | | (0.014) |
| Number of Household Members | | | 0.002 | | | 0.005** |
| | | | (0.002) | | | (0.002) |
| Travel Time to School | | | 0.001 | | | -0.000 |
| | | | (0.028) | | | (0.033) |
| Learner Likes Mathematics a Lot | | | 0.024** | | | 0.020 |
| | | | (0.011) | | | (0.012) |
| Learner has repeated at least one Grade | | | -0.003 | | | 0.031** |
| | | | (0.015) | | | (0.014) |
| Is a member of a sports team | | | 0.020 | | | 0.010 |
| | | | (0.013) | | | (0.014) |
| In a member of a Cultural Group | | | -0.011 | | | -0.014 |
| | | | (0.014) | | | (0.014) |
| Dummy: Baseline Mathematics Score Missing | 0.296*** | 0.261*** | 0.285*** | 0.145** | 0.102 | 0.123* |
| | (0.054) | (0.057) | (0.063) | (0.061) | (0.062) | (0.063) |
| Dummy: Age Missing | | -0.355*** | -0.312** | | -0.320*** | -0.396*** |
| | | (0.111) | (0.124) | | (0.110) | (0.122) |
| Dummy: Race Missing | | -0.120** | -0.126** | | -0.083 | -0.103 |
| | | (0.054) | (0.060) | | (0.079) | (0.083) |
| Dummy: Home Language Missing | | 0.002 | -0.008 | | 0.033 | -0.010 |
| | | (0.055) | (0.053) | | (0.083) | (0.084) |
| Dummy: Mother Alive Missing Ite | | | 0.120 | | | -0.067 |
| | | | (0.153) | | | (0.129) |
| Dummy: Mother Matric Missing Item | | | -0.026 | | | 0.002 |
| | | | (0.017) | | | (0.020) |
| Dummy: Father Alive Missing Item | | | -0.086 | | | -0.059 |
| | | | (0.072) | | | (0.036) |
| Dummy: Father Matric Missing Item | | | 0.023 | | | 0.016 |
| | | | (0.017) | | | (0.018) |
| Dummy: Household Computer Missing Item | | | -0.010 | | | -0.019 |
| | | | (0.019) | | | (0.021) |
| Dummy: Household Internet Missing Item | | | -0.003 | | | -0.012 |
| | | | (0.015) | | | (0.015) |
| Dummy: Household Members Missing Item | | | -0.044 | | | -0.018 |
| | | | (0.033) | | | (0.034) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dummy: travel Time to School Missing Item | | | | -0.001 | | | | 0.000 |
| | | | | (0.029) | | | | (0.033) |
| Dummy: Like Mathematics A Lot Missing | | | | -0.010 | | | | 0.229* |
| | | | | (0.141) | | | | (0.132) |
| Dummy: Grade Repeat Missing | | | | 0.003 | | | | -0.045* |
| | | | | (0.024) | | | | (0.024) |
| Dummy: Baseline Sports Missing | | | | -0.011 | | | | 0.055 |
| | | | | (0.048) | | | | (0.053) |
| Dummy: Baseline Cultural Group Missing | | | | 0.032 | | | | -0.048 |
| | | | | (0.039) | | | | (0.038) |
| Dummy: Raven's Missing | | | 0.062* | 0.058 | | | 0.116*** | 0.096** |
| | | | (0.037) | (0.037) | | | (0.042) | (0.043) |
| Dummy: English Missing | | | 0.078* | 0.086 | | | 0.017 | 0.063 |
| | | | (0.045) | (0.057) | | | (0.050) | (0.057) |
| Dummy: Typing Missing | | | -0.058*** | -0.060* | | | 0.008 | -0.022 |
| | | | (0.020) | (0.036) | | | (0.019) | (0.033) |
| Constant | 0.488*** | 0.063* | 0.438*** | 0.384*** | 0.264*** | -0.041 | 0.163 | 0.209 |
| | (0.012) | (0.037) | (0.129) | (0.137) | (0.011) | (0.037) | (0.126) | (0.140) |
| Observations | 453 | 453 | 453 | 453 | 453 | 453 | 453 | 453 |
| R-squared | 0.024 | 0.707 | 0.741 | 0.754 | 0.013 | 0.627 | 0.668 | 0.687 |

Notes: Robust standard errors are in parentheses. Statistical significance of the coefficient is given by *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Distribution of the difference in basic numeracy (Section A) mathematics scores between baseline and endline

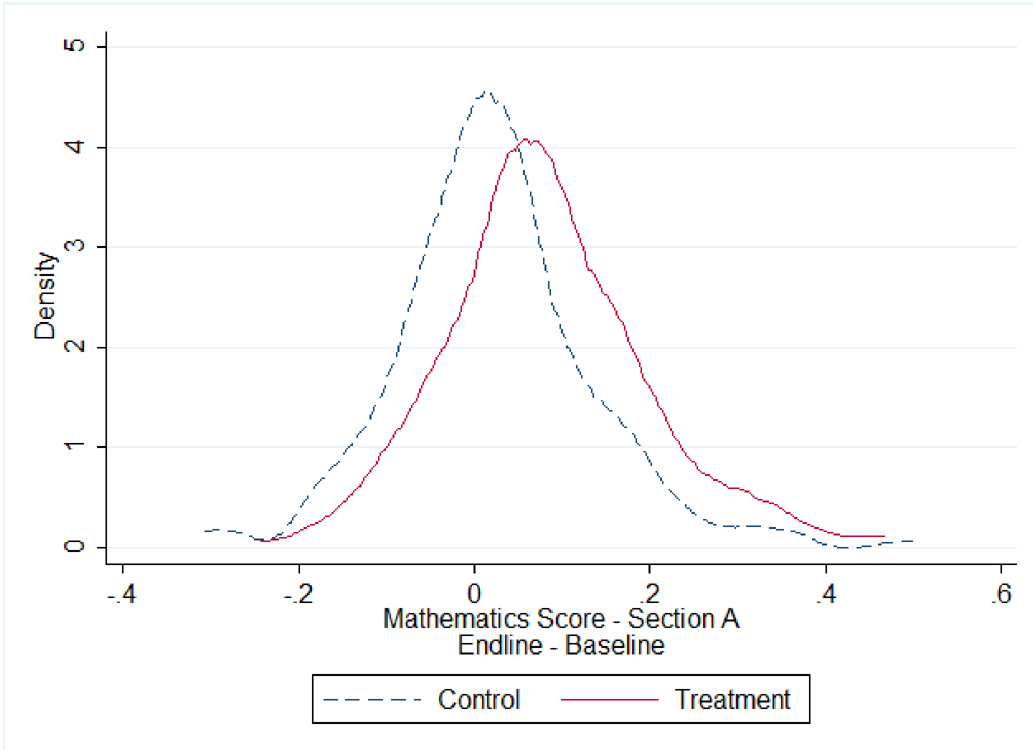



Figure A2: Bar graph of the mean score achieved on Section A of the endline Mathematics test by quintile and drop-out status in July

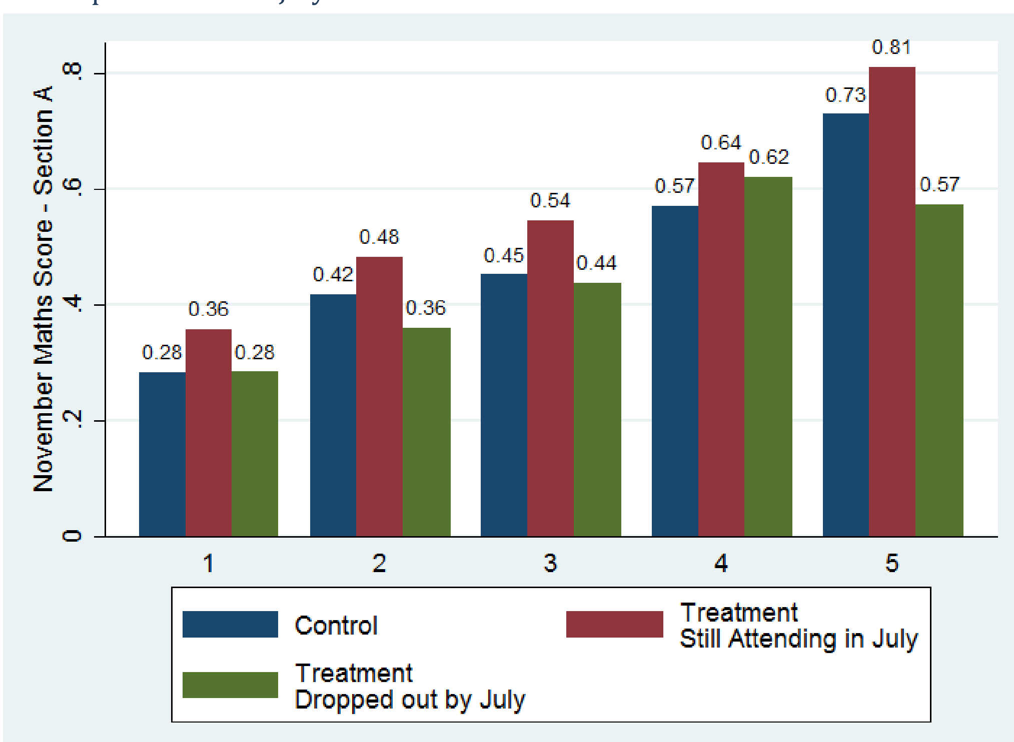


Figure A3: Bar graph of the mean score achieved on Section B of the endline Mathematics test by baseline Mathematics quintile and drop-out status in July

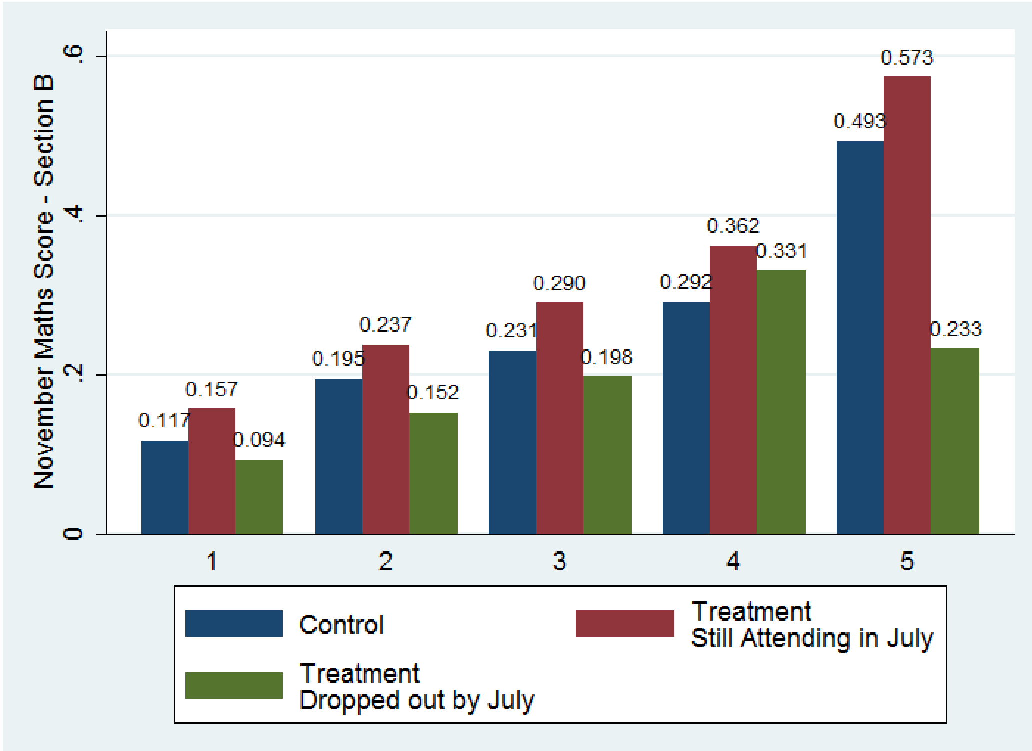



Figure A4: Histogram showing the Number of Skill Mastered by Learners on Khan Academy

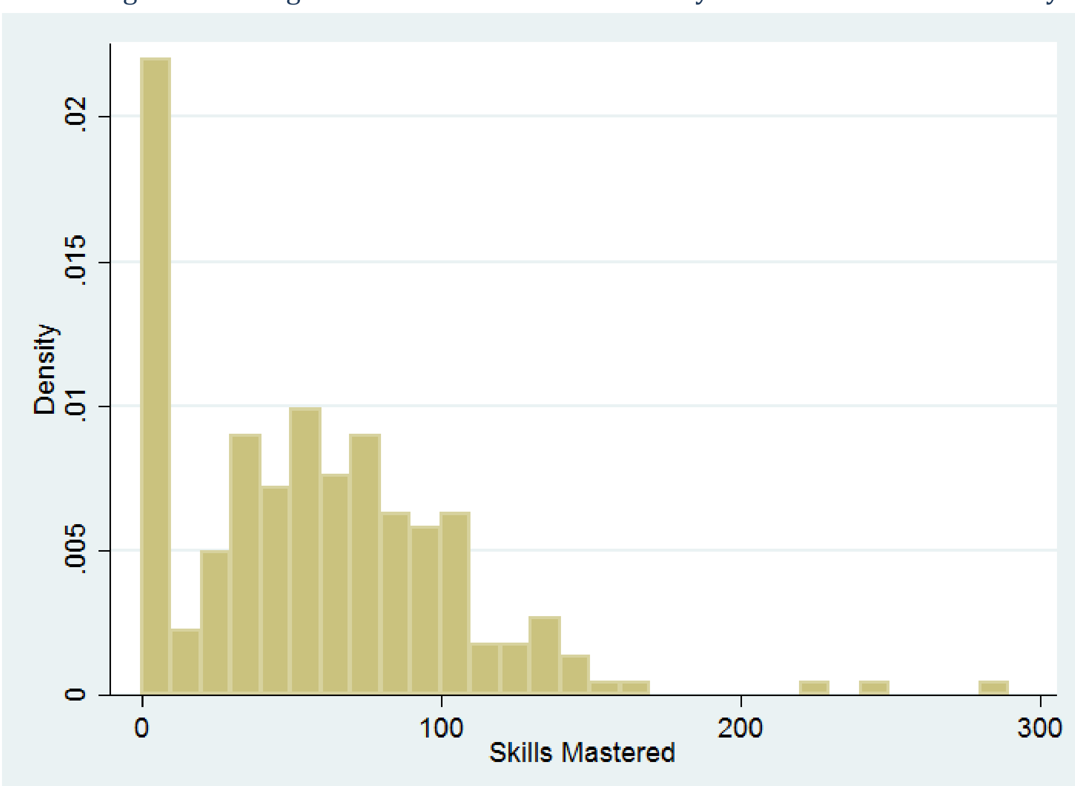


Figure A5: Histogram showing the distribution of the total time Learners spent doing exercises on the Khan Academy website

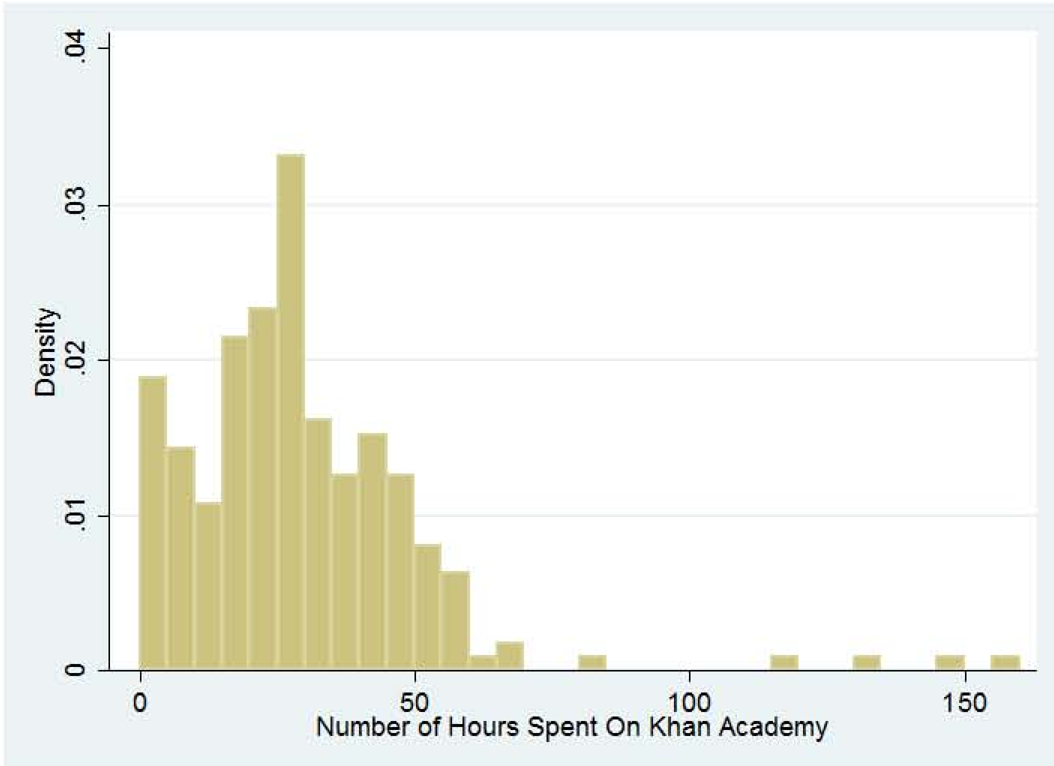


Table A3: Table to show the basic numeracy scores at baseline and endline Mathematics outcomes by gender

| | All | | | Applied | | | Control | | | Treatment | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Male** | **Female** | **Difference** *(Male-Female)* | **Male** | **Female** | **Difference** *(Male-Female)* | **Male** | **Female** | **Difference** *(Male-Female)* | **Male** | **Female** | **Difference** *(Male-Female)* |
| Baseline Mathematics - Section A | 0.439 | 0.444 | -0.005 | 0.475 | 0.446 | 0.029** | 0.478 | 0.453 | 0.025 | 0.479 | 0.454 | 0.025 |
| | (0.179) | (0.172) | (0.008) | (0.167) | (0.163) | (0.014) | (0.162) | (0.168) | (0.022) | (0.178) | (0.166) | (0.023) |
| | 907 | 991 | 1898 | 217 | 330 | 547 | 104 | 127 | 231 | 88 | 143 | 231 |
| Endline Mathematics - Section A | 0.493 | 0.483 | 0.011 | 0.536 | 0.501 | 0.034** | 0.496 | 0.481 | 0.015 | 0.582 | 0.53 | 0.052* |
| | (0.197) | (0.181) | (0.009) | (0.206) | (0.178) | (0.017) | (0.203) | (0.174) | (0.025) | (0.207) | (0.186) | (0.027) |
| | 824 | 941 | 1765 | 212 | 328 | 540 | 101 | 126 | 227 | 85 | 141 | 226 |
| Endline Mathematics - Section B | 0.241 | 0.243 | -0.003 | 0.29 | 0.273 | 0.017 | 0.263 | 0.265 | -0.002 | 0.326 | 0.293 | 0.033 |
| | (0.165) | (0.16) | (0.008) | (0.186) | (0.172) | (0.016) | (0.167) | (0.167) | (0.022) | (0.209) | (0.183) | (0.026) |
| | 824 | 941 | 1765 | 212 | 328 | 540 | 101 | 126 | 227 | 85 | 141 | 226 |
| Difference Section A (Endline - Baseline) | 0.044 | 0.035 | 0.009* | 0.051 | 0.053 | -0.002 | 0.017 | 0.028 | -0.011 | 0.094 | 0.072 | 0.021 |
| | (0.108) | (0.106) | (0.005) | (0.123) | (0.11) | (0.01) | (0.117) | (0.106) | (0.015) | (0.122) | (0.111) | (0.016) |
| | 791 | 916 | 1707 | 207 | 317 | 524 | 100 | 123 | 223 | 84 | 138 | 222 |

Notes: Columns with the male and female means have standard deviation in parentheses. Standard errors are in parentheses in columns with the Difference between male and females. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Table to show the number of Skills mastered, time spent on Khan Academy and drop-out rates of Girls and Boys

| | Male | Female | Difference *(Male-Female)* |
|---|---|---|---|
| Skills Mastered | 60.780 | 54.936 | 5.844 |
| | (54.793) | (41.703) | (6.517) |
| Hours On Khan Academy | 29.463 | 29.271 | 0.192 |
| | (26.449) | (18.761) | (3.041) |
| Dropped out by end of July | 0.247 | 0.163 | 0.084 |
| | (0.434) | (0.371) | (0.053) |
| Dropped out by end of November | 0.360 | 0.238 | 0.121* |
| | (0.483) | (0.427) | (0.060) |
| **Number of Observations** | **89** | **147** | **236** |

Notes: Column 1 and column 2 have standard deviation in parentheses. Standard errors are in parentheses in last column. *** p<0.01, ** p<0.05, * p<0.1

# 14. Appendix B

## Missing Outcome Variables

Missing observations, particularly, missing outcome variables are a potential source of bias if there are underlying factors which affect attendance at tests and also the effectiveness of treatment. There were 472 in the evaluated sample. At baseline 461 (97.7 percent) of these learners in the evaluated sample wrote the mathematics tests, whilst only 453 (96.0 percent) learners from the treatment and control group wrote the endline Mathematics test. There are two learners in the evaluated sample for whom there is neither baseline nor endline Mathematics test data. Table B1 shows the mean baseline Mathematics score for learners who have missing and non-missing endline mathematics test scores divided into treatment and control groups. It is evident that the treatment learners with missing outcome variables scored significantly (p = 0.034) lower than learners in control who had missing observations. If as a result of this

Table B1: Table to show the mean baseline Mathematics scores of learners with missing endline mathematics data in control and treatment

| | Missing Endline Mathematics Data | Non-Missing Endline Mathematics Data | Total |
|---|---|---|---|
| **Control** | 0.38 (0.11) 8 | 0.47 (0.17) 223 | 0.46 (0.17) 231 |
| **Treatment** | 0.26 (0.11) 9 | 0.47 (0.17) 221 | 0.46 (0.17) 230 |
| **Total** | 0.32 (0.12) 17 | 0.47 (0.17) 444 | 0.46 (0.17) 461 |

Notes: The first number in each cell is the mean baseline Section A mathematics scores, in parentheses are the standard deviation, whilst the last number specifies the number of observations.

Out of the 17 (3.6 percent of sample) learners for whom there is baseline, but no endline data available, there is midline mathematics test data available for 11. Of the remaining 6, 5 had moved to a different school and one was missing. Table B2 presents the average improvement on Section A between baseline and midline for learners with and without endline mathematics observations in treatment and control.

Table B2: Table to show the mean difference in Section A test scores between baseline and midline of learners with missing endline mathematics data in control and treatment

| | Missing Endline Mathematics Data | Non-Missing Endline Mathematics Data | Total |
|---|---|---|---|
| **Control** | 0.087 (0.112) 5 | 0.043 (0.095) 209 | 0.044 (0.096) 214 |
| **Treatment** | 0.073 (0.105) 6 | 0.084 (0.106) 211 | 0.083 (0.105) 217 |
| **Total** | 0.079 (0.103) 11 | 0.064 (0.102) 420 | 0.064 (0.102) 431 |

Notes: The first number in each cell is the mean baseline Section A mathematics scores, in parentheses are the standard deviation, whilst the last number specifies the number of observations.

It is evident that there is no significant difference in improvement from baseline to midline between control and treatment learners with missing endline data. This suggests that any possible bias would be small, due to the small number of missing observations, and would come from an underestimation for the improvement of the average control group scores over time as opposed to an overestimation of the increase in average treatment scores over the course of the year.

## Missing Indicator Method

There are two main reasons for the inclusion of baseline characteristics as covariates in a regression to estimate a treatment effect. Baseline characteristics are used as controls in calculating the treatment effect as this increases the power, particularly in the case of a small sample (White and Thompson, 2004). In this thesis baseline characteristics are also used to calculate differential treatment effects across various subgroups.

If baseline characteristics are included as controls, all observation that are missing information for one or more covariates would be excluded. This has two implications. Firstly, the decrease in sample size results in a loss of statistical power. Additionally, if observations are excluded non-randomly, and the underlying factors are also correlated with the outcome variable or variable of interest, then this is also a possible source of bias.

Table B3: Table to show the number of missing observations on the baseline and endline Mathematics tests as well as missing item numbers for various groups of covariates.

| | Number of Non-Missing Observations | Cumulative Non-Missing Observations |
|---|---|---|
| Treatment and Schools | 472 | 472 |
| Endline Mathematics | 453 | 453 |
| Baseline Mathematics | 461 | 444 |
| Individual Characteristics (Age, Gender, Race and Home Language) | 463 | 440 |
| Raven's Test Results | 387 | 366 |
| English and Typing Test Results | 366 | 348 |
| Home and Parent Characteristics | 247 | 183 |
| School Related Characteristics | 442 | 180 |

### Missing Baseline Mathematics Test Scores

Table B4: Table to show the average results at midline and endline of learners that has missing baseline mathematics data

| | | Missing Baseline Mathematics Data | | Non-Missing Baseline Mathematics Data | |
|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment |
| Midline - Section A | Mean | 0.52 | 0.51 | 0.51 | 0.55 |
| | Standard Deviation | (0.09) | (0.17) | (0.17) | (0.19) |
| | N | 3 | 214 | 214 | 217 |
| Midline - Section B | Mean | 0.17 | 0.24 | 0.24 | 0.26 |
| | Standard Deviation | (0.1) | (0.14) | (0.14) | (0.16) |
| | N | 3 | 214 | 214 | 217 |
| Endline - Section A | Mean | 0.40 | 0.49 | 0.49 | 0.55 |
| | Standard Deviation | (0.12) | (0.19) | (0.19) | (0.2) |
| | N | 4 | 223 | 223 | 221 |
| Endline - Section B | Mean | 0.18 | 0.27 | 0.27 | 0.31 |
| | Standard Deviation | (0.16) | (0.17) | (0.17) | (0.19) |
| | N | 4 | 223 | 223 | 221 |

### Missing Baseline Raven's SPM, English and typing test results

There were a large number of missing observations on the Raven's, English Literacy and typing test baseline results. There were 85 (18 percent) learners for whom there is no baseline Raven's score, whilst data is missing for 106 (22 percent) learners for the English and typing scores.

Table B5: Table to show the mean baseline Mathematics scores of learners with missing baseline Raven's data in control and treatment

| | Missing Baseline Raven's Test Data | Non- Missing Baseline Raven's Test Data | Total |
|---|---|---|---|
| **Control** | 0.42 (0.14) 47 | 0.47 (0.17) 184 | 0.46 (0.17) 231 |
| **Treatment** | 0.43 (0.16) 33 | 0.47 (0.17) 197 | 0.46 (0.17) 230 |
| **Total** | **0.43 (0.15) 80** | **0.47 (0.17) 381** | **0.46 (0.17) 461** |

Notes: The first number in each cell is the mean baseline Section A mathematics scores, in parentheses are the standard deviation, whilst the last number specifies the number of observations.

Learners who wrote the baseline Raven's test score 4.4 percent higher on the baseline mathematics test, significant at the 5 percent level. This suggests that weaker learners were more likely to have been absent for the after-school tests.

The following tables show the proportion of learners that wrote the Raven's SPM test within each gender, race, home language and age category.

Tables B6: Tables to show the proportion of baseline Raven's tests written by gender, race, home language and age

| Gender | |
|---|---|
| Male | 0.81 (0.39) 194 |
| Female | 0.83 (0.38) 278 |
| Total | 0.82 (0.38) 472 |

| Race | |
|---|---|
| African | 0.86 (0.35) 193 |
| Coloured | 0.79 (0.41) 263 |
| Total | 0.82 (0.38) 456 |

| Home Language | |
|---|---|
| Afrikaans | 0.82 (0.39) 44 |
| English | 0.80 (0.40) 236 |
| isiXhosa | 0.85 (0.36) 181 |
| Total | 0.82 (0.38) 461 |

| Age | |
|---|---|
| 13 | 0.84 (0.37) 232 |
| 14 | 0.83 (0.38) 176 |
| 15 | 0.75 (0.44) 40 |
| 16 | 0.81 (0.40) 16 |
| Total | 0.83 (0.38) 464 |

# 15.Appendix C

## Learner - Midyear Programme Feedback

During the first two weeks of September, all learners who attended their class on the day of the survey were asked to fill in an anonymous online feedback form. A total of 194 responses were received. As this survey was conducted in September, any learner who had dropped out during the first two terms would not be represented. Additionally, as learners who dropped out were replaced by other on the waiting list, this number would also include some learners who applied, but are not in the control or treatment groups. Selected questions and their responses are reported here.

Figure C1: Which session do you attend?

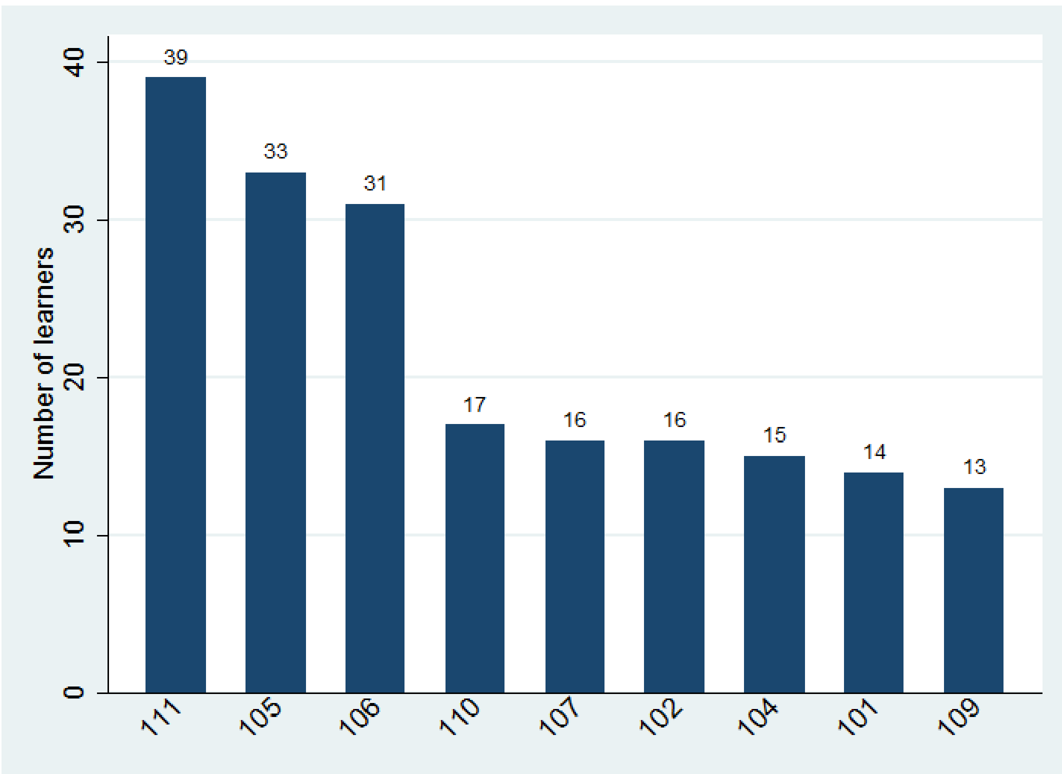


Figure C2: What do you enjoy most at Khan Academy classes?

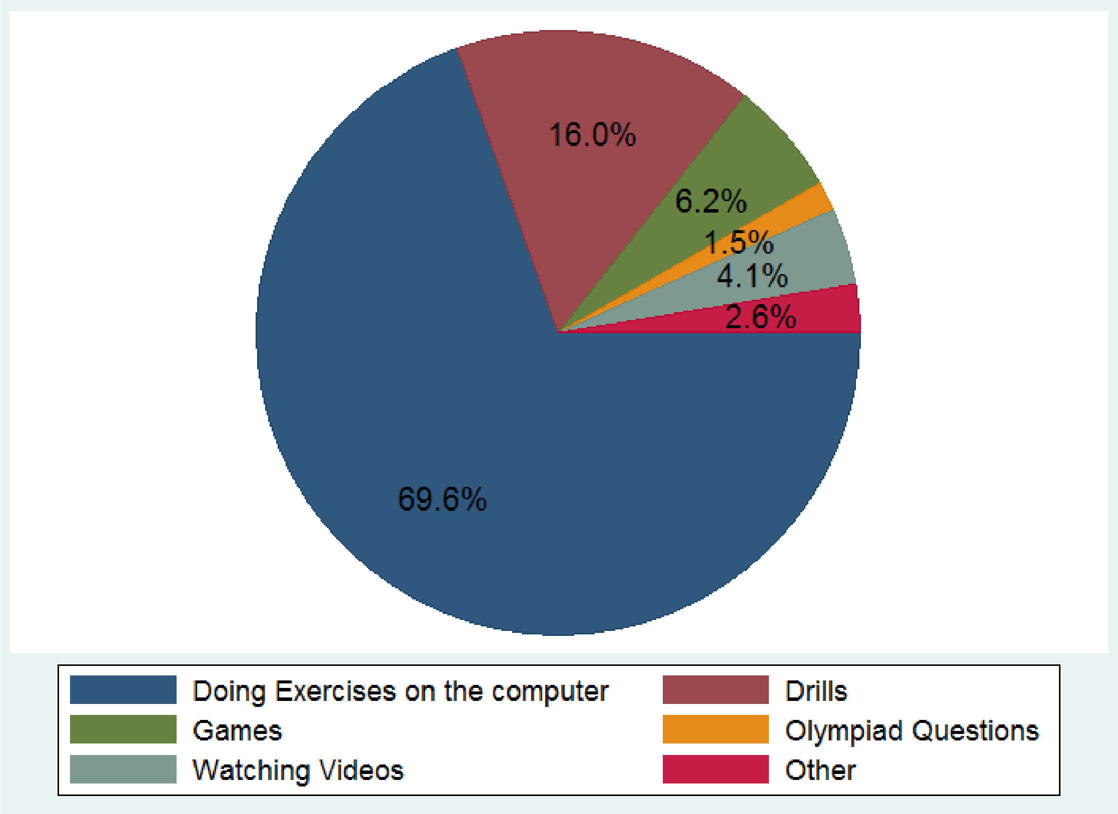



Figure C3: Would you prefer to watch a video or have your coach teach a lesson to the class?

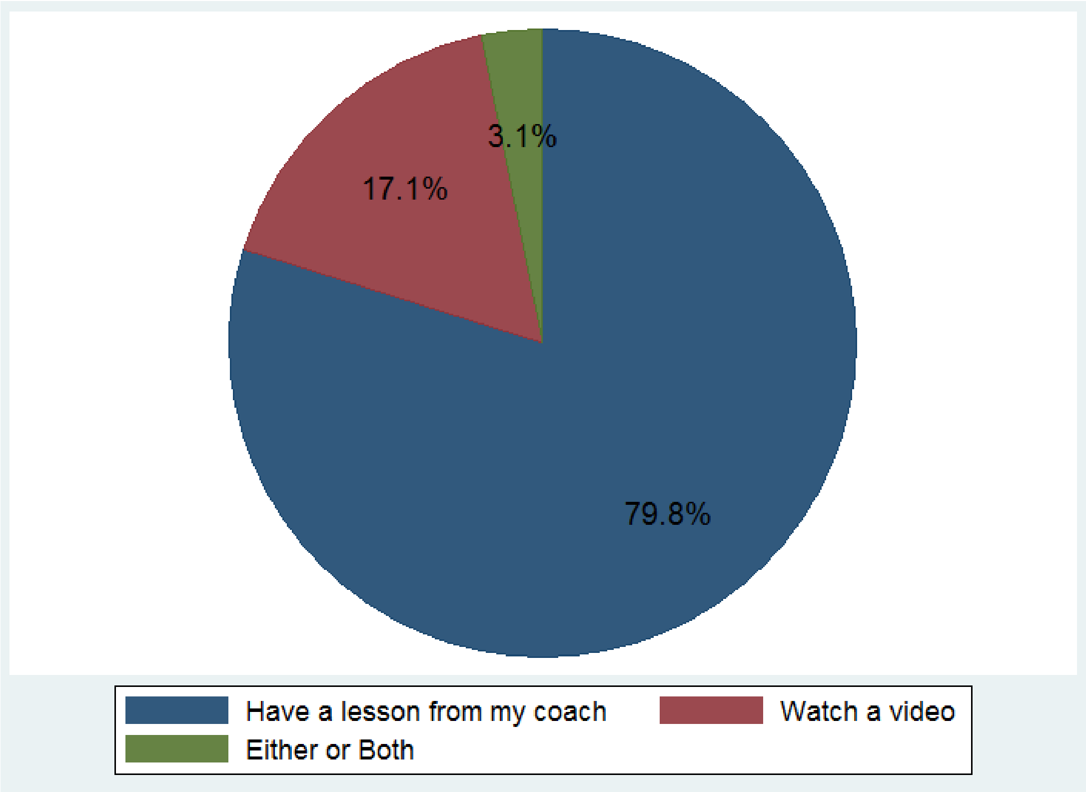


Figure C4: Do you like Maths?

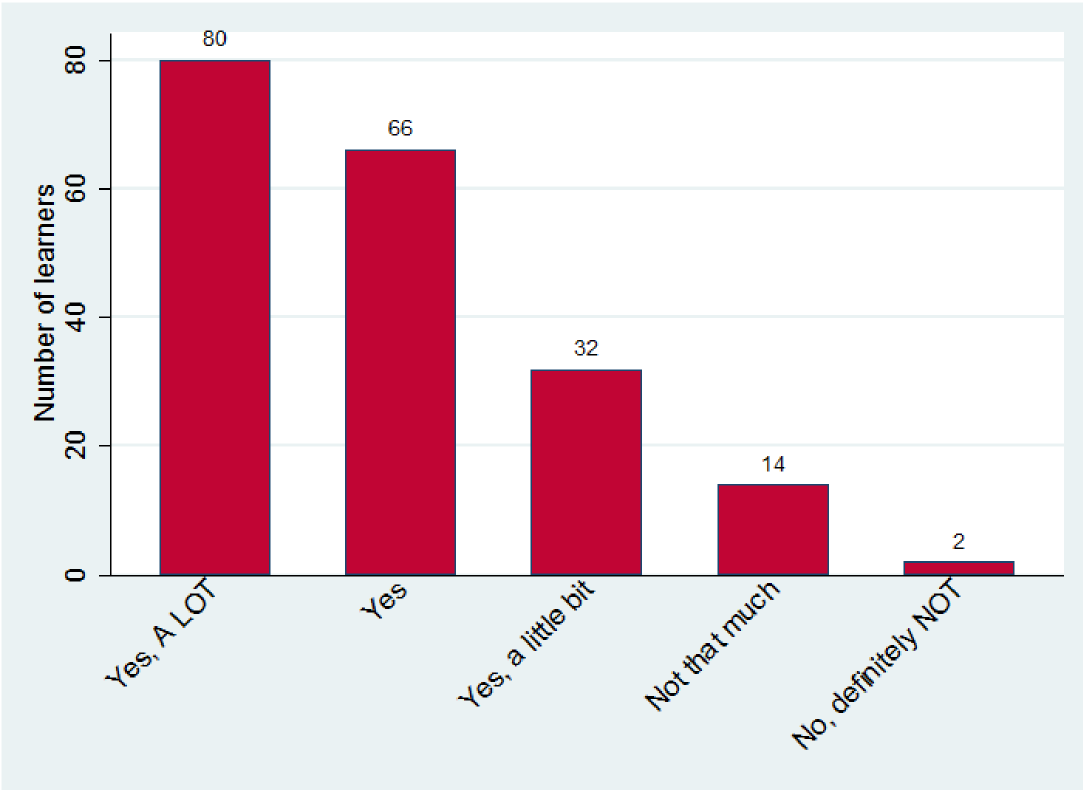



Figure C5: Do you enjoy Maths more now than you did at the beginning of the year?

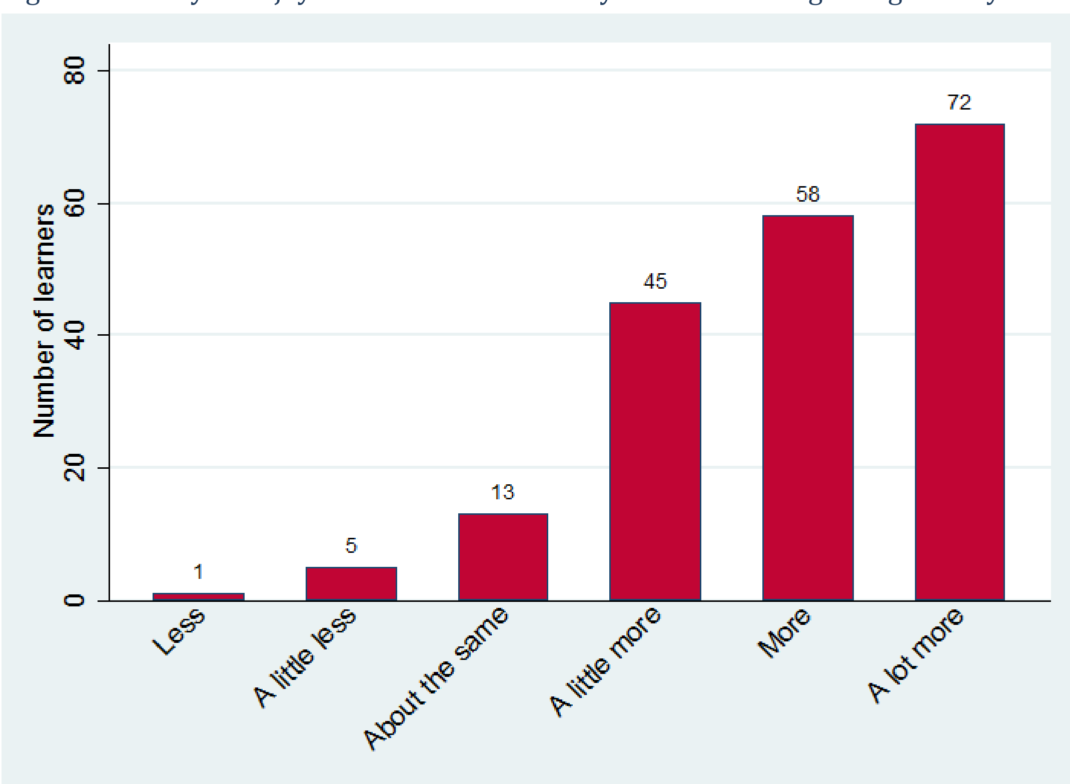


Figure C6: Have the things you learnt at Khan Academy helped you in Maths at school?

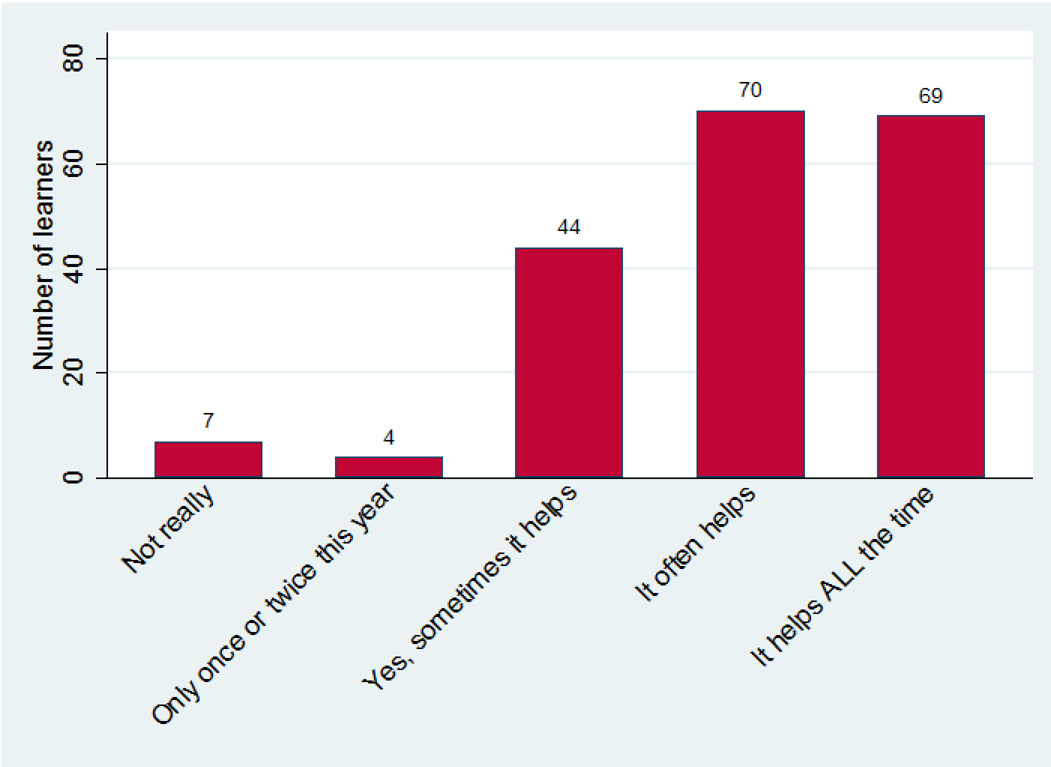



Figure C7: Have your computer skills improved this year because of Khan Academy?

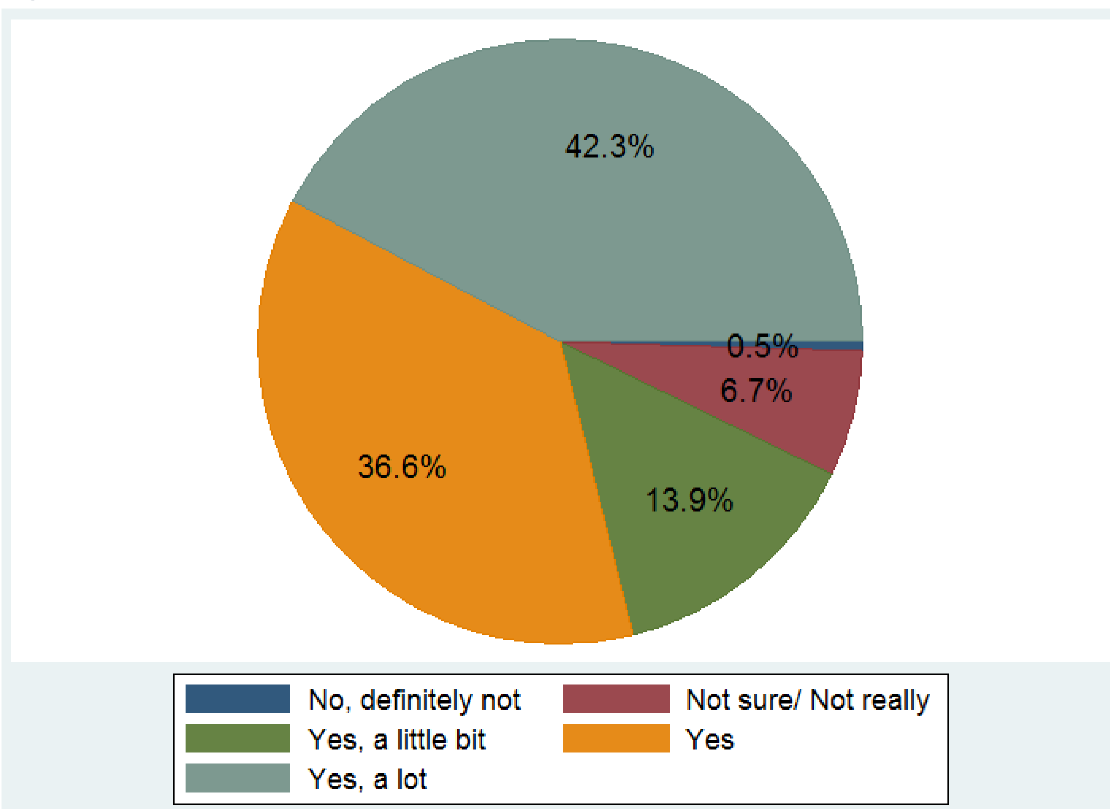


Figure C8: Do you have a best friend in your Khan Academy class?

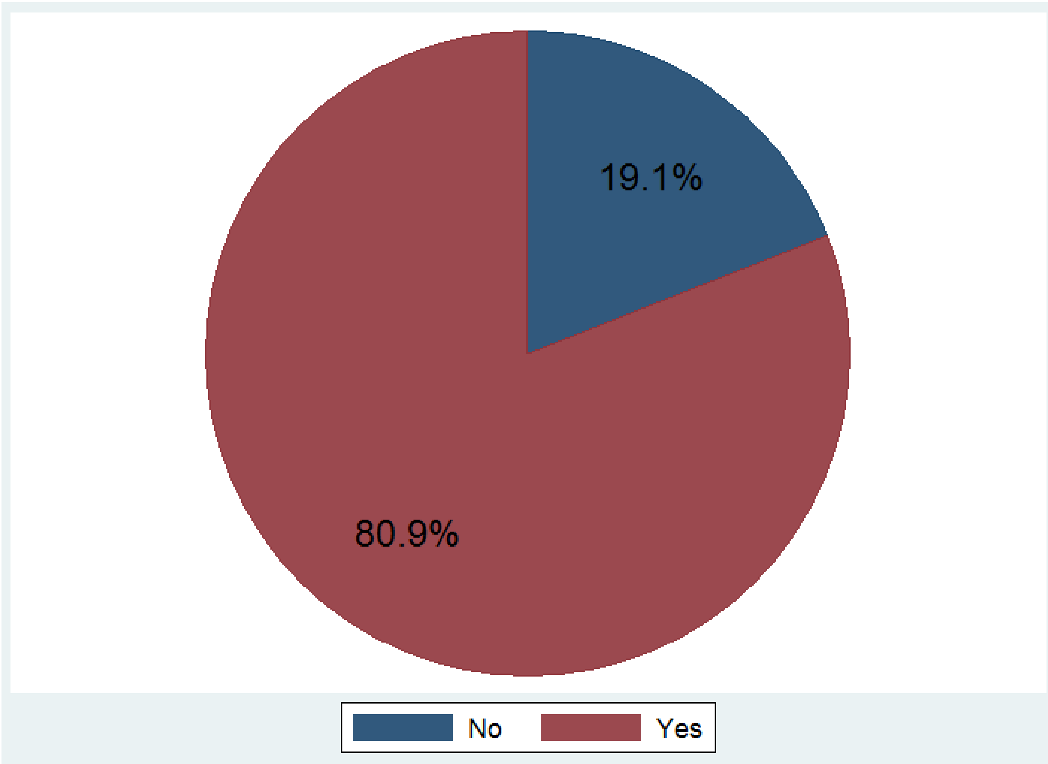



Figure C9: Have you ever told any of your friends about Khan Academy?

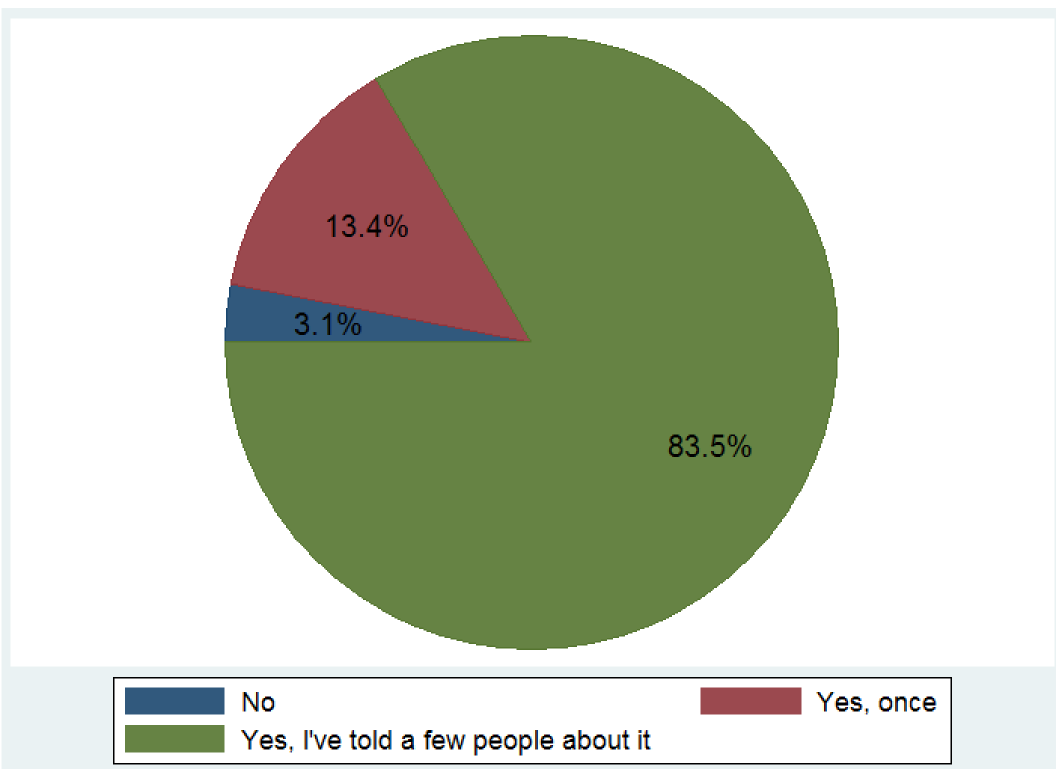


Figure C10: Have you ever used Khan Academy outside of your Khan Academy classes?

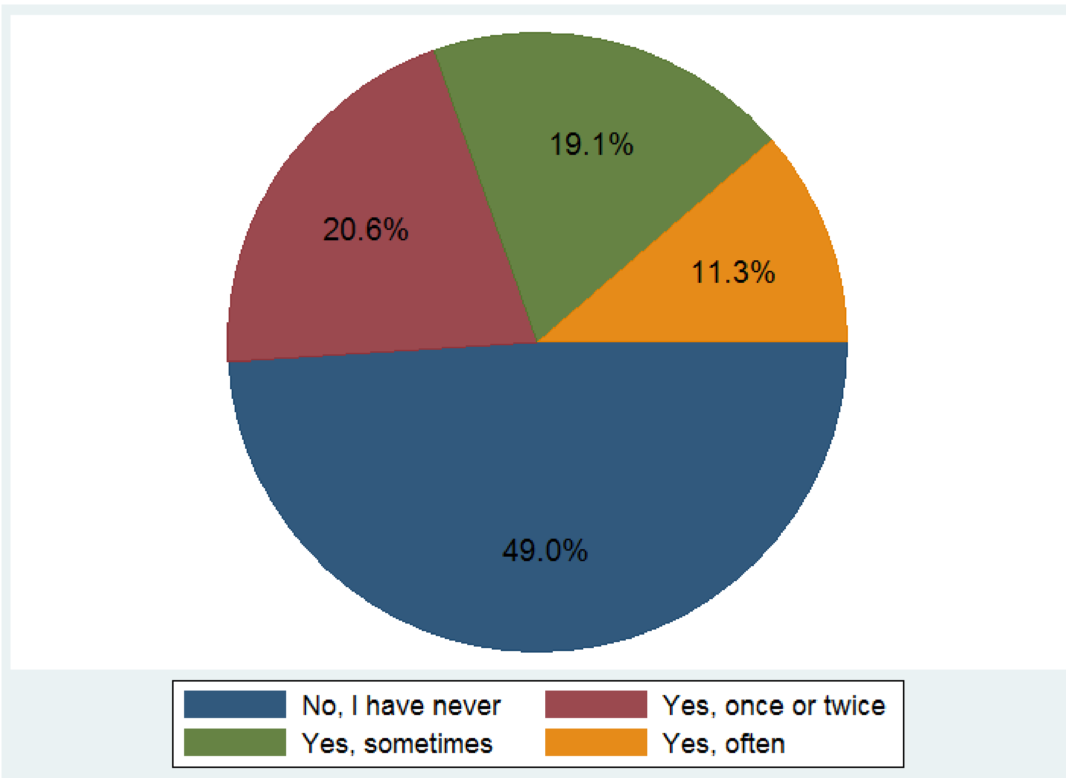



Figure C11: Who decided that you should apply for Khan Academy?

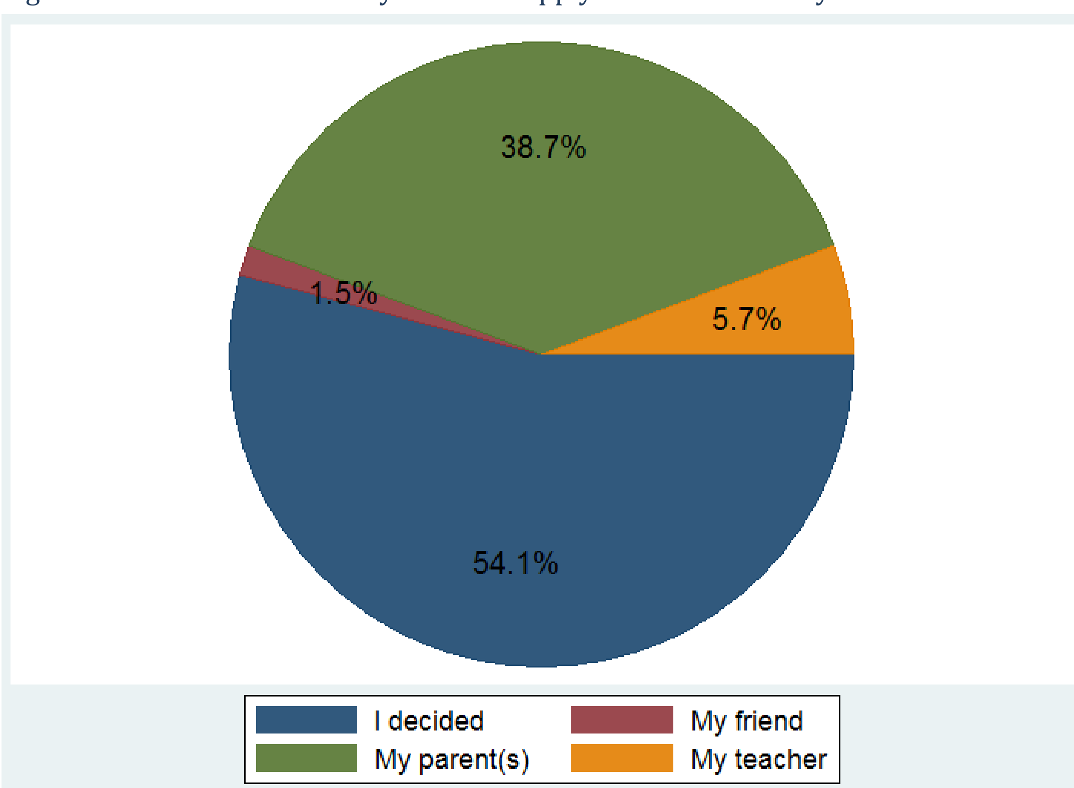
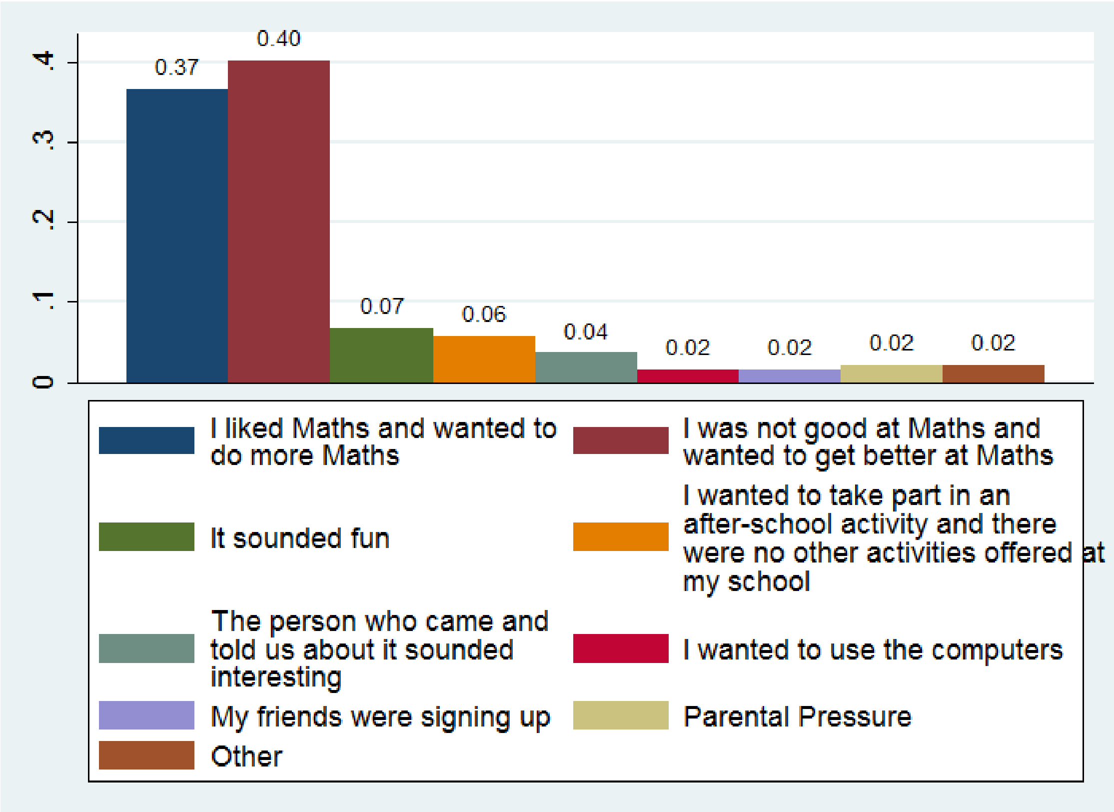


Figure C12: What was the main reason you wanted to join Khan Academy at the beginning of the year?




## Other comments

There a number of themes that emerged from the comments. Overall, the majority of learners enjoyed the programme and were satisfied with the service they were receiving. Many reiterated that they had found the Numeric classes helpful and would like to continue in Grade 9.

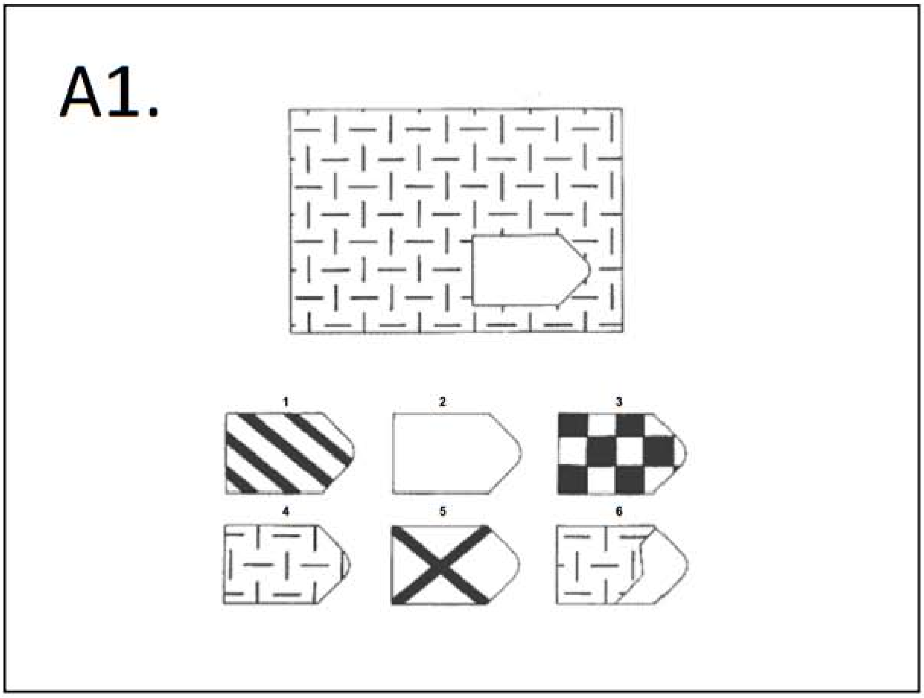
Food was mentioned repeatedly. A number of learners felt they were not being given enough food or asking for different types of food. Also, if there had been any problems with the delivery of food at any point throughout that year, this was mentioned. Essentially, this aspect of the programme was very important to many of the learners.

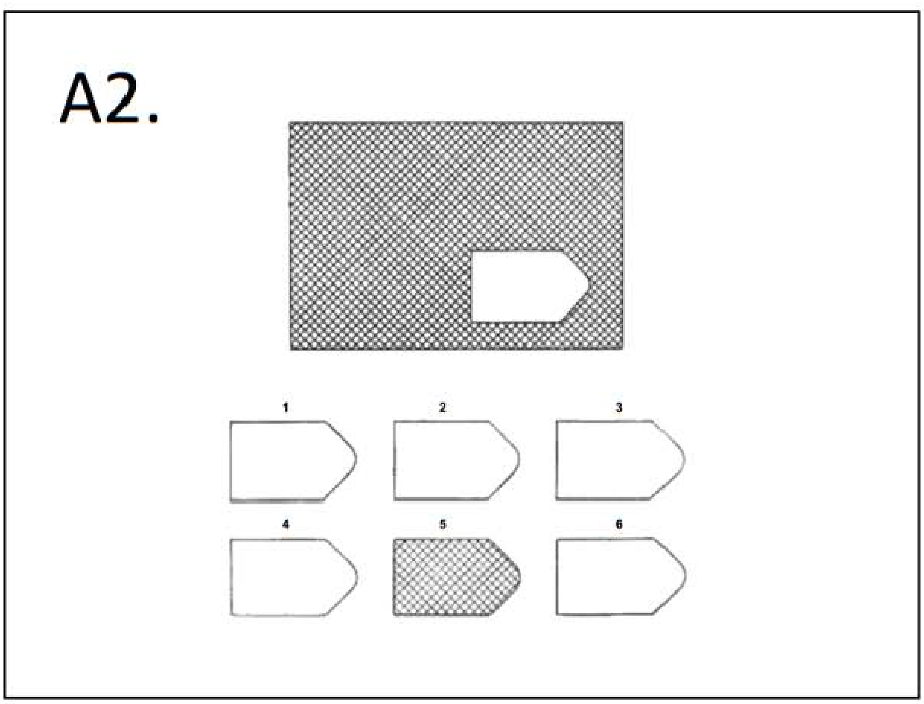
Numeric had held a Mathematics Camp in June which 50 learners attended. The learners that attended all thoroughly enjoyed it and many asked them to hold a similar event.

## 16.Appendix D

### Raven's Standard Progressive Matrices

The first two questions of the Raven's Standard progressive matrices are shown below as an example.



## English Literacy Test

The answer sheet to the English test is shown below

| | | | |
|---|---|---|---|
| | **English** | | **IID:** *(filled in by enumerator)* |
| | **SCHOOL:** | | |
| | **NAME and SURNAME:** | | |
| 1.<br>2. | **INSTRUCTIONS:**<br>Answer all the questions.<br>Answer the questions on the question paper. | | |
| | **QUESTION 1**<br>Please listen to the News Report read out to you and answer the questions that follow by circling the correct answer. | | |
| 1.1 | a) In the morning<br>b) At midday | c) In the early evening<br>d) late at night | |
| 1.2 | a) 1<br>b) 6 | c) 2<br>d) 102 | |
| 1.3 | a) arsonists<br>b) astronauts | c) the army<br>d) vandalists | |
| 1.4 | a) they died<br>b) they had to leave | c) they walked their dog<br>d) they called the police | |
| | **QUESTION 2**<br>Please circle answer which best describes the following words: | | |
| 2.1 | **Basic** | a) Not good<br>b) Fundamental<br>c) Holy<br>d) Beautiful | |
| 2.2 | **Aid** | a) Help<br>b) Robber<br>c) Ship<br>d) Couch | |
| 2.3 | **Audio** | a) Gratitude<br>b) Template<br>c) Sound<br>d) Argumentation | |
| 2.4 | **Encounter** | a) Meet<br>b) Erase<br>c) Shake<br>d) Complain | |
| 2.5 | **Repugnant** | a) Arctic<br>b) Fascinated<br>c) Disgusting<br>d) Jelly-like | |
| 2.6 | **Sentiment** | a) Travel Route<br>b) Blackmail<br>c) Defect<br>d) Opinion | |

## QUESTION 3

Read the following text and answer the questions that follow

**Words that you might wonder about:**
**Iron ore:** the metal or rock from which we get iron.
**Holds:** place on a ship for cargo (goods that are transported)
**Hull:** body of the ship
**Tug:** small powerful boat used to tow others
**Oil slick:** smooth patch of oil on the sea

### PETER, PAMELA AND PERCY SURVIVE THE OIL

On 12 June 2000, a large cargo ship, the *Treasure*, asked for shelter in Cape Town harbour – the beginning of what was to become a major oil-pollution disaster, but one with a happy ending.

The *Treasure* was carrying iron ore from Brazil to China and the captain thought that one of the ship's holds was leaking.

The South African authorities inspected the ship at sea but saw no damage. Inside the harbour, divers inspected the ship underwater and found a huge hole in its hull. The tug, *John Ross*, immediately towed the *Treasure* out of the harbour for repairs at sea.

But on 23 June the *Treasure* sank in 50 m of water near Melkbosstrand, leaking 400 tons of oil before divers could plug the hole. The huge oil slick from the ship floated across the waters and polluted beaches around Cape Town – and especially the ecologically sensitive beaches of Robben Island and Dassen Island. These islands are the breeding grounds for more than half of the African Penguin colonies in the world.

(Adapted from *Language, Literacy and Communication Grade 5*: Jean du Plessis, Penny Hansen, Bev Rau and Audrey Wrightman, NASOU 2001 pages 112-113)

| | |
|---|---|
| 3.1 | Please arrange the following words in Alphabetical order : **Oil, African , Harbour, Underwater**<br>____________________________________________ |
| 3.2 | What is the *Treasure*?<br>a) The oil which the ship was transporting<br>b) The name of the ship<br>c) The African penguins<br>d) The shelter provided by the Cape Town harbour |
| 3.3 | How many years ago did this disaster occur?<br>a) 10<br>b) 13<br>c) 50<br>d) 2000 |
| 3.4 | In which season did this disaster take place?<br>a) Summer<br>b) Autumn<br>c) Winter<br>d) Spring |
| 3.5 | Upon inspection, what was found to be wrong with the ship?<br>a) Its cargo was a pollutant.<br>b) It needed to come to the harbour.<br>c) It was too close to the penguins.<br>d) Its hull was damaged. |
| 3.6 | Where was the ship taken to be mended?<br>a) Underwater<br>b) To Dassen Island<br>c) Into the harbour<br>d) Out to sea |

Question 1 was a listening exercise. The questions and the article that were read out to the learners are shown below:

# Questions

1. When did the explosion take place?
2. How many people were hurt in the fire?
3. Who do the police think started the fire?
4. What happened to the people living near the factory?

## News Report[67]

Late last night, there was an explosion in a fireworks factory in Digbeth, a district of Birmingham.

Nobody was in the building at the time of the explosion but a couple out walking their dog were badly burned. It took firemen six hours to get the fire under control.

One hundred and two people living near the factory were evacuated and Digbeth has been closed off until Wednesday evening.

The police think that arsonists started the fire deliberately and they would like anyone who saw anything suspicious to call them on 021 684 5583 as soon as possible.

[67] News Report and Questions adapted from Entry Functional English listening tasks x 4 at Skills Workshop by Hazel Hughes. Used with Permission.

## Computer Literacy Test

### Computer Questions

### A. 1

How would you go about safely shutting down your computer.

a) Turn it off using the power button on your computer
b) First click the start button on your desktop
c) Pull out the power plug
d) Press the escape key on your keyboard

### A. 2

How would you get this box to appear

a) Double click on the icon
b) Right click on the icon
c) Left click on the icon and then press Enter
d) Highlight the icon

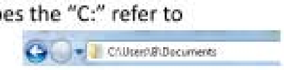
### A. 3

Which of the following can be placed in a "folder" ( )?

a) a file
b) a folder
c) a shortcut
d) all of the above

### A. 4

What does the "C:" refer to



a) a file
b) a folder
c) a hard drive
d) a command

### A. 5

Which file extensions are both used for picture files?

a) .txt and .doc
b) .jpg and .gif
c) .mp3 and .pdf
d) .xls and .gif

## A. 6

If you have Microsoft Paint open, what would you do to close the programme.

a) Click the cross in the top right corner 

b) Click the save button 

c) Press the Esc key 


d) Click the start menu button 


## A. 7

Which of these can store the most data?

a) A CD-ROM 

b) An external hard drive 

c) A flash drive 

d) A DVD disk 

## A. 8

Which is the largest storage capacity?

a) 16 MB

b) 16 kB

c) 16 GB

d) 16 bytes

## A. 9

To run, all computers must have:

a) A word processor

b) An operating system

c) A printer

d) An anti-virus programme

## A. 10

What is considered to be the "brains" of a computer?

a) The keyboard

b) The central processing unit (CPU)

c) The power cord

d) The screen

## A. 11

Match the text on the right with the keys on the left needed to type it.

| | |
|---|---|
| a) Caps Lock | Happy Birthday Xolisani! |
| b) Tab | if you can<br>keep your head |
| c) Enter | CAREFUL |
| d) Shift | mistakee → mistake |
| e) Backspace | 1. tree. 2. kite |

## A. 12

When do you use a router?

a) When you want to create a presentation
b) When you want to access the Internet
c) When you want to draw a picture
d) All of the above

## A. 13

Which one is an email address?

a) http://vianet.com/index.htm
b) D:nnEmailnStandard
c) joesmart@billme.com
d) Chaminade.org/teachers/mailaddresses

## A. 14

Which of these is a browser?

a) Microsoft Windows
b) Microsoft Excel
c) Paint
d) Internet Explorer

## D. 1

1. Have you ever seen this sign before

2. What does one do at www.google.com
   a) Look for friends
   b) Search for websites
   c) Write and receive e-mails
   d) Make calculations

## D. 2

1. Have you ever seen this sign before

2. What is Wikipedia?
   a) A translation website
   b) An online bookstore
   c) Free online file storage
   d) A free online encyclopedia

## D. 3

1. Have you ever seen these signs before

2. What can one do on facebook?
   a) Chat to friends
   b) Upload photos
   c) Play games
   d) All of the above

## D. 4

1. Have you ever seen this sign before
2. What can one do with Skype?
   a) Plan your holiday online
   b) Make calls online
   c) Upload and share pictures
   d) Play games

## D. 5

1. Have you ever seen this sign before
2. What can one do on YouTube?
   a) Play games
   b) Write and share stories
   c) Watch and share videos
   d) Draw pictures

## D. 6

1. Have you ever seen this logo before
2. What can one do at www.gumtree.co.za?
   a) Buy second-hand goods and look for jobs
   b) Find information about plants and farming
   c) Download recipes and Do-It-Yourself tricks
   d) Share the latest music and movies