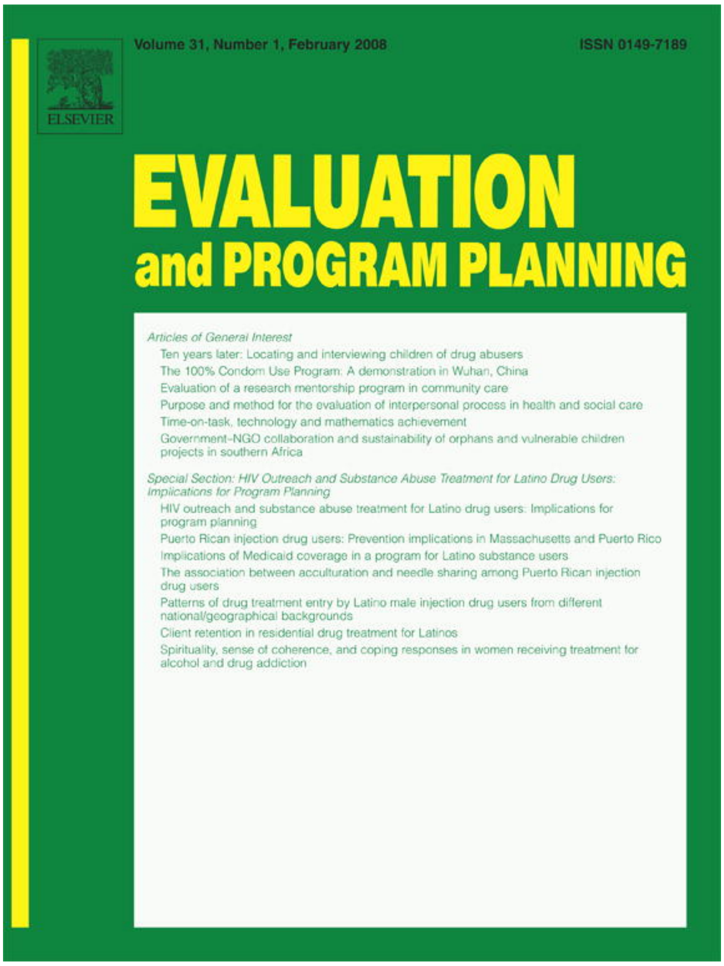
Volume 31, Number 1, February 2008
ISSN 0149-7189
ELSEVIER
EVALUATION
and PROGRAM PLANNING
Articles of General Interest
Ten years later: Locating and interviewing children of drug abusers
The 100% Condom Use Program: A demonstration in Wuhan, China
Evaluation of a research mentorship program in community care
Purpose and method for the evaluation of interpersonal process in health and social care
Time-on-task, technology and mathematics achievement
Government-NGO collaboration and sustainability of orphans and vulnerable children projects in southern Africa
Special Section: HIV Outreach and Substance Abuse Treatment for Latino Drug Users: Implications for Program Planning
HIV outreach and substance abuse treatment for Latino drug users: Implications for program planning
Puerto Rican injection drug users: Prevention implications in Massachusetts and Puerto Rico
Implications of Medicaid coverage in a program for Latino substance users
The association between acculturation and needle sharing among Puerto Rican injection drug users
Patterns of drug treatment entry by Latino male injection drug users from different national/geographical backgrounds
Client retention in residential drug treatment for Latinos
Spirituality, sense of coherence, and coping responses in women receiving treatment for alcohol and drug addiction

ELSEVIER

Evaluation and Program Planning 31 (2008) 41–50

EVALUATION and PROGRAM PLANNING

www.elsevier.com/locate/evalprogplan

Time-on-task, technology and mathematics achievement

Johann Louw[a,*], Johan Muller[b], Colin Tredoux[a]

[a]Department of Psychology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa
[b]School of Education at the University of Cape Town, Private Bag, Rondebosch 7701, South Africa

Abstract

Information and communication technologies hold much promise for use in education in developing countries. This study reports on an evaluation conducted on the introduction of computers in the delivery of the mathematics curriculum in one of the provinces of South Africa. Although the request was for an outcome evaluation very early in the implementation of the program, it was tailored in such a way as to fulfill a more formative role. Despite substantial variability in implementation, and in most cases with very weak exposure of the learners to the intervention, sufficient evidence emerged to indicate that this mode of curriculum delivery may be effective. Improvement in mathematics performance was related to a range of variables: some concerned classroom teaching practices, some referred to social differences between the learners, and some to the specific intervention. The strongest of these predictors in the sample was the strength of the intervention: the more time learners spent on using the software to study mathematics, the more improvement they showed from 1 year to the next in their performance in the subject.

Keywords: Education; Mathematics; Technology; South Africa; Evaluation

1. Introduction

Current discussions in educational settings acknowledge the value of incorporating information and communication technologies (ICTs) into school curricula. This value however is not added automatically, and careful planning and sufficient support is required to improve learning and teaching. For example: teachers must receive sufficient training in how to use computers as well as in how to use content-specific software (BECTA, 2000; Howell & Lundall, 2000); the choice of software must be appropriate to the learning tasks (Chang, Sung, & Chen, 2001; Veerman, Andriessen, & Kanselaar, 2000); pupils and teachers must have good access to the new technology (Wenglinsky, 1998); and schools should have sufficient financial support to maintain and sustain the level of provision (Howell & Lundall, 2000). If this is done, there is ample evidence that ICTs can enhance learner performance and develop teachers professionally (e.g. Dwyer, 1994; Interactive Educational Systems Design, 1996; Jurich, 1999; SIIA, 2000; Soloway, Lockhead, & Clement, 1982).

There are few examples of ICTs being used as tools in developing countries, where the adequacy of implementation and support is often not to be taken for granted, and there are even fewer well-designed evaluations of their implementation and use. In this paper, we report on the outcome evaluation of an ICT schools-based intervention in poor, disadvantaged schools around Cape Town in South Africa. As we will show, the evaluation tells a tale about ICT use and effectiveness that occurs frequently in countries around the world.

South African learners are able to take mathematics and science at one of two levels for their exit certificate: at the higher grade or at standard grade. Higher grade passes are the prerequisite for entry into many university courses. The low number of high school learners in South Africa who matriculate (graduate at the end of senior school) with mathematics and science as subjects at the higher grade (university entrance) level has been a source of concern for some time. On the one hand, the number of candidates taking mathematics as their final year subject increased by 90.4% between 1991 and 2003, yet the number taking it

*Corresponding author. Fax: +27 21 6504104.
E-mail address: Johann.Louw@uct.ac.za (J. Louw).

0149-7189/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.evalprogplan.2007.11.001

Table 1
Number of learners passing mathematics and science at the higher grade, 1991–2003

Year	Total no. of grade 12 passing candidates	Mathematics higher grade passes	Science higher grade passes
1991	210,314	20,677	23,109
1992	243,611	21,558	24,172
1993	226,943	20,365	23,835
1995	283,742	29,475	34,895
1996	279,487	22,416	25,462
1997	264,795	19,575	26,658
1998	272,488	20,130	26,473
1999	249,831	19,854	24,191
2000	283,294	19,327	23,344
2001	277,206	19,504	24,280
2002	305,774	20,528	24,888
2003	322,492	23,412	26,067

Source: Clynick and Lee (2004, p. 32).

at the university entrance level dropped by 32.95% over the same period. Moreover, as the following table shows, the number of mathematics passes at the university entrance level has remained virtually static over the same period. The science passes follow the same pattern (Table 1).

When it comes to race differences, race being generally accepted as a proxy for educational disadvantage, the results are even starker. Only 1% of all African matriculation candidates in 2003 passed mathematics at the higher grade with a D symbol or higher, usually required to get into the more prestigious university courses, at a time when African candidates outnumbered white candidates by seven to one (Kane-Berman, 2006). And when it comes to the Western Cape Province, where the Khanya Project was launched, the number of African passes in mathematics at the higher grade in 2002 was a scant 200 (Clynick & Lee, 2004, p. 50). It is not hard to see why, by 2002, not only government but academia and the private sector had all become seriously worried about the capacity of the public school system to produce high level mathematics and science graduates for the country.

Quite why mathematics and science scores have remained so impervious to attempts to reform schooling during the first 10 years of democracy in South Africa is the subject of much debate. To be sure, there is little disagreement about the causes: a legacy of poor resourcing, poor teacher preparation, and a curriculum that is not explicit about the performance standards expected are amongst the most common causes cited (Taylor, Muller, & Vinjevold, 2003). By 2001, the desire to make a tangible difference to this parlous state of affairs was common cause. If computers and software could improve learners' achievement in mathematics and science, computers in South African classrooms might prove invaluable to a country that faced such low levels of achievement.

1.1. The Khanya project

The Khanya Technology in Education Project (for more information, see www.khanya.co.za) was officially established in May 2001 by the Western Cape Education Department (WCED). The degree of initial support for it can be gauged by the very generous funding it received. The Khanya Project is an ambitious plan to strengthen curriculum delivery via the use of information and communication technology (ICT). Considering that there are nearly 1500 schools in the Western Cape, and that every two extra schools included in the project will cost an extra R1 million (approximately $150,000, and a figure which seems to exclude on-going support and maintenance), the scale and scope of the project is not small. It is targeted at secondary schools in the Western Cape, and by 2006 it had helped 595 schools to use technology effectively. This paper reports on an evaluation of only one aspect of this intervention, relating to the delivery of the mathematics curriculum in grades 11 and 12 in the poorest schools in the region.

To qualify for inclusion in the project, schools had to be "poor", according to the WCED poverty index of schools, and reasonably well-managed, again according to the Department's "management index" of schools. Once a school qualified, a secure room ("the computer laboratory") had to be found that was able to house approximately 40 computers per school, again meeting certain minimum criteria. In some schools, computers were installed in the classrooms.

These computers provided the platform to support proprietary commercially available software developed by a local software company, to deliver the mathematics curriculum. The software is called MasterMaths (MM), a tutoring system that consists of the following components: (a) the MM tutor (a person) who oversees and coordinates students' engagement with the system; (b) the tutor administrator, which is the software used by the tutor to control the functioning of the system; (c) the teaching/learning modules, including test modules, which students have access to through software called the M^2 browser; and (d) module notes and worksheets that are intended to reinforce teaching and learning (see www.m2maths.co.za). The modules accessible through the browser cover the mathematics curriculum from grade 7 to grade 12 and there are 293 teaching modules available.

2. Method

2.1. Planning the evaluation

A year after the project was officially established, an evaluation tender was advertised, and awarded to the first two authors. That there was something substantial to evaluate a year after the project's establishment is rare in the history of education projects in this country. Furthermore, the project was ambitious in asking for an outcome evaluation, in terms of improvement in learner performance in mathematics. It is unusual for a newly launched program, especially such a complex one, to ask for an outcome evaluation so early in its existence. While this is indeed commendable, there is also a serious risk attached to it. Programs have unique developmental trajectories, and outcome evaluations usually are not recommended in the early stages of their development.

Negotiations with stakeholders indicated that an outcome evaluation was not possible in all schools. In consultation with these stakeholders, we negotiated to conduct the study in five overlapping phases:

1. Conduct an evaluability assessment.
2. In a sample of schools, conduct an assessment of the technical installation and training provided to teachers.
3. Obtain expert opinion on the software.
4. Conduct an implementation assessment.
5. Carry out an early evaluation of outcomes.

Space does not allow for an examination of all these steps in this paper. We concentrate on the outcome evaluation component, because we believe there are some interesting procedures and findings to report. Where appropriate, aspects of the earlier phases will be alluded to, where this is required for an understanding of the outcome evaluation.

Suffice it to say that the evaluability assessment, conducted along the guidelines provided by Wholey (2004), convinced us that there was general acceptance and support for Khanya as a project among all stakeholders regarding the project and evaluation goals and objectives, the project's performance criteria, and the project and evaluation priorities, to proceed with the evaluation. Furthermore, an independent snapshot, conducted via interviews and site visits, of how the project had been implemented in a small sample of schools, revealed that its installation of technology seemed to be as good as could reasonably be expected.

2.2. *An evaluation of outcomes*

In the evaluability assessment, consensus was reached among the stakeholders that the outcome question of interest was: does learner performance in grade 12 mathematics improve as a result of curriculum delivery via MM? The logic behind the intervention was as follows: the principal cause of the low achievement levels in mathematics was assumed to be the low capacity of teachers, and that ICTs could compensate for low-capacity teachers. Indeed, these were the core suppositions driving Khanya, and we confirmed this via interviews with primary stakeholders. Furthermore, since the source of the problem was seen as poor capacity of teachers, it resulted in poor curriculum delivery. We took that to mean that Khanya's computers and software were expected to provide the coverage of the curriculum that poorly trained teachers were not able to provide. In other words, if the symptom was poor learner attainment, and the cause was poor curricular coverage because of poorly trained teachers, then the cure was to improve curricular "dosage" by means of IT hardware and software. This was the logic we set out to evaluate.

A quasi-experimental design was used to obtain comparative learner performance data in five treatment/experimental and five control schools.

To set up a non-equivalent control group design, five schools were randomly selected from a list of the 146 schools that had started to receive the intervention prior to 2003. In addition, they had to be operational (in Khanya terms) for at least 12 months. Five of these 146 schools were selected at random, subject to two exclusion rules: the technology had to be installed by the project, with no other ICT curriculum delivery in place at the school; and the schools had to have a below average poverty index. These five schools also were the ones at which we conducted our snapshot investigation of the technology installation, so that we were confident that for the five experimental schools the installation had gone as planned.

A further five schools, which had been selected for participation in the project, but where the curriculum was not yet being delivered via the software, were identified and matched to the five experimental schools in terms of geographical location and poverty index. They acted as the non-treatment or control group, and were to reach the implementation and delivery stage by November 2003, by which time the data collection for this study was to have been completed.

As indicated earlier, the main outcome of interest was learner performance in mathematics. We took as measures of performance marks obtained at the end of grade 11 in mathematics (as a "before" measure), and marks obtained in the grade 12 final examination (as an "after" measure). The grade 12 examinations are set nationally, and are thus comparable across schools. Unfortunately, the "before" measure was less satisfactory, since schools set their own, often inconsistent, examinations. Nevertheless, two main considerations settled the matter: standardised mathematics tests for grade 12 were not available in South Africa, and it would have been too costly and time-consuming to develop them for this study; and learner performance in grade 12 mathematics was exactly what the intervention was designed to affect.

2.3. *Sample*

A total of 580 learners (271 in the control group, and 309 in the experimental condition) constituted the study sample. Fifty-one percent of the sample were females, and 49% males. Approximately 43% reported that they usually spoke Afrikaans at home, and 10% English. The remaining 47% were presumed to speak an African language at home. Several indices that were collected in the self-report questionnaire suggest that the study sample was drawn from a low to moderately low socio-economic group, confirming the school's standing on the WCED's poverty index. Seventy-nine percent reported that they have either no books at home, or a total number that would fill one shelf; and that approximately six people lived in their home. More than 70% of the parents did not finish high school.

2.4. Measures

At the school level, data were collected directly from the WCED's data bank. Information on more than 70 variables for the schools in our study was collected. These included indices of the quality of school management, absenteeism, and repeat rates (overall, and for mathematics in particular), the total percentage of the school budget invested in ICTs, and the degree to which the school principal supported the Khanya project.

At the learner level, we administered a lengthy questionnaire to all grade 12 mathematics learners at the 10 schools. A great many variables were assessed, including pre-existing level of mathematics ability, degree of access to computers outside of school, time spent on computers both inside and outside of school, and on the MM system, computer literacy, confidence in using IT, motivation for mathematics, degree of enjoyment of learning mathematics, intention/commitment to study further (i.e. after school), home language, and parental encouragement/placing of importance on performance at mathematics.

The grade 11 marks in mathematics for learners were obtained directly from the schools.

The grade 12 marks in mathematics were downloaded from the WCED data base and were sent directly to the evaluation team in a data file.

Computer usage logs were downloaded at three of the experimental schools (these were not available for the other two).

3. Results

The overwhelming majority of pupils in the study sample started grade 12 mathematics at the standard grade level (541/578 = 94%), and this majority increased by the end of the 2003 school year (550/578 = 95%). The number of higher grade enrolments was too low to include in further meaningful analysis, and all further analyses are restricted to pupils who were registered at the standard grade level at the end of grade 11 in 2002. (In fact, this is one of the objectives of the Khanya project: to increase the percentage of learners taking mathematics at higher grade.)

The overall level of pre-intervention performance in mathematics across the 10 schools was comparatively poor. The average mark across experimental and control schools was 34%, as is shown in Table 2. There was also considerable variation between schools. Thus, 58% of pupils scored less than 40% on the final grade 11 mathematics examination, but for some schools this figure was as high as 96%, and for others it was as low as 15%. The experimental and control schools differed significantly on the average grade 11 mathematics mark (a one-way analysis of variance was conducted, which yielded the following statistics: $F = 35.44$; d.f. = 1,523; M.S.E. = 366.72; $p<0.001$; $\eta^2 = 0.06$), with the experimental schools scoring an average of almost 10% more than the control schools, as is shown in Table 2. However, this quite large difference of approximately one-half a pooled standard deviation was largely due to the very poor performance of one control school in the grade 11 examination (which had an average grade 11 mark of 15%).

Table 2
Performance of control and experimental schools in grade 11 (2002) and grade 12 (2003) mathematics

	Grade 11			Grade 12		
	Mean	*S*	*N*	Mean	*S*	*N*
Control groups	29.01	21.44	253	27.5	19.74	260
Experimental groups	38.97	16.75	272	35.66	20.92	274

The difference between the experimental and control groups of schools on the grade 11 mathematics results means that it was not possible to compare them directly on the grade 12 mathematics result. This problem was dealt with by using the difference between the grade 11 and grade 12 results for all comparisons of experimental and control schools, and in all attempts to model performance in mathematics on the basis of demographic, motivational, or program variables.

3.1. Comparison of groups on difference scores, grades 11–12

Difference scores between the WCED grade 12 final mathematics result and the school grade 11 final mathematics result were computed for individual learners, and the experimental and control groups compared on their respective averages for these scores. This comparison assesses the ability of the Khanya intervention to improve performance of learners in the experimental schools, relative to performance of learners in the control schools.

In technical terms, the decrease over time, i.e. between the final examinations of grade 11 and grade 12, taking the experimental and control schools as a whole, is statistically significant, and the difference between the experimental and control schools, taken as a whole, is statistically significant, but the interaction is not (a repeated measures mixed model analysis of variance shows this: for the time factor $F = 12.94$; d.f. = 1,516; M.S.E. = 111; $p<0.0003$; $\eta^2 = 0.003$; for the experimental/control factor $F = 30.6$; d.f. = 1,516; M.S.E. = 671; $p<0.0001$; $\eta^2 = 0.05$; and for the interaction $F = 2.07$; d.f. = 1,516; M.S.E. = 111; $p>0.15$). In other words, the results for the grade 12 examination are lower, on average, than the results for the grade 11 examination, and the group of experimental schools consistently scores higher than the group of control schools. The group of experimental schools starts grade 12 scoring higher than the control schools and they end grade 12 scoring higher than the control schools.

Earlier we pointed out that the low performance of the control schools on the pre-test was largely due to the very poor and unusual performance of one control school on the grade 11 examination. The comparatively large numbers of students enrolled in this school, and their poor performance in the grade 11 mathematics examination negatively weighted the results of the group of control schools. The comparison of control and experimental conditions set out above may therefore profit from an examination of the specific school pairings we mentioned earlier. By examining school pairings, we will be able to isolate this strong biasing effect to the single pairing which involves this particular school. The other pairings may reveal a different picture about the efficacy of the Khanya intervention.

Accordingly, each of the school pairings was examined, comparing experimental and control schools within the pairing on the difference between grade 11 mathematics and grade 12 mathematics results. As explained earlier, matching was effected on the basis of geographical location, poverty index, and school management index. Table 3 reports the results for comparisons of the matched schools, including one-way analyses of variance tests of differences between each matched pair. The table shows that three of the individual comparisons are statistically significant, and of these two are in the expected direction, and one in the opposite direction. The evidence in favour of the effectiveness of the intervention is thus not clear from these comparisons.

3.2. *Strength of the intervention: description and analysis*

The key aspect of the Khanya intervention around mathematics is the amount of time that learners are exposed to the MM program. If learners in experimental schools receive low exposure to the program, one would expect there to be little improvement relative to learners in the control school. Conversely, if learners in experimental schools receive high exposure to the program one expects there to be improvement in mathematics performance relative to learners in the control schools. We attempted to obtain information from log files on computer network servers in the experimental schools in order to chart the strength of the intervention, and in order to understand the possible relationship between the intervention and performance at mathematics in grade 12. Unfortunately, we were only able to obtain log files from three of the experimental schools. One of the experimental schools inadvertently deleted their log files, and another appeared to have a poorly functioning laboratory (despite our earlier check on its installation), and did not appear to have any log files at all (the server was down for six months during that year).

Analysis of the log files reveals that learners spent very little time logged on to the MM program, and they logged on comparatively infrequently. Table 4 shows that the average time logged on to the MM program was only 158 min, and that on average learners logged on to MM approximately seven times. These are not estimates of the total intervention experienced by learners in the three experimental schools in question, since the log files reflect the period from April to September 2003 only. Also, learners use the MM system in pairs, so the amount of time spent on the system may not accurately reflect the amount of intervention experienced by learners. However, the statistics reported in the table are so low as to raise serious concerns about the implementation of the intervention.

It is also clear from Table 4 that pupils spend different amounts of time using the MM system. This suggests a second analytic framework for evaluating the Khanya intervention. Pupils who spend more time on the system should show better mathematics performance than pupils who spend less time, all other things being equal. The difficulty, of course, is ensuring that "all other things are equal". It would not be a very revealing analysis to chart the relationship between grade 12 mathematics performance and amount of time spent on MM, since learners

Table 4
Descriptive statistics for "strength" of intervention delivered to learners

	Mean	*S*	Minimum	Maximum	*N*
MasterMaths total time spent	158.13	102.6	5	459	126
MasterMaths total no. of sessions	7.4	4.3	1	20	126

Note: values in the first row are minutes, and values in the second row are number of sessions longer than 5 min.

Table 3
Difference score statistics for individual matched pairs of experimental and control schools

School pair	Experimental difference score	Control difference score	*F*	d.f.	*p*	M.S.E.	η^2	Direction
A	−1.66	10.65	32.54	1,170	<0.0001	120.1	0.16	E<C
B	−1.34	−20.68	63.83	1,159	<0.0001	194.1	0.29	E>C
C	−7.2	−6.36	0.07	1,35	>0.8	77.33	0.002	E = C
D	−4.00	−19.51	48.36	1,90	<0.0001	85.8	0.34	E>C
E	−3.65	−3.58	0.0007	1,60	>0.98	104.8	0.001	E = C

who do better at mathematics may be expected to spend more time on MM (i.e. the direction of the relationship is directly opposite to that predicted). This could be due to greater motivation or higher mathematics self-esteem, or indeed any number of other factors. It seems best to chart the relationship between improvement in mathematics performance across grades 11 and 12 and time spent on MM, as such an analysis attempts to remove pre-existing differences in mathematics performance between learners, and by implication, the variables that are associated with individual differences in performance.

We therefore conducted correlational analyses of the relationship between improvement in mathematics performance, and the amount of time spent on the MM system. This analysis revealed that the relationship between amount of time spent on MM and improvement in mathematics performance is positive, statistically significant, and moderate in strength ($r = 0.37$; $n = 125$; $p < 0.001$). The same is true for the relationship between number of sessions spent on the MM system and improvement in mathematics performance ($r = 0.36$; $n = 125$; $p < 0.001$). This may be considered somewhat surprising, since there is a fairly severe restriction of range on both of these indices of exposure to the MM system: most learners spend fewer than 100 min and fewer than 10 sessions on the system in total. (An anonymous reviewer has questioned our interpretation here, arguing that the skewness in our data is likely to have inflated the correlation between time spent on Master Maths and improvement in Mathematics performance. He/she suggested re-calculating the correlation after log-transforming the predictor. We did this, and found a slightly increased correlation of $r = 0.39$.) The relationship between the strength of the intervention and improvement in mathematics performance is likely to be much stronger when the strength of intervention is allowed to range more widely, i.e. when learners spend more time on average on the system, and when there is more variation in the amount of time they spend on the system. One should also note that the increase in performance shown as a function of time, and sessions spent on MM is counter to the general decrease between grades 11 and 12 that the cohort shows as a whole.

In order to explore further the nature of this relationship, we repeated the analysis for each of the three schools listed above, but focused only on the outcome variable recording amount of time spent on the MM system. Table 5 reports Pearson correlations and other appropriate statistics.

Table 5
Pearson correlation coefficients for the relationship between time spent on MasterMaths and improvement in mathematics performance, per school

School	Coefficient (*p*-value, pairs)
A	$r = 0.32$ ($p > 0.08$, $n = 29$)
B	$r = 0.47$ ($p < 0.009$, $n = 29$)
C	$r = 0.46$ ($p < 0.001$, $n = 68$)

These results reinforce the impression that there is indeed a relationship between the intervention and improvement in mathematics performance. The relationship is present in each of the three schools, but it is much stronger (and statistically significant) in two of the three. It is also clear in each of the three cases that the strength of the relationship is likely to be underestimated, since there is restriction of range on the predictor variable (time spent on MM).

3.3. *Predictors of improved mathematics performance*

We have established in the preceding analysis that the intervention appears to have a positive effect on mathematics performance, specifically improvement between grades 11 and 12. However, it is unlikely that it is the only predictor of improvement in mathematics performance. It is also possible, although unlikely, that learners who experienced higher "strengths of MM dosage" are alike in other ways, and that those similarities explain the improvement in mathematics performance rather than the intervention itself. In order to answer these questions, and also simply to understand what variables are associated with improvement, and what outcomes might be produced by the Khanya intervention other than mathematics improvement, we conducted a number of correlational and regression analyses.

3.3.1. *Non-performance differences between control and experimental schools*

The Khanya intervention intends to produce improvement in mathematics performance. On the way to produce this improvement, however, it should produce changes in other variables that could lead to the desired improvement, for instance motivation, or mathematics self-esteem. We did not have a clear idea of what these variables would be, nor it is clear in the program theory underlying the intervention. We therefore examined differences between the control and experimental schools on a wide variety of non-performance outcomes, as an attempt to elaborate or chart the routes through which Khanya may or may not bring about improvement in mathematics performance. Table 6 lists variables on which the experimental and control groups differed significantly.

It is clear from the table that the experimental group scores higher on a number of items than the control group. There are some they should obviously be scoring higher on, such as whether "the teacher uses a computer to demonstrate ideas in mathematics", but there are also a number of items that relate to the teaching of mathematics on which they also score higher. Thus, the experimental group appears to be shown more frequently how to do mathematics problems, they report that they copy notes from the board more frequently, that they use calculators more frequently, that the teacher requires more homework from them, and that the teacher explains mathematics rules

Table 6
Non-performance differences between the experimental and control schools

Mathematics is difficult (E < C)
The teacher shows us how to do mathematics problems (E > C)
We copy notes from the board (E > C)
We work on mathematics projects (E < C)
We use calculators (E > C)
We use computers (E > C)
The teacher gives us homework (E > C)
The teacher uses a computer to demonstrate ideas in mathematics (E > C)
The teacher explains the rules and definitions (E > C)
The teacher ask us what we know related to the new topic (E > C)
How many times per week do you get mathematics homework? (E > C)

Note: the term "E > C" indicates that the experimental schools recorded a higher score than the control schools.

and definitions more frequently. They find mathematics less difficult than the control group finds it, and they do fewer mathematics projects than the control group.

On the face of it, there are a number of differences between the experimental and control schools, and many of these are on variables that seem plausible candidates for improving mathematics performance. The difficulty is in deciding whether they are a function of the Khanya intervention, or whether they are pre-existing differences due to different school or teaching practices, or less plausibly, whether they are differences produced by something other than the Khanya intervention, but contemporaneous to it. We will make an attempt in a later section of this paper to examine whether the effects of the Khanya intervention—as indexed by the time spent on MM by learners—are dependent, or perhaps even an artefact of these variables, but that is as far as we can take it in this analysis without conducting further research.

3.3.2. Predictors of improved performance: zero order relations

The surveys we conducted among learners at the 10 target schools in 2003 were used to obtain information about a wide variety of variables. We correlated mathematics improvement scores (i.e. the difference scores indexing change between grade 12 and grade 11) with these variables, and found a number of statistically significant correlations. These are reported in Table 7. A number of variables are correlated with improved performance in mathematics, including time spent on MM, some motivational variables (e.g. the perception that effort in mathematics will lead to better results), some teaching variables (e.g. the reported extent to which teachers check homework), and some specific perceptions about the role of computers in learning mathematics (e.g. the belief that the use of computers can have a positive effect on schoolwork).

Of course, these variables should not be considered independently, as many are inter-correlated. Similarly, some of the variables that are significant predictors are likely to be proxies for other phenomena or processes (e.g. how much Afrikaans the learner speaks at home is likely to be a rough index of socio-economic status, at least in this sample), and they should not be interpreted as routes to mathematics improvement.

Table 7
Correlations of predictors with criterion (improvement in mathematics performance)

Predictor	Correlation with criterion
Time on MasterMaths	0.37
Sex	0.16
Afrikaans speaking	0.22
Perception that effort at mathematics leads to success	0.21
Mathematics performance leads to jobs	0.1
Friends think it is important to do well in mathematics	0.14
Copy notes from the board often	0.2
Teacher checks homework	0.14
Learners use the board in mathematics lesson	0.19
Teacher explains rules and definitions	0.1
Look at textbook while teacher explains	0.17
Computer makes work more enjoyable	0.15
Computer makes schoolwork take more time	0.14
Computer has an effect on schoolwork	0.21

All correlations reported in this table are significant at $p < 0.05$.

3.3.3. Predictors of improved performance: multiple regression analysis

As we indicated immediately above, the correlations of individual variables with the mathematics improvement score should be read with a caveat in mind, namely that the variables are not independent. One way to better reconcile the list of predictor variables is to conduct a multiple regression analysis. (We considered a more sophisticated method of analysis, hierarchical linear modelling, which appears to be better suited to our multi-level design, but we rejected this. Several weaknesses in our design militated against this choice, e.g. we had too few clusters at some levels [only three experimental schools with a full variable set] to make this a statistically powerful option.) This analysis has the advantage of partialling the contributions of individual predictors for their correlations with other predictors in the analysis, and for pointing to variables that it may be useful to study in greater depth, or to include in more sophisticated modelling of mathematics improvement. Very important for our purposes is to understand whether the strength (and statistical significance) of a key intervention variable, viz. time spent on MM, is dependent on other predictor variables, and whether it will remain a significant predictor once we have controlled for the effect of other predictor variables. There are disadvantages to multiple regression analysis too, though, and perhaps the most important of these in the present case is that it is used as an empirical replacement for program model development. Precisely which variables end up in the final model we report here has much to do with random sampling variation, and analysis of a different sample may well identify different model variables.

We chose a forward stepwise method as a selection method for the multiple regression analysis. We chose also to separately analyse three subsets of our data, namely (i) the set comprising all five control schools, (ii) the set containing all five experimental schools, and (iii) the set containing only the experimental schools for which we had data regarding the use of MM. The latter set was much smaller than either of the former two, but was important to analyse, since it contained the clearest indicator of intervention strength. It was also important to separate the control from the experimental schools, as we wanted to examine the comparative importance of variables relating to the use of computers in learning mathematics.

Table 8 shows the model selected by the stepwise procedure in the case of the control schools. Teaching and motivational variables dominate in the list, and have higher values of β on average than other variables (roughly indicating greater importance). Several demographic variables are also in the model, but not much should be read into these variables, we suggest, as they are likely to be sampling artefacts.

Table 9 shows the model in the case of the experimental schools. The stepwise algorithm identified teaching variables and specific positive perceptions about using computers to do mathematics and schoolwork as statistically significant predictors. However, it should be noted that the amount of variance explained by the full model is comparatively small. Also, none of the sampling variables entered the model, as they did in the model for the control schools.

Table 8
Predictors selected by a stepwise algorithm for the model of mathematics improvement: control schools

Variable	β
Effort leads to success at mathematics[a]	0.24
Learners use the board	0.23
We look at textbook while teacher explains	0.21
How often do you speak Afrikaans at home?	0.21
Teacher checks homework	0.18
How many people live in your home?	−0.14
Sex	0.12

All β coefficients are statistically significant at $p<0.05$. Some coefficients have been reversed according to scale direction. Full model $r^2=0.37$; $F=14.1$; d.f. = 8,194; $p<0.0001$.

[a]This variable is a composite of a number of scale items. It is intended to assess the extent to which learners believe that success in mathematics derives from effort and factors within the control of the respondent.

Table 9
Predictors selected by a stepwise algorithm for the model of mathematics improvement: experimental schools

Variable	β
We often have a quiz or test in class	0.29
I like using computers to learn mathematics	0.19
We often check each other's homework	0.16
The computer makes schoolwork easier	0.17

All β coefficients are statistically significant at $p<0.05$. Some coefficients have been reversed according to scale direction. Full model $r^2=0.14$; $F=5.5215$; d.f. = 4,166; $p<0.0001$.

Table 10
Predictors selected by a stepwise algorithm for the model of mathematics improvement: experimental schools with data on amount of intervention

Variable	β
Total time spent on MasterMaths	0.39
My mother thinks it is important for me to do well in mathematics	0.27
The computer makes schoolwork easier	0.26
We often have a quiz or test in class	0.21

All β coefficients are statistically significant at $p<0.05$. Some coefficients have been reversed according to scale direction. Full model $r^2=0.33$; $F=9.58$; d.f. = 4,77; $p<0.0001$.

Table 10 shows the model in the case of the experimental schools for whom intervention data (time spent on MM) were available. Four variables were selected for inclusion by the stepwise algorithm, namely total time spent on MM, the perceived importance attached by a learner's mother to the learner's performance in mathematics, one variable related to teaching practices, and one variable related to perception of the role of computers in schoolwork. The MM variable was highly weighted in the model (just as it was in all the earlier analyses involving it), and the variance resolved by the model was considerably higher than that for the experimental group which did not include the MM variable. In order to check further the salience of the MM variable, a regression model consisting of all the variables in Table 10 except the MM variable was computed, and compared to the full model. The reduced model accounted for only 15.5% of the variance, compared to the 33% of the full model, and this difference was statistically significant ($F=18.8$; d.f. = 1,76; $p<0.001$). In other words, the intervention variable contributed substantial and significant variance to the model, over and above that contributed by other variables.

We may sum up these findings as follows: a number of variables predict improvement in mathematics performance. Some of these relate to teaching practices in the classroom, some relate to social differences between learners, and some relate to the specific intervention implemented by the Khanya project. The strongest of these predictors in the present sample is the strength of the intervention, and its predictive capacity does not appear to be reducible to any other variable in our analysis.

4. Conclusion

The Khanya project aims ultimately at improving the mathematics performance of grade 12 learners through the delivery of a computer based intervention. A key outcome is thus whether the final results achieved in the WCED grade 12 mathematics examination show an improvement as a function of the intervention. In this paper, we report analyses that suggest that the answer to this question is a cautious "yes". In the first place, data from a quasi-experiment provided equivocal support. In two matched comparisons, schools that had received the intervention maintained their grade 11 level of performance in the final grade 12 examination, whereas control schools that did not receive the intervention declined. However, this was not the general pattern of results in the quasi-experiment, as two experimental schools showed no improvement over the matched controls, and one showed an advantage for the control school over the experimental school. We must conclude that there is only equivocal support for the effectiveness of the intervention on the basis of the quasi-experiment.

In the second place, we examined the relation between the "strength" of the intervention and improvement in grade 12 mathematics performance. This was achieved by collecting log files from computers in three of the experimental school laboratories, and correlating student records of MM usage in those log files with improvement in mathematics across grades 11 and 12. The first thing we noted was that learners are in fact spending very little time on MM in these schools, averaging less than 3 h over the period April 2003–September 2003. However, the amount of time that learners spent was significantly correlated with improvement in mathematics performance. This was true for each of the schools we examined, and for their combination. Since learners were in fact spending very little time on MM, the data showed considerable restriction of range, and it is very likely that the correlation we observed is an under-estimate. This is a clear, but not conclusive indication that the Khanya intervention improves mathematics performance in grade 12 learners.

5. Lessons learned

In evaluation terms, a number of observations can be made. The major one probably relates to the request from the program management for an outcome evaluation, virtually within 2 years of launching the program. Our earlier reports on implementation confirmed our initial reservation: the program was still "settling down", as it was still developing and perfecting its delivery system, amongst other things. It was very early in the development of the program to expect it to deliver unequivocal outcomes. Nevertheless, in our pre-evaluation discussions with management officials, we became convinced that this request was not unreasonable, and thus agreed to proceed with an evaluation design that took this into consideration. From the program's side, however, they had to acknowledge the limitations of what such an early outcome evaluation could achieve, and indeed the risks involved. The discussion of risk also formed a major part of our initial conversations with other stakeholders, as part of our evaluability assessment. In these conversations, we made sure that significant stakeholders, such as the Finance Department, were prepared to accept the formative nature of the evaluation; that it was aimed at improving the program rather than delivering a definitive statement of its effectiveness.

We believe that this strategy was successful. As we assessed the technical installation at the schools, and the implementation, these findings were fed back to the program, which, in all cases that we are aware of, acted on them. More importantly, the final report was framed in a formative tenor, and was accepted as such by all stakeholders. It enabled the program managers to act on problems of implementation, even though these emerged as part of an outcome evaluation. The major finding that more time spent on the program led to improvement in mathematics performance, obviously was very encouraging to the program and the Western Cape Education Department. It is worth noting too that this finding supports the internationally established finding in empirical curricula studies that "time-on-task" is a central component of the construct "opportunity to learn" (Burstein, 1993; Shavelson, McDonnell, & Oakes, 1989).

Since the evaluation (2004), the Khanya project has gone from strength to strength. In 2004, it won the Standard Bank Public Sector Innovation Award, and in 2005, a Silver Award at the Western Cape Premier's Service Excellence Awards. Quite recently (October 2006), it was nominated in two categories of the 2006 African ICT Achievers Awards (Top Public Sector Department Embracing ICT in Africa, and Top Civil Society to Bridge the Digital Divide in Africa).

The challenges we faced in implementing the evaluation design should be obvious to the readers. What started out as a quasi-experimental design, with schools as the unit of analysis, ended up as a different quasi-experimental design with individuals as the unit of analysis. Difficulties in sustaining an evaluation design are not that unusual, but technology provided opportunities to address these challenges. The fact that computers captured user logs (albeit in a less-than-perfect way) turned out to be the key feature in our evaluation, and in providing useful results in terms of outcome. Indeed, this was one of our strongest recommendations to program management: to ensure that these log files are maintained and captured regularly. These data sources provided distinctive indicators of individual levels of "dosage", and they served the evaluation well.

Thus, we conclude this paper in the same way that we concluded our report on the program. We caution against an interpretation of these results that regard them as in any way delivering a final judgement on its critical aspects. Although it is an evaluation of outcomes, this phase of the study, like the evaluability, technology and implementation assessments, should have a formative function. That is, it is intended to assist the program to perform better. This is an evaluation for program improvement, aimed at providing early feedback to Khanya on how it is doing in terms of achieving (some of) the objectives it set for itself.

References

BECTA. (2000). *ImpaCT2: Emerging findings from the evaluation of the impact of information and communications technologies on pupil attainment*. National Grid for Learning (NGfL) and BECTA. Retrieved October 2002 from: 〈www.becta.org.uk/research/reports/impact2/index.html〉.

Burstein, L. (Ed.). (1993). *The IEA study of mathematics III: Student growth and classroom processes*. Tarrytown, NY: Pergamon Press.

Chang, K. E., Sung, Y. T., & Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, *17*, 21–33.

Clynick, T., & Lee, R. (2004). *From Laggard to World Class: Reforming mathematics and science education in South Africa*. Johannesburg: The Centre for Development and Enterprise, Research Report No. 13.

Dwyer, D. (1994). Apple classrooms of tomorrow. What we've learned. *Educational Leadership*, *51*, 4–10.

Howell, C., & Lundall, P. (2000). *Computers in schools: A national survey of information communication technology in South African schools*. Cape Town: Education Policy Unit, University of Western Cape. Retrieved October 2002 from: 〈http://www.education.pwv.gov.za/teli2/research/ICT%20audit/Prelims.pdf〉.

Interactive Educational Systems Design Inc. (IESD). (1996). *Report on the effectiveness of technology in schools* '95–96. Commissioned by the Software Publishers Association. Retrieved October 2002 from: 〈http://www.spa.org/project/edu_pub/summary.htm〉.

Jurich, S. (1999). Computers in the classroom: How effective? *Technologia: International Journal of Technologies for the Advancement of Knowledge and Learning*, *1*(2), 31–35.

Kane-Berman, J. (13 April 2006). Health, education struggle amid booming economy. *Business Day*.

Shavelson, R., McDonnell, L., & Oakes, J. (Eds.). (1989). *Indicators for monitoring mathematics and science education*. Santa Monica, CA: RAND.

SIIA. (2000). *Research report on the effectiveness of technology in schools*. Software Industry Information Association (SIIA). Retrieved October 2002 from: 〈http://www.siia.net/sharedcontent/store/e-edtech-sum00.pdf〉.

Soloway, E., Lockhead, J., & Clement, J. (1982). Does computer programming enhance problem solving ability? Some positive evidence on algebra word problems. In R. J. Seidel, R. E. Anderson, & B. Hunter (Eds.), *Computer literacy*. New York: Academic Press.

Taylor, N., Muller, J., & Vinjevold, P. (2003). *Getting schools working: Research and systemic school reform in South Africa*. Cape Town: Pearson Education South Africa.

Veerman, A. L., Andriessen, J. E. B., & Kanselaar, G. (2000). Collaborative learning through computer mediated argumentation. In C. Hoadley, & J. Roschelle (Eds.), *Proceedings of the computer support for collaborative learning (CSCL) 1999 conference* (pp. 682–686). Stanford University: Lawrence Erlbaum Associates.

Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics*. Educational Testing Service. Retrieved October 2002 from: 〈http://www.ets.org/research/allreports.html〉.

Wholey, J. S. (2004). Evaluability assessment. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed., pp. 33–62). San Francisco: Jossey-Bass.