# The Effect of an Individualized Online Practice Tool on Math Performance - Evidence from a Randomized Field Experiment[1]

**Carla Haelermans[ξψ2] and Joris Ghysels[ξ]**

ξ Top Institute for Evidence-Based Education Research (TIER), Maastricht University, PO Box 600, 6200 MD, the Netherlands

ψ Centre for Innovations and Public Sector Efficiency Studies (IPSE Studies), Delft University of Technology, the Netherlands

## Abstract

This paper explores the effect of using an individualized interactive online practice tool on basic math skills of 7th grade students with a randomized field experiment. The results show that practicing with the online tool leads to a substantial and significant increase in math performance growth. On top of that, a positive and significant relation between additional minutes practiced per week and math performance is revealed. The effect is robust to adding student characteristics that influence their practice behavior and to adding usage and attitude towards the tool of the non-randomized teachers. So, the effect holds, despite the fact that there is large heterogeneity in teachers' usage and attitude towards the practice tool and despite the fact that there is large variation in practice behavior by students. Moreover, there are low implementation barriers, and a cost-benefit analysis indicates that the potential cost savings of this method are very large.

**JEL-Classification** – I20, I21, C93.

**Key words –** Randomized Field Experiment; IT in Education; Online Practice Tool; Math Skills; Secondary Education.

---

[1] The authors would like to thank Jan-Hein de Wit and Dorien Stals from Dendron College for making this field experiment possible. Additionally, they express gratitude for the contributions of Freek Weeda and Theo Schijf. This paper has greatly benefited from the comments and feedback from Trudie Schils and Anders Sternberg.

[2] Corresponding author: Carla.Haelermans@maastrichtuniversity.nl, PO Box 616, 6200 MD Maastricht, the Netherlands. T: +31 43 388 4458

## 1. Introduction

Basic math and language skills are of major importance in daily life. However, many students do not have sufficient basic math and language skills (Funnekotter, 2012; KNAW, 2009). In the Netherlands, both a special parliamentary commission (Commissie Meijerink, 2008) and the Royal Netherlands Academy of Arts and Sciences (KNAW, 2009) concluded that urgent action was required, because an increasing number of students lack the necessary math and language skills. This has led to the introduction of compulsory math and language exams in the already existing national graduation exam program for secondary education.

Many scholars conclude that individual differentiation is the key to higher student performance (e.g. Hattie, 2009), but traditional classroom settings only partly allow schools to differentiate their teaching between individual students. This is often seen as the reason for the lack of skills. The combination of the increase in computer use in education, the need for individualization in the learning process and the decrease in math and language skills has led to the development of individualized IT-tools aimed at developing these skills. Accordingly, many schools started using individualized IT-tools to increase students' math and language skills. Individualized IT-tools focus on an individual learning path for the student, adapting the exercises available for the student to the skills that he or she is lacking.

However, the existing literature contains only few experimental studies on the effect of IT-tools on math and literacy performance (Arroyo, Park Woolf, Royer, Tai, & English, 2010; Borman, Benson, & Overman, 2008; Pilli & Aksu, 2013; Rouse & Krueger, 2004) and, hence, it is unclear whether these schools chose an effective teaching program.

Therefore, the purpose of this paper is to analyze the effect of individualized educational software developed in the context of the above described policy change in the Netherlands. We conduct a randomized experiment with an interactive online practice tool and analyze the effects on math performance[3] of students in 7$^{th}$ grade (age 12, first year of secondary school). In doing so we were able to also take into account the intensity of treatment and the influence of the teacher. We show that the effect of the online practice tool is about 0.28 of a standard deviation. Furthermore, we show that the non-compulsory training complements class-based training and is effective regardless of the math class and the teacher attitude.

---

[3] Language performance was also part of the experiment, but this will be analyzed in a companion paper. The language experiment shows no significant results on performance or performance growth, and the potential reasons for the lack of significant results will be explored in the companion paper.

Many of the previous evaluations of information technology (IT) are wide in scope, as they evaluate, for example, increased budgets for IT either for schools or for households. They rely on the assumption that users have sufficient skills to implement and use IT to their benefit and that it does not matter how IT is used, in order to benefit educational outcomes. Yet, in practice, these general evaluations offer mixed results (e.g. no significant effect of IT: Goolsbee and Guryan (2006); positive effect of IT: Machin et al. (2007); negative effect of IT: Angrist and Lavy (2002), Leuven et al. (2007)). A second part of the literature on IT focuses on the comparison of computer directed versus traditional classroom teaching. A couple of meta-analyses with strict selection criteria with respect to methodology used in the individual studies (Cheung & Slavin, 2012, 2013; Kulik & Kulik, 1991), show that in general, computer directed instruction does have small positive effects on student performance, compared with traditional classroom teaching, for both math and language. Two older studies by Becker (1990) and MacGregor et al. (1988) both show no effects of computer directed education.

IT is particularly suited to provide individualized differentiation (from now on: individualization), as algorithms offer the opportunity to develop individual learning paths. Incorporating the differences in level, interests and learning styles between students proves to improve students' motivation (Tomlinson, 2004), and neglecting these differences might lead to decreased performance of certain students (Tomlinson & Kalbfleisch, 1998).

Evaluations of IT-based individualization programs in math range from general teaching to remedying programs and cover both general student audiences and students with learning disabilities. In general, evaluation outcomes tend to be positive. Burns et al. (2012) show that significantly fewer of the students at risk for math difficulties beforehand where still at risk after using a computer delivered math fact intervention. Similar results are found by Pilli and Aksu (2013). Banerjee et al. (2007) report on the positive outcomes of an experiment with an IT-based math remedying program, introduced in public schools of two cities in India, which illustrates that the benefits of IT-individualization are not confined to students from highly technologized societies.

The before mentioned three studies all analyzed 3$^{rd}$ and/or 4$^{th}$ grade students. There is only one academic publication using a similar age group as in the study at hand. Arroyo et al. (2010) analyzed 250 7$^{th}$ and 8$^{th}$ grade students that used a digital skill drill method, or traditional practicing on paper, 15 minutes per day next to math classes, for four days, and find a significant positive effect of digital practicing. A report by the US Department of Education shows results on a similar age group in the recent experimental evaluation of various mathematics software packages, mainly used as substitute for math classes, among 6th grade students in the US. The report came to the conclusion that none of the tools reviewed was effective (Dynarski et al., 2007). However, the large variation in the way schools implemented the software packages and the variation in intensity of treatment make it difficult to compare this study with previous literature and compare results.

Although it is possible that publication bias distorts the conclusions on specific programs more than in the case of the general IT evaluation, the former group of evaluations offers a range of positive experiences to build on. However, a potential hindrance to the rapid expansion of educational innovation through IT is acceptance by teachers. On the one hand, teachers often do not want interference in their classroom, and especially elder teachers often do not believe in the benefits of IT training. On the other hand, interventions and innovations are often imposed by the management, without consulting the teachers, which also might lead to resistance by teachers. As we study the introduction of software that does not require a large teacher investment, this potential hindrance is not reflected in this study. The "Mousework[4]" program we put to the test is an online, skill-drill program that students are free to use at home and that complements math classes without being tightly linked to the pace of teaching. The experimental setup allows for studying the effect of the program both with and without taking into account the teacher. This is of particular importance, since students' performance is largely influenced by the teacher (Hattie, 2009).

In analyzing this tool, this paper contributes to the literature in four ways. First, the randomized experimental design allows for a causal analysis of practicing with an interactive online tool on basic math skills of students in 7th grade. We can control for extremely rich information on these students, their teachers and the context of the experiment. Second, our research contributes to the discussion of individualized differentiation in education. Third, we show effects of non-compulsory training that complements class-based training. This implies that the training can take place, and that students can gain in performance, regardless of the math class and the teacher attitude. Fourth, another advantage of this study is the availability of data on the intensity of the treatment; i.e. we do not just have a dummy variable to measure who had access to the tool (although that is the main focus of this paper) but we also know if, when, what, how often and how long students used the practice tool. This allows for very detailed analysis of students' practice behavior.

[4] The Dutch name is "Muiswerk".

The remainder of this paper is structured as follows: Section 2 presents the context of the experiment, e.g. the purpose, contents and organization of the online practice tool, the identification strategy, the use of the online practice tool, and measuring the math skills of secondary students. Section 3 presents the empirical model, the baseline results and the regression results. In Section 4 we discuss additional evidence on the effect of the online practice tool, such as the analysis of factors influencing students' practicing behavior, an analysis of minutes practiced as continuous explanatory variable instead of the treatment dummy and a cost benefit analysis. Section 5 concludes the paper and discusses the findings.

## 2. Context of the Experiment

### *2.1 Dutch Educational Context*

The setting of the experiment is Dutch secondary education, in which a recent law change took place in 2010. The government announced a change in the national graduation exam requirements by introducing compulsory exams in math and language skills from 2015/2016 on, next to the already existing national exams, and by setting references levels, levels which students are supposed to have achieved at the end of each year. Currently, students write national exams at the end of secondary education in the subjects that they followed until graduation year. Regardless of which subjects the students followed, all students will have to write the exams on math and language, from 2015/16 on. Failure to pass these math and language exams means that the student will not be able to graduate. Therefore, most schools are developing a policy to make sure that all students will have the required level for language and math when they write their national exams in the last year of secondary education. The school under study chose to use an interactive online practice tool to improve the language and math level of its students, of which we will only study the effect on math skills in this experiment. Considering the large discussion on using IT tools for practicing purposes in education, the context and results of the experiment are not only relevant for Dutch schools, but for all schools, regardless in which Western country they are located.

### *2.2 The school under study*

The school under study, Dendron College, has about 2000 students in total and is - to Dutch standards - a mid-sized school for secondary education (junior high and high school).

Dendron College offers secondary education in all tracks[5] and is tracking students from the first year on in several prevocational, general and pre-university tracks. In school year 2012/2013, there were 430 students in 17 first year classes (equivalent to seventh grade in the US). Of these classes, 13 were part of the experiment (355 students), ranging from the more theoretical prevocational track to the pre-university track. Only the four most extremes of the first year classes, which were the three classes from the basic prevocational track and the one pre-university highly gifted class, were excluded from the experiment. The basic prevocational students were excluded for ethical reasons, as these students are lacking all basic math skills, and excluding part of these classes from practicing because of the experiment was unethical to both the researchers and the school. The one pre-university highly gifted class was excluded due to comparability issues, as this was the only class with highly gifted students and the counterfactual could therefore not be created.

The age of the students in the experiment ranges from 11 to 14 (differences are mainly due to grade repetition), about 6 percent of students are diagnosed with dyslexia by an external organization and about 56 percent are girls (see Table I). Furthermore, students come from 25 different primary schools and most students either state not to be religious or are Christian. Religion is used as a proxy for family background characteristics. As such the school is a typical representative of schools outside of the highly urbanized, central region of the Netherlands (the "Randstad").

[Table I around here]

### *2.3 Purpose, Contents and Organization of the Interactive Online Practice Tool*

The purpose of the interactive online practice tool is to help individuals practice their math skills, while being able to individualize, and give users direct feedback (Muiswerk, 2013). Although the program is mainly being used in the Netherlands, it also has an international version and is used by several international schools both in Europe and other parts of the world. The program is interactive and person specific. Students work at their own level and get those exercises that will help them improve the sub aspects of math they are not knowledgeable in yet, while some exercises are meant to keep up their already gathered

[5] Dutch secondary education has a tracking system from 7$^{th}$ grade on, with 3 different tracks (prevocational education, which consists of 4 sub tracks where level 1 is the lowest (mainly practical) track and level 4 the highest (mainly theoretical) track, general higher education and pre-university education).

knowledge. The school uses this tool to make sure each student achieves the highest possible level of math, given his/her abilities, and maintains the level achieved. It offers all students online access to the tool for use after school hours, at home. Currently, most math teachers at Dendron College are not using the tool in class, although the tool is also developed to that end.

First year secondary students make a math skills pretest at the start of the school year, in September. This test determines their level of different sub aspects of math, which in turn determines the types of exercises they have to start practicing with at home[6]. Regularly, students make a short computer test at school to determine for which exercises their skills are still lacking and for which exercises their knowledge level is good enough for the moment. These tests also determine the type and level of exercises a student can access in the online tool.

In the program, students can practice with all different topics of basic math and arithmetics. They practice by making exercises, and after a subset of exercises they get a score for this part. Then, depending on this score, the student might get new exercises or repetition of the same type of exercises. The program functions in a highly individualized manner, as it starts with explanation screens (online instruction), offers feed-back and it provides the student with either repetition or new learning modules on the basis of previous performance of the individual student. It works without teacher interventions, but teachers have access to a reporting module and some may incorporate knowledge of “Mousework” performance in their interaction with the students.

Math teachers are supposed to motivate students to practice with ”Mousework” at home and for checking students’ practicing behavior. However, not all math teachers are in agreement with the management to use this interactive online practice tool school-wide and some of these teachers refuse to act in accordance with the responsibility to check the students’ practice behavior. Therefore, we will control for the teachers’ attitude towards “Mousework” when studying the impact of online practicing. Moreover, we will look into explanations for practicing behavior in Section 4.

[6] A student questionnaire in spring 2013 shows that of the students in the experiment group only 5 students do not have a computer at home to practice with. However, IP address data shows that these students have practiced with the tool at school, where there are computers available for students that do not have one at home.

### *2.4 The field experiment*

*Identification strategy*

The main problem with determining the effect of a practicing tool is the potential correlation of unobservable factors with both the practicing behavior and the outcome variables, such as basic math performance. In this study we use exogenous variation in the possibility to practice, by fully excluding the control group from access to the online practice tool, and the time practiced per week through an experimental set-up.

Figure I shows the timeline of the field experiment, which consists of a pre-experiment period and the experiment period itself. In spring of their final year in primary education, students register at their school of choice for secondary education. The secondary school uses the results of the standardized national exit exam and the recommendation made by the primary school teacher to assign students to the first year classes before the summer break of 2012. At the school under study, the assignment to classes is done randomly within the boundary of the ability grouping that forms part of the Dutch system of secondary education ("early tracking") and the option for each student to select friends with whom to be placed in the same class. In summer, week 29/2012, classes were randomly assigned to treatment and control group. Only two types of first year classes (5 prevocational classes and 8 higher general/pre-university classes) took part in the experiment. Two classes of each type where assigned to the control group (113 students), whereas the other 9 classes are the treatment group (242 students). In week 33/2012 the school year started, and in the second week of the school year all students and their parents were informed about the experiment by means of a letter. In the fourth week of the school year, one of the researchers was present at the information evenings for parents to provide them with additional information regarding the experiment. Because the school provides children in the control group with extra lessons in math and language in the second half of the school year, all parents agreed to their child's participation in the experiment.

The pretest took place in week 37/2012. Table II shows that there is no significant difference in performance between the treatment and the control group at the pretest. The experiment lasted 8 weeks and the posttest took place in week 46.

[Figure I around here]

Table II presents the observable characteristics of the treatment and control group, for all students that wrote the pretest (N=337[7]), as well as the *t*-statistics/Mann-Whitney statistics on the differences between the groups. Apart from the randomization, these statistics indicate that we can trust (with a significance level of 5%) the treatment and control group to represent the same population.

[Table II around here]

### *2.5 The Use of the Interactive Online Practice Tool during the Experimental Period*

The school uses the interactive online practice tool in addition to their math classes. All students, both in the treatment and control classes, are being taught mathematics using the math method that has been used over the previous years. For first year secondary students (7th grade) there are 4 math classes of 50 minutes each week, and students make an average of between 45 and 60 minutes of homework per week[8]. Practicing the basic skills with the online tool is an additional activity that takes place outside the school. The small tests the students write regularly at school, determine which exercises they have to take and which skills are still lacking. The exercises available to the student in the online tool are based on the results of the pretest and these weekly tests. As these 13 classes are being taught by 7 different teachers (and the 9 treatment classes by 5 different teachers), there is large variation in practice behavior between classes in the treatment group, which we will show below to correlate with the teachers' attitude towards the tool (see Section 4.1).

Students have their personal account with login information to practice at home and are only allowed to make the small tests at school. Therefore, it is hardly possible for students from the control group to gain access to the online tool. Also, as students are allowed to select friends to be placed in the same class with, it is unlikely (especially during the first months of the first year of secondary education) that student have very good friends in other classes and will share their login code that gives access to the online practice tool. Lastly, most students complain about the tool being 'boring', which makes it even more unlikely that there will be contamination effects of students in the control group gaining access to the online practice tool.

[7] Note that 18 students did not write the pretest, for various reasons. The majority of these students also did not write the posttest.

[8] Information on homework gathered via a student questionnaire in which a multiple choice question was used to ask them how much time they spend on homework for math, *excluding* the time they spend on "Mousework".

Comparison of data on the attendance of tutoring classes of this cohort with earlier cohorts that did not work with this online math tool shows that the share of students who follow additional tutoring lessons for math, is similar over the years. Furthermore, the tool focuses on basic math and arithmetic skills, and these skills are not (yet) official part of tutoring classes, as these classes mainly focuses on the official subjects taught in secondary school. This provides information on the use of the tool namely that the tool is not replacing tutoring classes, but is used next to these classes.

### *2.6 Measuring Mathematical Skills and Practice Behavior*

The mathematical skills are measured using a digital standardized math test[9], which is written by all first years students at T0 and T1 (see Figure I). The standardized tests are tests that measure whether students have mastered the required national math level they are supposed to have, given their age and given the fact that they finished primary school. The tests are not directly related to the exercises in the practice tool, and although the tool is designed to improve students' math level that is examined by this test, students are also supposed to gain this knowledge in regular math classes (however, math classes are of course not as differentiated towards the specific level of that individual student). All students, both treatment and control group, practiced with digital multiple choice math assignments in the testing program in the week before the pretest was administered, to make sure they knew what to expect when writing the pretest. The test contains multiple choice questions and students are allowed to use scrap paper for their calculations, but no digital calculator. Both tests were identical and lasted for about 20 minutes, although most students did not need all this time. Test scores can range from 0 to 100, 100 being the absolute maximum. For the analysis, we calculate both the absolute and the percentage growth in score between T0 and T1 which are both normally distributed and range from -31 to 35 and from -0.63 to 1.23, respectively. We study two types of performance measures: first, the growth in absolute score and the growth in absolute score compared with the starting score, and second, the growth in score per second, meaning the number of questions answered correct per time unit. This is easily calculated as we know how much time each student spent writing the tests and measures the automation of math skills. The absolute and percentage growth in score per second are also

[9] Ideally, we would also like to see if there is an effect on regular math performance, next to the potential effect on performance on this specific arithmetic math test. However, regular math tests are developed by the own teacher, and are not comparable across classes, not even within the same year and the same track. Therefore, we cannot make a valid comparison of the regular math performance of treatment and control group.

normally distributed and range from -0.03 to 0.11 and from -0.40 to 3.71. Growth in score shows the average improvement of students whereas growth in score per second is a combined measure of speed and accuracy. In the extended analysis from Section 3.3 on we only present results for this latter measure, as this gives more specific information on both goals of the online tool, namely improved performance and improved automation of certain skills (i.e. increased speed).

In addition to studying the effect of access to the online practice tool, it is also descriptively interesting to study the intensity of the treatment, by including the effect of the number of minutes practiced per week on growth in math score. Below, we will look into the number of minutes practiced per week, which is determined by subtracting the weeks of school holidays from the total number of regular school weeks during the experimental period (i.e. excluding test weeks). Students could still practice at home during school holidays, but the data shows that this was hardly the case. This leaves us with 7 weeks between T0 and T1 (see Figure I).

### *2.7 Measuring teacher attitude towards the tool and how use of the tool*

In the beginning of the experiment a short questionnaire was handed out among math teachers, to gather information on their attitude towards the online tool and on the way they use the online tool (if they use it at all), for example in class or to check up on students. The questionnaire consisted of 21 multiple choice questions, among which 18 statements which are measured on a 5-point Likert scale. Using factor analysis, we found that of these 18 statements, 6 considered the attitude towards the online tool, and we use the average of these 6 statements to measure "math teacher attitude" (this combined measure has a Cronbach's alpha of 0.78, where an alpha of 0.7 or more is acceptable (Field, 2013). Another 7 statements all considered the use of the tool by the teacher, and these were combined (weighted average) into the new variable "math teacher use of tool" (this combined measure has a Cronbach's alpha of 0.84)[10].

## 3. Empirical Analysis and Results

### *3.1 Methodology*

To identify the Average Treatment Effect (ATE) of access to the online practice tool on growth in test scores we use the notation first used by Rosenbaum and Rubin (1983). We

[10] Questionnaire and data files are available upon request from the corresponding author.

observe a student $i$ in class $j$'s percentage growth in test score $y_{ij}$ and the treatment, a students' access to the interactive online practice tool, determined at the class level, $d_{ij}$, which results in the following equation:

$$y_{ij} = d_{ij}y_{ij}(1) + (1 - d_{ij})y_{ji}(0), \quad (1)$$

Where $y_{ij}(1)$ is the percentage growth in test score and the percentage growth in test score per second for treated students in treatment classes and $y_{ij}(0)$ is the percentage growth in test score and the percentage growth in test score per second for untreated students in control classes. Since the randomization ensures the independence between the treatment and potential outcomes, we identify the ATE as follows:

$$\tau_1 = E[y_{ij}(1) - y_{ij}(0)]. \quad (2)$$

We can estimate the ATE using either simple $t$-statistics or using a linear regression. The linear regression is estimated as follows:

$$y_{ij} = \alpha_i + \tau_2 d_j + \beta_i X_i + (\varepsilon_i + u_j), \quad (3)$$

Where $d_j$is the treatment status of class $j$, $X_i$ are the students' observable characteristics, such as ability variables, and student characteristics, which are independent of the treatment, $\varepsilon_i$ are the residuals at the student level and $u_j$ are the residuals at the class level. Because of the randomization at class level, we use a class random effects model throughout the analyses presented in this paper.

### *3.2Baseline Results*

The first results we present are the simple $t$-statistics of the effect of treatment on the growth in scores. Table III presents both the absolute and the percentage growth in test scores and in test score per second for treated and untreated students for the posttest. In Table III, we see that the treatment group has a significantly higher growth, both absolute and in percentage, than the control group for both outcome variables. This implies that practicing at home with the online tool is beneficial, both for growth in test score in itself and for growth in score per second. Not only have the skills improved significantly, but they have also become more automated. The standardized version of the effect on percentage growth test score per second points towards a medium effect of 0.38 of a standard deviation[11], given the interpretation of Cohen's *d* (Cohen, 1988). However, educational interventions typically give effects of around 0.2, indicating that this standardized effect of 0.38 is substantially large, given the sector of research.

Additional analysis on the Dutch reference levels shows that in T0 only about 6 percent of the students achieved the reference level they were supposed to have achieved after finishing primary education (see section 2.1). During the experiment, this increased by about 7 percentage point for the reference group, whereas the share of students who have achieved this reference level for the control group remained constant. This difference is significant at the 5% level.

[Table III around here]

### *3.3 The Returns to Practicing Online*

The next step is to analyze the returns to practicing online (regardless of how often and how much the treatment student practices) using regression analysis. The results of these analyses of the returns to practicing online are presented in Table IV. In Table IV, we present 5 models: Model 1 gives the basic model in which no covariates are included, estimated by simple OLS. Model 2 is the same model as Model 1, however, we do now control for the clustering of students in classes by using a class random effects model and by clustering the standard errors at the class level. It is our expectation that the models will not produce qualitatively different results, because the randomization procedure should produce comparable research samples and, hence, class clustering should not alter the effect of the intervention. An –unexpected- difference may signal a lack of comparability due to randomization at the class level. In Model 3, we include the test score for the text comprehension test of T0, as the math questions can be very linguistic, and students who have problems with text comprehension might also score lower for math. In Model 4, we add variables that account for student ability and past education, such as scores for the math part of the standardized exit exam of primary education, as well as the total score of that test, the advice given by primary school, and an indicator for dyslexia, and we add student specific

[11] Note that the standardized effect sizes of the other outcomes range between 0.25 (Absolute growth test score per second) to 0.43 (absolute growth in test score).

characteristics such as age, gender, oldest child, religion, family type and primary school. In the last model we add the attitude of the math teacher towards the online practice tool and the extent to which the math teacher uses the tool, for example for checking up on students. As such, model 5 combines measures of student ability with indicators of past education (e.g. primary school) and current education (e.g. class group, math teacher).

The results presented in Table IV show that the percentage growth in score per second is around 15 percentage points higher for treatment students, compared with control students, when we only control for class. Note that this number is similar to the simple OLS regression in which we not include class random effects, which implies that the class structure does not influence the results that much. The percentage growth in score per second decreases to 12 percentage points in the fifth model, where we control for all the covariates. The latter corresponds to a medium small effect of 0.28 of a standard deviation. Hence, the significance of the effect proves robust to adding different types of student specific information. In other words, even when taking into account various student and teacher characteristics which may contribute to math learning, the intervention is shown to add to math performance. Regarding identification, the maintained effect suggests that randomization worked and produced a fairly independent effect. Detailed results of the regressions presented in Table IV can be found in Appendix 1[12].

[Table IV around here]

## 4. Additional evidence

### *4.1 Factors influencing students' practicing behavior*

A closer inspection of the treated students shows that there are large differences in practice behavior between these students, ranging from an average of 0 minutes per week to 35 minutes per week[13]. To study the reasons why practice behavior differs so much, we analyze practice behavior in a class random effect regression. Table V shows the results of the analysis of the relation between student and teacher variables and the average number of

[12] As practice behavior differed largely among students, we checked whether the effect was driven by the group of students that had practiced the most during this period (on average more than 10 minutes per week), vs. the group that practiced the least (less than 5 minutes per week) and the group in between. The results of the analysis of these separate groups show that this is not the case, independently of the group with respect to minutes practiced, experiment group students perform significantly better than control group students.

[13] Note that practicing minutes only count when at least one exercise is finished.

minutes practiced per week. Similar to Table IV, we control for class grouping by using a class random effects model. Table V shows that the only student characteristic that significantly influences practice behavior is situation at home. Students that have a more difficult situation at home, i.e. have divorced or deceased parents, practice significantly less. The latter could also be caused by the fact that these students usually live at multiple places and the set maximum of 5 IP addresses per account, which can be used to log in to, constrained their online practice environment.

Table V also shows the positive influence of teacher attitude towards to tool and teacher use of the tool on student practice behavior. If teachers use the tool in their classes and check up on students, student practice significantly more minutes per week than if teachers do not use the tool.

[Table V around here]

### *4.2 The Returns to Practice Time*

Apart from the analysis of the returns to practice in general, it is also interesting to analyze the intensity of the practice behavior and its relation with the percentage growth in score per second. Table VI shows models 2 to 5 as were presented in Table IV, except that the treatment dummy has been replaced by a continuous variable which measures the average number of minutes practiced per week. Note that control group students obviously have 0 minutes practiced per week.

Table VI shows that the effect of practicing an additional minute per week is only significant in the two last models, in which we include student and teacher characteristics. This result is in line with the results presented in Table V, as there are significant differences in practice behavior for certain types of students and different teacher characteristics. When focusing on the result in the most extensive model, we see that additional practice of one second increases percentage growth in score per second by 0.009, which means that half an hour per week of additional practice leads to a significant increase of 16 percentage points in percentage growth of score per second. Detailed results of the regressions presented in Table VI can be found in Appendix 2.

Although we show a positive relationship between additional practice and the growth in math score (per second) it is of course very likely that this relationship is not linear. However, our data does not show a non-linear effect when including quadratic terms. Future research is needed to confirm if this holds if larger datasets are used and if students have a larger average in minutes practice per week.

[Table VI around here]

### *4.3 Cost effectiveness*

Determining the cost effectiveness of this online practice tool can be done from the schools point of view, but also from the society's point of view. The costs of the online practice tool for math are approximately 25 euros per student per year. The total costs for this tool for the group of 430 first year students for the school is around 10 000 euros. For the school, an alternative measure to foster math skills, is the introduction of an additional math class. The additional costs of hiring a teacher who practices basic math skills with the student for at least an hour a week would be a lot higher than using this online tool. Given that there are 18 first year classes, one would need an additional full time teacher each year, which will bring about costs of at least 40 000 euros per year.

With respect to gains for society, we see that in the test of September 2012, about 85 percent of the first year students in our dataset performed lower on math than they are supposed to, according to the national reference levels (Commissie Meijerink, 2008), having finished primary education successfully. If practicing with the online tool would increase the average math level such that the majority of these students would perform at the expected level by graduation, the societal cost saving could be very large. Additional analysis on the reference levels shows that in the treatment group the share of students that is performing at the reference level has significantly increased by 7 percentage points, compared with the control group which remained constant. Given that each student costs about 7 000 euros per year (Teule, 2012), each student that does not fail the national exam at the end of secondary education because they fail their math exam, saves the government 7 000 euros. Furthermore, retention in grade in graduation year, due to failing the math exam, will delay the student for at least one year in entering either vocational education or higher education, which in turn delays labor market entry by at least one year. The opportunity costs of the student will therefore be a lot higher than the 7 000 euros for the government. In any case, to be cost effective from society's point of view, the introduction of "Mousework" in the starting year of secondary education only needs to allow two students to graduate in time instead of delaying their graduation with one year. Given that it can be expected to help 30 students across the threshold (7% of 430), the latter seems highly likely, although future research following-up on students throughout secondary education should confirm that expectation before any solid statements of this type can be made.

## 5. Conclusions and discussion

### *5.1 Conclusions*

In this paper, we study the average treatment effect of offering an individualized interactive online practice tool on basic math skills of 7th grade students. A randomized experiment was carried out during 8 weeks among 337 students. Treatment group students were given the opportunity to practice with the tool at home, next to their math classes. Control group students were not granted access to the online tool and only attended the regular math classes. The results show that there is a significant effect of practicing with the online tool on both (percentage) growth in score and (percentage) growth in score per second (accuracy and speed). The effect on percentage growth in score per second is 10 percentage points, which corresponds to a medium small effect of 0.28 of a standard deviation. This effect is robust to adding different types of student specific information that influences practice behavior and information on the use of the tool and attitude towards the tool of the teacher. A cost benefit analysis shows large potential gains, since this tool might prevent many students from failing their national graduation exams and consequently having to repeat the last grade. Retention in grade is costly in both direct and indirect ways. The direct costs lie in the costs per student per year for the government, approximately 7 000 euros per year. The indirect costs are the costs of a delayed entrance in the labor market, which is relevant for both the individual and for society as a whole.

### *5.2 Discussion*

Our contribution to the literature is the randomized experimental design, combined with a relatively large number of students and rich information on these students, their teachers and the context of the experiment. The research design allows for causal analysis and for estimating the average treatment effect (ATE) of practicing with an interactive online tool on basic math skills of first year secondary students. Furthermore, this paper contributes to the literature on individualization, the use of skills drill methods and the use of IT in (secondary) education. Lastly, a strong point of this study is that it shows that non-compulsory training has an effect, regardless of teacher attitude and that we have detailed information on how students use this non-compulsory training.

Although the experiment was conducted at only one school, test results of math tests of two other schools (where no experiment was conducted, but which are similar to our experiment school) that use the same online practice tool show a similar increasing trend in results after students have practiced with the tool. This may indicate that the results can be generalized to other secondary schools in the Netherlands, with a similar student population, that use this tool.

Data on the remainder of the school year for our sample, which shows that the average amount of minutes practiced per week has increased after the experimental period, indicates that the effect is not driven by student motivation caused by the fact that they were part of an experiment. Furthermore, the student questionnaire shows that students do not like to practice with the tool, and informal conversations with students reveals that control group students were actually *happy* that they were selected as control group. These two observations indicate that we are measuring the effect of online practice.

The innovation evaluated in this paper offers a dual conclusion regarding the role of teachers. First, using an online practice tool that is independent of teaching in class gives positive outcomes, even if the teachers have mixed feelings about its use. Second, the analyses also illustrate, however, that teachers with a positive attitude towards the tool tend to contribute to the success of the tool itself, by fostering the intensity of its use.

Future research should shed further light on the distribution of the efforts and gains between various types of students and show to what extent the individualization the tool offers is a crucial element for its success. Moreover, it may clarify how exactly teachers can stimulate students to increase their practice time, so as to deepen the effect of the tool. Lastly, future research should focus on the potential motivational role that parents can play if they would have access to their child's practice behavior.

All-in-all, the results of this experiment are in line with the findings of most experimental studies discussed in the introduction. We show that an online skill-drill tool which is used by students at home without a close link with the math teaching at school, can be an effective and efficient way to improve math skills. Moreover, barriers to implementation seem relatively low, because students use it at home and the tool does not require a large learning effort by teachers, because they do not need to adapt their teaching to the tool.

## Appendix 1

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | | Model 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of obs. | | 337 | | | | 337 | | | | 337 | | | | 323 | | | | 282 | | |
| Number of groups | | | | | | 13 | | | | 13 | | | | 13 | | | | 11 | | |
| R-squared within | | | | | | 0.00 | | | | 0.00 | | | | 0.07 | | | | 0.09 | | |
| R-squared between | | | | | | 0.23 | | | | 0.23 | | | | 0.57 | | | | 0.79 | | |
| R-squared overall | | | | | | 0.02 | | | | 0.03 | | | | 0.12 | | | | 0.14 | | |
| Sigma_u | | | | | | 0.13 | | | | 0.13 | | | | 0.00 | | | | 0.00 | | |
| Sigma_e | | | | | | 0.46 | | | | 0.46 | | | | 0.47 | | | | 0.44 | | |
| Rho | | | | | | 0.07 | | | | 0.08 | | | | 0.00 | | | | 0.00 | | |
| percentage growth score per second math | Coef. | St. Err. | *t* | *p* | Coef. | St. Err. | *t* | *p* | Coef. | St. Err. | *t* | *p* | Coef. | St. Err. | *t* | *p* | Coef. | St. Err. | *t* | *p* |
| Treatment dummy | 0.16 | 0.09 | 1.82 | 0.09 | 0.16 | 0.09 | 1.89 | 0.06 | 0.16 | 0.09 | 1.90 | 0.06 | 0.15 | 0.05 | 3.04 | 0.00 | 0.12 | 0.05 | 2.24 | 0.03 |
| Score text comprehension | | | | | | | | | 0.00 | 0.00 | 0.64 | 0.53 | 0.00 | 0.00 | 1.33 | 0.18 | 0.00 | 0.00 | 0.99 | 0.32 |
| Primary school ability test math | | | | | | | | | | | | | 0.00 | 0.00 | -0.31 | 0.76 | 0.00 | 0.00 | -0.53 | 0.59 |
| Primary school ability test total score | | | | | | | | | | | | | 0.00 | 0.01 | -0.66 | 0.51 | 0.00 | 0.01 | -0.27 | 0.79 |
| Dyslexia | | | | | | | | | | | | | 0.11 | 0.14 | 0.82 | 0.41 | 0.11 | 0.17 | 0.69 | 0.49 |
| Primary school advice: pre university/higher general education/prevocational education level 2 | | | | | | | | | | | | | 0.26 | 0.20 | 1.31 | 0.19 | 0.40 | 0.18 | 2.21 | 0.03 |
| Primary school advice: pre university/higher general education | | | | | | | | | | | | | 0.17 | 0.08 | 2.19 | 0.03 | 0.20 | 0.08 | 2.40 | 0.02 |
| Primary school advice: pre university | | | | | | | | | | | | | 0.18 | 0.11 | 1.74 | 0.08 | 0.23 | 0.13 | 1.71 | 0.09 |
| Age | | | | | | | | | | | | | 0.06 | 0.07 | 0.94 | 0.35 | 0.04 | 0.07 | 0.58 | 0.56 |
| Gender | | | | | | | | | | | | | -0.04 | 0.07 | -0.59 | 0.55 | 0.01 | 0.07 | 0.11 | 0.91 |
| Oldest Child | | | | | | | | | | | | | -0.09 | 0.07 | -1.36 | 0.17 | -0.06 | 0.05 | -1.08 | 0.28 |
| Religion: Hinduism | | | | | | | | | | | | | -0.19 | 0.11 | -1.75 | 0.08 | -0.33 | 0.09 | -3.52 | 0.00 |
| Religion: Muslim | | | | | | | | | | | | | -0.46 | 0.14 | -3.32 | 0.00 | -0.52 | 0.13 | -3.90 | 0.00 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Religion: Christian | | | | | | | | | | | | | 0.01 | 0.08 | 0.15 | 0.88 | -0.04 | 0.09 | -0.45 | 0.65 |
| Situation at home: parents divorced | | | | | | | | | | | | | -0.01 | 0.06 | -0.12 | 0.90 | -0.02 | 0.08 | -0.31 | 0.76 |
| Situation at home: mother or father deceased | | | | | | | | | | | | | 0.34 | 0.24 | 1.45 | 0.15 | 0.30 | 0.21 | 1.43 | 0.15 |
| Teacher attitude | | | | | | | | | | | | | | | | | 0.08 | 0.07 | 1.15 | 0.25 |
| Teacher use of tool | | | | | | | | | | | | | | | | | -0.03 | 0.06 | -0.46 | 0.65 |
| Constant | 0.33 | 0.07 | 4.60 | 0.00 | 0.16 | 0.09 | 1.89 | 0.06 | 0.18 | 0.26 | 0.71 | 0.48 | 1.68 | 3.47 | 0.48 | 0.63 | 0.66 | 4.14 | 0.16 | 0.87 |

Reference primary school advice: prevocational education level 2

Reference religion: none

Reference situation at home: both parents at home

Primary school fixed effects added standard errors clustered at the class level

## Appendix 2

| | Model 1 | | | | Model 2 | | | | Model 3 | | | | Model 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of obs. | | 337 | | | | 337 | | | | 323 | | | | 282 | | |
| Number of groups | | 13 | | | | 13 | | | | 13 | | | | 11 | | |
| R-squared within | | 0.002 | | | | 0.000 | | | | 0.022 | | | | 0.035 | | |
| R-squared between | | 0.281 | | | | 0.297 | | | | 0.462 | | | | 0.805 | | |
| R-squared overall | | 0.018 | | | | 0.020 | | | | 0.069 | | | | 0.100 | | |
| Sigma_u | | 0.120 | | | | 0.126 | | | | 0.000 | | | | 0.000 | | |
| Sigma_e | | 0.464 | | | | 0.464 | | | | 0.463 | | | | 0.442 | | |
| Rho | | 0.062 | | | | 0.068 | | | | 0.000 | | | | 0.000 | | |
| percentage growth score per second math | Coef. | Robust St. Err. | *t* | *p* | Coef. | Robust St. Err. | *t* | *p* | Coef. | Robust St. Err. | *t* | *p* | Coef. | Robust St. Err. | *t* | *p* |
| Minutes practiced per week | 0.004 | 0.004 | 0.880 | 0.377 | 0.004 | 0.004 | 0.890 | 0.375 | 0.007 | 0.004 | 1.720 | 0.086 | 0.009 | 0.005 | 1.930 | 0.053 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Score text comprehension | | | | | 0.002 | 0.003 | 0.640 | 0.519 | 0.003 | 0.004 | 0.770 | 0.442 | 0.001 | 0.004 | 0.370 | 0.710 |
| Primary school ability test math | | | | | | | | | -0.001 | 0.002 | -0.320 | 0.746 | -0.001 | 0.002 | -0.440 | 0.662 |
| Primary school ability test total score | | | | | | | | | -0.002 | 0.007 | -0.230 | 0.815 | 0.002 | 0.007 | 0.330 | 0.738 |
| Dyslexia | | | | | | | | | 0.115 | 0.088 | 1.300 | 0.193 | 0.122 | 0.110 | 1.110 | 0.266 |
| Primary school advice: pre university/higher general education/prevocational education level 2 | | | | | | | | | 0.151 | 0.249 | 0.610 | 0.544 | 0.199 | 0.128 | 1.550 | 0.120 |
| Primary school advice: pre university/higher general education | | | | | | | | | 0.161 | 0.080 | 2.000 | 0.045 | 0.195 | 0.074 | 2.630 | 0.009 |
| Primary school advice: pre university | | | | | | | | | 0.235 | 0.092 | 2.570 | 0.010 | 0.246 | 0.093 | 2.650 | 0.008 |
| Age | | | | | | | | | 0.017 | 0.063 | 0.270 | 0.788 | 0.002 | 0.062 | 0.030 | 0.980 |
| Gender | | | | | | | | | -0.023 | 0.071 | -0.330 | 0.744 | 0.022 | 0.071 | 0.310 | 0.759 |
| Oldest Child | | | | | | | | | -0.066 | 0.052 | -1.270 | 0.203 | -0.043 | 0.045 | -0.960 | 0.338 |
| Religion: Hinduism | | | | | | | | | -0.392 | 0.185 | -2.130 | 0.034 | -0.563 | 0.160 | -3.520 | 0.000 |
| Religion: Muslim | | | | | | | | | -0.421 | 0.130 | -3.240 | 0.001 | -0.436 | 0.128 | -3.400 | 0.001 |
| Religion: Christian | | | | | | | | | 0.021 | 0.077 | 0.280 | 0.783 | 0.000 | 0.000 | 0.000 | 0.000 |
| Situation at home: parents divorced | | | | | | | | | 0.035 | 0.060 | 0.590 | 0.557 | 0.017 | 0.066 | 0.250 | 0.802 |
| Situation at home: mother or father deceased | | | | | | | | | 0.387 | 0.197 | 1.960 | 0.050 | 0.364 | 0.180 | 2.020 | 0.043 |
| Teacher attitude | | | | | | | | | | | | | 0.078 | 0.047 | 1.680 | 0.094 |
| Teacher use of tool | | | | | | | | | | | | | 0.005 | 0.041 | 0.110 | 0.909 |
| Constant | 0.403 | 0.066 | 6.100 | 0.000 | 0.259 | 0.203 | 1.270 | 0.203 | 0.795 | 3.475 | 0.230 | 0.819 | -1.335 | 3.908 | -0.340 | 0.733 |

Reference primary school advice: prevocational education level 2

Reference religion: none

Reference situation at home: both parents at home

Standard errors clustered at the class level

## References


Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *Economic Journal, 112*, 735-765.

Arroyo, I., Park Woolf, B., Royer, J. M., Tai, M., & English, S. (2010). Improving Math Learning through Intelligent Tutoring and Basic Skills Training. In V. Aleven, J. Kay & J. Mostow (Eds.), *ITS 2010, Part i, LNCS 6094* (pp. 423-432). Berlin Heidelberg: Springer-Verlag.

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

Becker, H. J. (1990). Effects of Computer Use on Mathematics Achievement. Findings from a Nationwide Field Experiment in Grades Five to Eight. Classes: Rationale, Study Design, and Aggregate Effect Sizes. Baltimore, MD.: Center for Research on Elementary and Middle Schools.

Borman, G. D., Benson, J. G., & Overman, L. (2008). A Randomized Field Trial of the Fast ForWord Language Computer-Based Training Program. *Educational Evaluation and Policy Analysis, 31*(1), 82-106.

Burns, M. K., Kanive, R., & DeGrande, M. (2012). Effect of a Computer-Delivered Math Fact Intervention as a Supplemental Intervention for Math in Third and Fourth Grades. *Remedial and Special Education, 33*, 3.

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198-215.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Commissie Meijerink. (2008). Over de drempels met taal en rekenen - Hoofdrapport van de expertgroep doorlopende leerlijnen taal en rekenen. Enschede.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., . . . Sussex, W. (2007). Effectiveness of reading and mathematics software products: findings from the first student cohort (Vol. 112). Wastington D.C.: US Department of Education / Institute of Education Sciences.

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics* (4 ed.). London: Sage.

Funnekotter, B. (2012, 11 juni 2012). Middelbare scholieren scoren massaal onvoldoende bij rekentoets, *NRC*.

Goolsbee, A., & Guryan, J. (2006). The Impact of Internet Subsidies in Public Schools. *Review of Economics and Statistics, 88*(2), 336-347.

Hattie, J. (2009). *Visible learning - A synthesis of over 800 meta-analysis relating to achievement*. New York, NY: Routledge.

KNAW. (2009). Rekenonderwijs op de basisschool. Advies en sleutels tot verbetering. Amsterdam: KNAW.

Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75-94.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economic and Statistics, 89*(4), 721-736.

MacGregor, S. K., Shapiro, J. Z., & Niemiec, R. (1988). Effects of A Computer-Augmented Learning Environment on Math Achievement For Students With Differing Cognitive Style. *Journal of Educational Computing Research, 4*(4), 453-465.

Machin, S., McNally, S., & Silva, O. (2007). New technologies in schools: Is there a payoff? *The Economic Journal, 117*, 1145-1167.

Muiswerk. (2013) Retrieved January 29th, 2013, from www.muiswerk.nl

Pilli, O., & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers and Education, 61*, 62-71.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. doi: 10.1093/biomet/70.1.41

Rouse, C. E., & Krueger, A. B. (2004). Putting Computerized Instruction to the Test: A Randomized Evaluation of a `Scientifically-Based' Reading Program. *Economics of Education Review, 23*(4), 323-338.

Teule, P. (2012, 25 July 2012). Wat kost zittenblijven nou echt? Retrieved January 30th, 2013, from http://sargasso.nl/wat-kost-zittenblijven-nou-echt/

Tomlinson, C. A. (2004). Research evidence for differentiation. *School Administrator, 61*(7), 30.

Tomlinson, C. A., & Kalbfleisch, M. L. (1998). Teach me, teach my brain: A call for differentiated classrooms. *Educational Leadership, 56*(3), 52-55.

## Tables

**Table I – Descriptive statistics of first year students of Dendron College in the school year 2012/2013**

| | *Obs.* | *Average* | *St. Dev.* | *Min* | *Max* |
|---|---|---|---|---|---|
| Primary school ability test: math | 339 | 56.55 | 24.30 | 4 | 100 |
| Primary school ability test: language (mother tongue only) | 338 | 56.31 | 24.38 | 3 | 100 |
| Primary school ability test: world studies | 304 | 55.21 | 25.51 | 0 | 100 |
| Primary school ability test: total score | 345 | 538.25 | 6.23 | 517 | 550 |
| Female student | 355 | 0.56 | 0.50 | 0 | 1 |
| Student diagnosed with dyslexia | 355 | 0.06 | 0.24 | 0 | 1 |
| Age (in completed years) | 355 | 12.30 | 0.48 | 11 | 14 |
| Oldest child in her/his household (1=yes) | 355 | 0.59 | 0.49 | 0 | 1 |

**Table II – *t*-statistics and Mann-Whitney Statistics of Treatment and Control Group**

| | Control group | | | Treatment group | | | T-statistic[14] |
|---|---|---|---|---|---|---|---|
| Variable | N[a] | Average | Std. Dev. | N[a] | Average | Std. Dev. | |
| Primary school ability test math | 104 | 53.53 | 23.95 | 219 | 58.39 | 24.36 | -1.69 |
| Primary school ability test language | 104 | 55.38 | 24.14 | 218 | 56.81 | 24.35 | -0.49 |
| Primary school ability test world studies | 96 | 54.17 | 25.04 | 194 | 55.69 | 26.11 | -0.47 |
| Primary school ability test total score | 106 | 537.73 | 6.29 | 222 | 538.56 | 6.17 | -1.13 |
| Female | 107 | 0.55 | 0.50 | 230 | 0.57 | 0.50 | -0.24 |
| Dyslexia | 107 | 0.06 | 0.23 | 230 | 0.07 | 0.25 | -0.32 |
| Age | 107 | 12.27 | 0.47 | 230 | 12.29 | 0.47 | -0.29 |
| Oldest child | 107 | 0.53 | 0.50 | 230 | 0.60 | 0.49 | -1.24 |
| Pretest in September (T0) | 107 | 56.05 | 13.60 | 230 | 54.18 | 13.91 | 1.15 |
| | | | | | | | *Mann-Whitney Z-score* |
| Variable | | Rank Sum | Expected | | Rank Sum | Expected | |
| Primary School Advice | 107 | 17433.0 | 18083.0 | 230 | 39520.0 | 38870.0 | -0.84 |
| Place of residency | 107 | 17966 | 18083.0 | 230 | 38987.0 | 38870.0 | -0.14 |
| Country of Birth | 107 | 18083.5 | 18083.0 | 230 | 38869.5 | 38870.0 | 0.00 |
| Nationality | 107 | 18090 | 18083.0 | 230 | 38863.0 | 38870.0 | 0.05 |
| Religion | 107 | 18320.0 | 18083.0 | 230 | 38633.0 | 38870.0 | 0.33 |
| Situation at home | 107 | 18027.0 | 18083.0 | 230 | 38926.0 | 38870.0 | -0.11 |
| Primary School | 107 | 18003.0 | 18083.0 | 230 | 38950.0 | 38870.0 | -0.10 |

[a] Note that not all students wrote the primary school ability test and that not all primary schools delivered detailed information on the subparts of that test to the secondary school. Therefore, a small number of observations is missing for these variables.

**Table III – Baseline Results: *t*-test of the effect of the experiment on various growth indicators**

| | *Control group (n=107)* | | *Treatment group (n=230)* | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Average** | **Std. Dev.** | **Average** | **Std. Dev.** | ***t*-statistic** | |
| **Absolute growth in test score T0-T1** | 2.645 | 10.971 | 7.422 | 11.533 | -3.594 | *** |
| **Percentage growth in test score T0-T1** | 0.072 | 0.233 | 0.170 | 0.269 | -3.236 | *** |
| **Absolute growth test score per second T0-T1** | 0.017 | 0.021 | 0.022 | 0.021 | -2.237 | ** |

[14] Apart from *t*-tests, all continuous and binary variables are also tested using the non-parametric Wilcoxon ranksum test, to compare discrete measures between small groups. Similar results are obtained using the two-side *t*-test and the Wilcoxon ranksum test.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Percentage growth test score per second** | | | | | | |
| **T0-T1** | 0.330 | 0.409 | 0.487 | 0.505 | -2.800 | *** |

* = significant at the 10% level **=significant at the 5% level ***=significant at the 1% level

**Table IV – The Returns to Practicing Online: Class random effects regressions.**

| Dep. Var= Percentage increase score per second math | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Treatment | 0.16 * | 0.16 * | 0.16 * | 0.15 *** | 0.12 ** |
| p-value | 0.09 | 0.06 | 0.06 | 0.00 | 0.03 |
| Covariates($X_{j\,i}$) | | | score text comprehension | score text comprehension, student ability, student characteristics | score text comprehension, student ability, student characteristics, teacher attitude towards tool |
| Observations (*n*) | 337 | 337 | 337 | 323 | 282[15] |
| Groups (classes) | n/a | 13 | 13 | 13 | 11 |

* = significant at the 10% level **=significant at the 5% level ***=significant at the 1% level

Standard errors clustered at the class level.

[15] One teacher, teaching two classes, did not fill out all statements in the questionnaire. Given that statements of different kinds were forgotten (and not only the ones that would reveal his/her positive or negative attitude towards the tool) we are confident that this does not influence the results. Furthermore, running models 1 to 4 with only these 11 classes yield almost identical coefficients and significance, compared with the 13 classes presented in Table IV (results available upon request).

**Table V – Factors influencing students' practicing behavior: Class random effects regression**

| Average number of minutes practiced per week | Coef. | t-statistic | |
|---|---|---|---|
| Primary school ability test math | 0.04 | 1.36 | |
| Primary school ability test language | -0.26 | -1.21 | |
| Dyslexia | -0.26 | -0.14 | |
| Primary school advice: prevocational education level 2 | Reference | | |
| Primary school advice: pre university/higher general education/prevocational education level 2 | -1.97 | -0.88 | |
| Primary school advice: pre university/higher general education | 1.85 | 0.42 | |
| Primary school advice: pre university | 2.61 | 0.56 | |
| Age | 2.24 | 1.49 | |
| Female | 1.23 | 1.48 | |
| Oldest child | -0.48 | -0.61 | |
| Religion: None | Reference | | |
| Religion: Muslim | -2.77 | -0.87 | |
| Religion: Christian | -0.69 | -0.69 | |
| Situation at home: both parents at home | Reference | | |
| Situation at home: parents divorced | -3.66 | -3.95 | *** |
| Situation at home: mother or father deceased | -5.86 | -3.65 | *** |
| Teacher attitude | 10.09 | 5.68 | *** |
| Teacher use of tool | 1.83 | 0.93 | |
| Constant | 89.37 | 0.82 | |

* = significant at the 10% level **=significant at the 5% level ***=significant at the 1% level

Standard errors clustered at the class level.

Selection: only the treatment classes for which we have full information on the teacher questionnaire, N = 177.

**Table VI – The Returns to Practice Time**

| *Dep. Var= Percentage increase score per second math* | *Model 1* | *Model 2* | *Model 3* | *Model 4* |
|---|---|---|---|---|
| Minutes per week practiced | 0.004 | 0.004 | 0.007* | 0.009** |
| p-value | 0.337 | 0.375 | 0.086 | 0.053 |
| Covariates($X_{j\,i}$) | Class random effects | Class random effects, score text comprehension | Class random effects, score text comprehension, student ability, student characteristics | Class random effects, score text comprehension, student ability, student characteristics, teacher attitude towards tool |
| Observations (*n*) | 337 | 337 | 323 | 282 |
| Groups (classes) | 13 | 13 | 13 | 11 |

* = significant at the 10% level **=significant at the 5% level ***=significant at the 1% level

Standard errors clustered at the class level.

**Figures**

**Figure I – Overview of the Field Experiment**