

# The Differential Effect of Basic Mathematics Skills Homework via a Web-Based Intelligent Tutoring System Across Achievement Subgroups and Mathematics Domains: A Randomized Field Experiment

## Citation for published version (APA):

Bartelet, D., Ghysels, J., Groot, W., Haelermans, C., & van den Brink, H. M. (2016). The Differential Effect of Basic Mathematics Skills Homework via a Web-Based Intelligent Tutoring System Across Achievement Subgroups and Mathematics Domains: A Randomized Field Experiment. *Journal of Educational Psychology*, 108(1), 1-20. <https://doi.org/10.1037/edu0000051>

## Document status and date:

Published: 01/01/2016

## DOI:

[10.1037/edu0000051](https://doi.org/10.1037/edu0000051)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 03 Nov. 2021

# The Differential Effect of Basic Mathematics Skills Homework via a Web-Based Intelligent Tutoring System Across Achievement Subgroups and Mathematics Domains: A Randomized Field Experiment

Dimona Bartelet, Joris Ghysels, and Wim Groot  
Maastricht University

Carla Haelermans  
Maastricht University and Delft University of Technology

Henriëtte Maassen van den Brink  
Maastricht University and University of Amsterdam

This article examines an educational experiment with a unique combination of 3 elements: homework, the use of information and communication technology and a large degree of freedom of choice (student autonomy). More particularly, we study the effectiveness of a web-based intelligent tutoring system (ITS) that a school offers to its students as optional homework, in a sample of 355 first-year secondary students, using an experimental design. The results show that whether students make this noncompulsory homework in the web-based ITS is dependent on their prior achievement and their teacher, and if they practice, they tend to choose easier modules. Students thus do not seem to optimize learning gains, but rather balance their perceived value of practicing and their “expectancy,” as found previously for regular (compulsory, non-ITS) homework. Regarding the value of choice, findings suggest that students may be poor judges of their most beneficial exercises pattern. Differentiation in ITS does not guarantee learning, unless scaffolding is tightly linked to it. In line with previous research, this research shows that practice does matter, though at varying intensities. Yet it also clarifies that motivation to make an effort is required, which adds to the lessons learned from homework research; both the perceived value and expectancy are at play and can be enhanced, for example, by teachers.

*Keywords:* randomized field experiment, junior high, noncompulsory homework practice program, mathematics subskills, differential treatment effect

*Supplemental materials:* <http://dx.doi.org/10.1037/edu0000051.supp>

Besides being increasingly important in daily life, basic mathematics skills form the foundation for the development of more advanced math skills during secondary education. As in many countries (Organisation for Economic Cooperation and Development [OCED], 2014), basic mathematics skills are therefore a source of social concern in the Netherlands. Currently, the mathematics skills level of many students entering secondary school in the Netherlands is lagging behind the national performance standards defined for mathematics by the Meijerink Commission (Koninklijke Nederlandse Akademie van Wetenschappen, 2009; Scheltens, Hemker, & Vermeulen, 2013). The report of this com-

mission describes which aspects of mathematics students should be familiar with at different grade levels of primary and secondary education (Commissie Meijerink, 2008). Van Groenestijn (2007) reported that, depending on the secondary education track, 10% to 50% of first-year secondary students (seventh grade) have mathematics skills equivalent to or lower than the skills they are expected to possess at the end of Grade 4.<sup>1</sup> Yet to succeed in solving algebraic equations or complex geometric problems, secondary students have to apply, integrate, and extend mathematics skills that they have learned in primary school (Ketterlin-Geller, Chard, & Fien, 2008).

In the Netherlands, the publication of the previously mentioned Meijerink report led to the introduction of a new mathematics skills exam at the end of secondary school, in addition to the graduation exam system already in place. Secondary schools are supposed to prepare students for the additional exam and have adjusted their educational policy accordingly. Some decided to offer additional math classes, some test students and offer remediation classes, and others rely on intelligent tutoring systems (ITSs) to make sure their students pass the exam.

---

This article was published Online First June 29, 2015.

Dimona Bartelet, Joris Ghysels, and Wim Groot, Top Institute for Evidence Based Education Research, Maastricht University; Carla Haelermans, Top Institute for Evidence Based Education Research, Maastricht University, and Centre for Innovations and Public Sector Efficiency Studies, Delft University of Technology; Henriëtte Maassen van den Brink, Top Institute for Evidence Based Education Research, Maastricht University, and Faculty of Economics and Business, University of Amsterdam.

Correspondence concerning this article should be addressed to Carla Haelermans, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: [carla.haelermans@maastrichtuniversity.nl](mailto:carla.haelermans@maastrichtuniversity.nl)

---

<sup>1</sup> In this article, grading is used according to the American school system.

ITSs are a specific type of computer-assisted instruction (CAI; a categorical name for “programs that use technology to enhance cognitive achievement”; Slavin, Lake, & Groff, 2009, p. 843), which are thought to be particularly effective because they are adaptive and interactive (Steenbergen-Hu & Cooper, 2013), and are often used in remedial education contexts (Crawford & Ketterlin-Geller, 2008; Gu, Fu, & Tong, 2010; Ketterlin-Geller et al., 2008).

However, although evaluations of ITS tools have shown positive results (Cheung & Slavin, 2013; Steenbergen-Hu & Cooper, 2013), evaluations hardly ever focused on the use of the tools outside of the school context, so there is not much known about its effectiveness in the absence of direct teacher supervision. Moreover, the same authors stress that remediation is currently the main use of the tools, and evaluations do not often shed light on the learning effectiveness of CAI and ITS in regular education for a general audience.

Furthermore, research on the effectiveness of homework posits that offering choice in homework is likely to be effective, when organized properly (Patall, Cooper, & Wynn, 2010). However, the latter findings remain within the context of written homework assignments and did not consider the framework of ITS. Hence, we do not know whether this works the same way for ITS homework.

Therefore, in this study, we analyze the use of such an ITS, called “Mousework,”<sup>2</sup> as an offer for homework for seventh grade students (i.e., first-year students in Dutch secondary education), using a randomized experiment. The school of study, Dendron College, grants its students access to the online version of the ITS and invites them to work with the tool at home for half an hour a week. At school, the tool is not used, but teachers are invited to follow-up on their students and organize, at regular intervals, computer tests the ITS uses to increase its adaptability to every student’s personal needs.

The analysis presented here adds to three strands of the literature. First, it evaluates the effectiveness of an ITS in an understudied context, that is, its use at home for all students in secondary education. As discussed before, previous evaluations hardly ever focused on the use of the tools outside of the school context and for the full student population. Consequently, it is of interest to see whether an ITS will prove effective beyond the target group of students with weakly developed math skills, that is, among students with regular or even excellent math skills.

Second, the analyses will apply the theory of homework effort of Trautwein et al. (Trautwein, Lüdtke, & Schnyder, 2006) to the particular field of online homework. According to Trautwein’s idea that, in order to make an effort, students need both to value the assignment and believe that they will be able to succeed in the task, we will develop hypotheses about the overall effort students put into the Mousework tool and their choice of particular subsets of exercises, linked to mathematical domains that vary in their perceived value and level of difficulty.

Third, the benefits of structured choice will be studied in an ITS context. As the ITS studied here allows students to make their own choices, both initially (whether to use it or not) and during the process (which sets of exercises to make), it seems like an excellent testing bed for the extension of the choice research to online homework. With this we contribute to earlier work done by Patall et al. (2010).

In the literature and theoretical framework section we will elaborate on the three strands of the literature we will rely on for our analyses. Then, the context of the experiment and the content of the ITS tool are described in more detail, followed by the data operationalization of the outcome measures, achievement-level groups, practice behavior, and the teacher variables. Subsequently, the methodology is explained, and the results are presented. Then, in the last section, we discuss the results and draw conclusions on the differential effectiveness of the ITS Mousework.

### Literature, Theoretical Base, and Hypotheses: Freedom of Choice in the Context of an Information and Communication Technology (ICT) Tool for Homework

To elaborate our research design, we drew on theory and previous empirical research regarding three particular topics: the use of ICT tools, the function of homework, and the value of freedom of choice in the learning process.

#### Theory and Literature on the Function of Homework and the Value of Freedom of Choice in the Learning Process

Regarding homework, the conceptual framework developed by Trautwein is our starting point. Trautwein developed a framework to analyze how homework contributes to the learning process (Dettmers, Trautwein, Lüdtke, Kunter, & Baumert, 2010; Trautwein et al., 2006) and tested its relevance empirically. The theoretical “model combines elements of expectancy–value theory (Eccles, 1983; Eccles & Wigfield, 2002), self-determination theory (Deci & Ryan, 2002; and others), and research on learning and instruction (Boekaerts, 1999; and others)” (Trautwein et al., 2006, p. 438). The model takes into account the three protagonists in the homework process (students, teachers, and parents) and covers six major groups of variables (achievement, homework behavior, homework motivation, student characteristics, parental behavior, and the learning environment). Dettmers et al. (2010) summarize the model as follows:

The model predicts that the effort students spend on their homework assignments (i.e., doing their best to solve the tasks assigned) is positively related to their achievement. In line with expectancy–value theory, homework effort is conceptualized as strongly influenced by expectancy and value beliefs, representing two aspects of homework motivation. The expectancy component reflects a student’s belief in being able to complete a given homework assignment successfully. The value component describes students’ reasons for doing a task in terms of the importance of succeeding in a specific domain, the enjoyment of engaging in the activity, the utility of the activity, and the costs associated with it. (Dettmers et al., 2010, p. 468)

For the learning situation we study in this contribution, the expectancy and value elements of the Trautwein framework are particularly relevant (see also Pintrich, 2000; Winters, Greene, & Costich, 2008). Given the large degree of freedom associated with

<sup>2</sup> The Dutch name of Mousework is *Muiswerk*. In addition to the Dutch version, an international version exists that is used by several international schools in Europe and in other parts of the world.

the homework assignment, students' motivation is crucial. Students had to make several successive decisions: first, whether they would engage in the homework assignment at all (use the tool), and subsequently, which of the modules (i.e., math domains) they would work with.

We hypothesize that for first-year students, who we study here, the value of additional practicing with basic mathematical skills may not be entirely clear. They may have been told that they will be required to do a specific numeracy test at the end of high school, but that is at least four years away, and hence is unlikely to be a strong motivator. Moreover, the fact that practicing is not graded takes away any instrumental motivation and may be perceived as a signal degrading the value of the task.

Yet there are also some counterbalancing aspects related to value. First, some teachers are positively inclined toward the tool and may have transmitted their positive attitude to their students. Second, some students have done more intermediate tests than others, which may have functioned as a positive signal of the value that the school attaches to practice even without grading it. Therefore, we will look into potential variation of the use and effectiveness of the tool related to the attitude of the teacher, the teachers' use of the tool,<sup>3</sup> and the frequency of intermediate testing.

A second set of hypotheses regards the expectancy of the students. For the use of the tool by students in general, we hypothesize that expectancy functions as a negative motivator for students with relatively bad prior math experiences (e.g., low grades in primary school), although the adaptive nature of the software makes sure that the exercises offered never exceed the practical skills of the student so as to frustrate her or him. A further aspect of expectancy regards the choice of modules. Expectancy may motivate students to choose those modules they believe they will be good at, and thus makes them shy away from the more advanced modules. More specifically, we expect a disproportional interest in the modules on numbers (the entry level of numeracy which is addressed at the start of primary school; Cowan & Powell, 2014; Geary, 2011) at the detriment of the other modules on measurement, especially proportions, which tend to be perceived as more difficult.

With regard to the role of the teacher, Trautwein et al. (2006) highlight the importance of control, although both positive (a signal of value) and negative (external rather than internal motivation) factors may be at play. Patall et al. (2010) showed, in a written homework context, that the provision of choice may increase both the completion of the homework and support learning (as revealed by a summative test). They stress, however, that although autonomy with regard to homework may be beneficial, choice should not be unlimited. A clear task structure and explicit goals are required (Patall et al., 2010, p. 898). Thinking along the lines of Patall et al., the "Muiswerk" tool offers a well-structured framework of homework choices. Students can decide about participation and choose the modules they want to engage in, but at the same time, practicing is structured by the software and the repeated formative tests, which also serve to update the level of exercise modules offered to the student. As such, the tool is an ICT alternative to the choice in written homework advocated by Patall et al., and may largely overcome a major hindrance mentioned by the authors: the additional workload for teachers to develop and manage various options in a homework assignment.

## Literature on the Use of ICT Tools

With regard to the literature on ICT for learning, the Muiswerk tool offers an example of a relatively rare species of ITS: tools used outside of the school. In their recent meta-analysis, Steenbergen-Hu and Cooper (2013) observed that almost all ITSs have been evaluated when used at school, where the context of supervision and autonomy is completely different from the homework setting.

Therefore, we also looked into previous research evaluating tools for CAI, the less advanced versions of ITS (e.g., less adaptive), for which a wider array of research on optional tools for mathematics instruction is available. The so-called supplemental CAI offers tools for noncore instruction. A recent meta-analysis by Cheung and Slavin (2013) included 74 high-quality studies on the effect of CAI technology on children's mathematics performance in K-12 classrooms—all evaluations that use experimental or quasi-experimental methods. Of these studies, the majority (70%) investigate the effect of supplemental CAI, though most are used at school for remediation, and the outcomes are mixed (e.g., see Dynarski et al., 2007, and Campuzano, Dynarski, Agodini, & Rall, 2009, for no impact, and see Gatti & Petrochenkov, 2009, for a positive impact).

However, the differential effect of a supplementary CAI on mathematics achievement for student groups in secondary school, differing in mathematics skills level at the pretest, has rarely been studied (Burns, Kanive, & DeGrande, 2012). Many previous studies have investigated the effect of supplementary CAI tools exclusively in low-achieving/at-risk student samples (e.g., ethnic minorities; low socioeconomic status [SES]). As a result, no conclusions can be drawn on differences between groups of students varying in mathematics skill level at the onset of the intervention, even though the Trautwein theory on homework would predict that these differences exist.

Moreover, mathematical development research stresses that children not only differ in the extent to which their mathematics skills are developed but also exhibit deficits in different domains (Käser et al., 2013). Researchers have theorized that mathematics achievement is not a unitary construct—instead, it consists of distinguishable components. These components are related, but supported by different cognitive processes, and consequently are sufficiently distinct that children who have severe deficits in one domain (e.g., numbers) do not necessarily exhibit severe difficulties in other mathematics domains (e.g., geometry; Dowker, 2005; Gersten, Jordan, & Flojo, 2005; but also see Cowan & Powell, 2014).

An intervention might be effective in enhancing students' skills in one mathematics domain but not in the others. This was also observed in the previously mentioned CAI study by Gatti and Petrochenkov (2009). The authors analyzed three specific subtests and revealed that the supplementary CAI tool fostered students' performance on the "operations and computation" and "process and applications" subtests, but not the "concepts and communications" subtest.

<sup>3</sup> The teachers' use of the tool indicates whether the teacher uses the tool to log in and check on his or her students' practice behavior. It explicitly does NOT mean that the tool is being used in class.

A consequence of selective effectiveness is that deficits in other domains may remain, and schools need to consider what supplementary actions to take to improve students' achievement in the nonaffected domains as well. One possibility is the adjustment of the intervention program followed by a reevaluation. Another option is the search for alternative educational resources that have a positive impact on the skills that are not affected by the first intervention. In the end, a balanced implementation of different interventions might prove to be more successful than the large-scale implementation of a single intervention.

Yet the particular setting of the present study should not be forgotten. We evaluate an ICT tool in a homework setting with a low degree of external incentives (no grading, variable level of follow-up by teachers, never comparable with collecting written homework at a specific moment). Therefore, the use of the tool relies on self-regulatory learning processes. Some of these processes are found to be less developed in children with low academic attainment or learning difficulties (Cleary, 2006; Núñez et al., 2005), and therefore it may be argued that low achievers are less inclined to use supplemental ICT tools (e.g., CAI or ITS; Steenbergen-Hu & Cooper, 2013, p. 985), which adds to the expectancy argument developed regarding homework completion.

## Hypotheses

Summarizing, we derive from the homework literature that the value and expectancy that students relate to the homework assignment are likely to be important determinants of their practicing behavior and, indirectly, their learning gains. Therefore, we hypothesize that students will tend to engage in modules perceived as easier, and, second, that students with relatively weakly developed math skills will tend to work less with the tool than other students because they are less confident in their ability to succeed in the assignment (expectancy). Consequently, the learning potential of the tool will most likely not be fully used, especially not by students who are performing weakly on earlier math tests. Furthermore, the homework and the freedom-of-choice literature suggests that teachers' attitudes are likely to influence practicing behavior, as does the clear structure of the tool, which can be expected to favor its use. Finally, the literature on supplemental CAI and ITS suggests that the tool may be effective, but similar tools have not yet been tested in an optional homework setting at secondary school addressing all students, rather than a group previously identified as requiring remediation.

Schematically, theoretical framework leads to the following hypotheses that we can test in the setting of the present study:

1. A teacher with a positive attitude and who uses the tool more actively than other teachers (to check on students' practice behavior) will inspire his or her students to practice more (adding to the perceived "value" of the homework).
2. Students with a relatively weak initial performance will tend to work less with the tool than the other students (even though they can be expected to benefit more from practice; expectancy trumps rational choice).
3. Students engage disproportionately in the easier modules (i.e., numbers) at the detriment of the more difficult

modules (i.e., measurement and proportions), irrespective of their level of expertise (i.e., they do not optimize their exercises portfolio from a learning perspective, and hence Mousework does not offer sufficient choice structure).

4. The effectiveness of the tool Mousework depends on the type of student (with respect to initial performance) and the module in which students have practiced: Practicing helps, and helps students with an initially low performance relatively more, but its effect is limited to the domains practiced.

## Context of the Experiment

### Participants

Participants are first-year secondary students (corresponding to seventh grade in the U.S. school system) attending Dendron College, a midsize school in the south of the Netherlands. Dendron College has about 2,000 middle and high school students who follow one of three major tracks. The Dutch educational system is characterized by a tracking system from seventh grade onward. At the start of secondary education, three tracks are distinguished, namely, a prevocational (*vmbo* [lowest]), a general higher education (*havo*), and a preuniversity (*vwo* [highest]) track. The prevocational education track consists of four subtracks, in which Level 1 is the lowest (mainly practical) track and Level 4 the highest (mainly theoretical) track. Which track students are enrolled in is largely determined by their score on the national Cito (the Dutch national testing and assessment company) test score, a standardized national test completed at the end of primary school (Grade 6) around the age of 12 (Inspectie van het Onderwijs, 2007).

In the school year from 2012 to 2013, the year that the experiment was conducted, 430 first-year students were enrolled in Dendron College. These students were clustered in 17 classes, of which 13 classes ( $N = 355$ ) were included in the experiment. In order to have a homogeneous sample, school management and researchers jointly decided to use only the 13 of 17 classes that were most similar, based on ability, for the experiment. Three prevocational track classes were excluded because of ethical reasons. These classes consisted of students characterized by extreme low academic skills, and denying them access to the tool was viewed as unethical. Therefore, all these students worked with the tool, and no valid counterfactual class could be found. In addition, a preuniversity class that consisted of only highly gifted students was excluded. It was the only class of its kind at Dendron College, and therefore it could not be compared with a counterfactual class. Lastly, another 19 students were excluded. These students either exhibited an extreme growth rate ( $>1,000\%$ ) on one or more domains after having a pretest score of 1 ( $N = 11$ ), or they had no posttest score ( $N = 8$ ), and so the outcome measure that is used in this study (test score growth) could not be computed. We deemed these scores invalid. No significant differences were observed when comparing the background characteristics of the excluded students with the remaining sample. The final sample included 336 students that were nested in 13 theoretical prevocational to preuniversity track classes.

The majority of students included in the sample indicated that they adhered to Christian beliefs or were not religious, and almost all were of Dutch nationality. This is typical for schools located outside of the highly urbanized, central region of the Netherlands. Additional descriptive statistics of the final sample are presented in Table 1. Students' ages ranged from 11 to 14 years, about 55% were girls, and 6% had a clinical diagnosis of dyslexia. It should be noted that for some of the variables, the number of observations deviates from the size of the final sample because of missing values.

### The Tool: Mousework

Aiming at improving or maintaining students' mathematics skills that are (supposed to be) acquired in primary school and that will be tested in the math exam at the end of secondary education, Dendron College choose the Mousework program. Dendron College offers the program for use at home (via the Internet)<sup>4</sup> and asks students to use the program for half an hour a week per subject. Dendron College does not promote the use of the program within the school. Accordingly, teachers reported that they were not to use it in class.

Mousework is a so-called ITS and is available for various subjects. Dendron College uses it for mathematics and Dutch, the mother tongue in the Netherlands. In this contribution, we restrict our analysis to the application for mathematics and refer to this part of the Mousework ITS as "the tool."

The didactical principles on which the tool builds are differentiation, making small steps ("scaffolding"), action (practicing) and variation, direct feedback, and the teacher as supervisor (Muiswerk Educatief, 2013).<sup>5</sup> The tool consists of tutored exercises, organized in submodules ("sets")<sup>6</sup> and modules that are organized within a specific mathematical domain.

Sets and modules are offered to the student in an adaptive and stepwise manner. At the starting point (typically the start of the year), students take an orientation test that determines, for the various mathematics domains of the tool, which sets and modules are offered to the student. At the end of a module, the tool evaluates progress and makes the student repeat (types of) exercises depending on the error level attained in the set. After this repetition within a set, a student can start a subsequent set within the offered modules. New modules are made available once a student takes a new test at school, which calls for a teacher to organize tests at regular intervals. Without these tests, students can finish the sets pertaining to the modules made available to them and can choose to start all over with particular sets, but they cannot make large progress within a particular mathematical domain, because higher level content (explanations and exercises) requires them to pass the threshold level at a test.

It is important to note the balance between choice and structure offered by the tool. Students take a test to get access to a menu of modules and sets within these modules. They decide which sets they want to engage in and are encouraged to finish these sets, because only at the end of a set can they get a reward (a positive signal) and move to another set within the module. Yet the fact that students can choose which set to start with also means that students have the option to determine which mathematical domain(s) they work on. The tool does not force students to elaborate a balanced portfolio of sets. If they do not aim at such a goal themselves, the

menu prompts them to do so, but in the end, external supervision (typically by the teacher) may be required to ensure it.

The tutoring of the tool happens in various ways. Before starting a particular set of exercises, students are invited to read a short explanation about the topic of the exercises (a few screens). While executing an exercise, students can ask for help at any time (get a hint in a pop-up), and after finishing the exercise, the student gets a positive signal when she had the answer right or more explanation when something went wrong. In the latter case, the feedback is as personalized as possible, because the tool contains a database of common errors and the specific feedback these require.

For the type of students we study in this contribution (first-year students), the tool provides exercises in three mathematics domains: numbers, proportions, and measurement.<sup>7</sup> These domains refer to the learning goals described in considerable detail in the report of the Meijerink Commission (Commissie Meijerink, 2008). The numbers domain contains exercises on, among others, subtraction and addition up to 100, 1,000, or 5,000; multiplication tables of 2 to 12; addition, subtraction; and multiplication of negative numbers; powers; and square roots. The proportions domain has exercises on percentages, fractions, decimals, discount, interest, inflation, scale increase/decrease, and so on. The measurement domain contains exercises on length, area, capacity, time, date, temperature, roman numbers, angles, working with the setsquare, circles, 3D figures, and so on. The exercises mainly contain word problems, although mathematical equations are also used, most often for the numbers domain, as well as some figures for proportions and measurement.

Technically, the tool requires students to provide answers to multiple choice and open questions, but rarely requires them to provide indications of the strategy applied to reach the answer. This is a deliberate choice of the developers, because they want their tool to be useful for the solution method preferred by the teacher (and previous teachers in primary education). In the explanation screens, the most common didactical approach is followed so as to maximize the match between the solution strategy offered at school and by the tool. To facilitate calculations, the tool offers a calculator, which is presented in three forms adapted to the level of difficulty of the exercises.

With regard to the theoretical notions that we developed we first want to reiterate that the Mousework tool differs from the ITS evaluated earlier (Steenbergen-Hu & Cooper, 2013) in various ways: The tool is offered as optional homework, is not used during class, and is meant for all students (not restricted to remediation). Yet in other ways it warrants expectations similar to those voiced

<sup>4</sup> A student questionnaire from the spring of 2013 shows that of the students having access to the tool (the experiment group), only five students did not have a computer at home to practice with. However, IP address data shows that these students have practiced with the tool at school after class hours, where computers were available for free.

<sup>5</sup> Note that we explain here the way the tool is used at Dendron College. Various other modi operandi are available (e.g., supervised use at school and nonrestricted practicing, i.e., open access to all modules).

<sup>6</sup> A set contains between eight and 20 exercises, with 12 as the mode.

<sup>7</sup> Note that the domains are officially called "numbers and operations," "ratio and proportion," or "measurement and geometry," and the Mousework program has sets and modules accordingly. For reasons of brevity, in this article we shorten these to "numbers," "proportions," and "measurement."

Table 1  
*Summary Statistics of Final Sample (N = 344) of First-Year Dendron College Students in the School Year 2012–2013*

	Observations	Mean	SD	Minimum	Maximum
Age (in completed years)	344	12.29	0.48	11	14
Female student	344	0.55	0.50	0	1
Students diagnosed with dyslexia	344	0.06	0.24	0	1
Oldest child	344	0.58	0.49	0	1
Both parents at home	344	0.84	0.36	0	1
Primary school exit exam (Cito): Math	329	57.35	24.08	4	100
Primary school exit exam (Cito): Language (Dutch)	328	56.63	24.49	3	100
Primary school exit exam (Cito): Study ability	329	62.24	25.00	2	100
Primary school exit exam (Cito): Total score	335	538.40	6.20	517	550

in earlier research, and thus allows for its empirical analysis in this particular setting. Various aspects of the use and effect of the tool can be linked into existing theory.

The frequency of intermediate testing at school, for a start, can be expected to have a positive effect on the subsequent use of the tool, because it functions as a signal of the value the teacher (and the school) assign to the tool. In general, we have already noted that we expect students not to value the tool highly because it aims at a faraway goal (an exam at the end of secondary school) and is not graded. Yet we observe variation in the way teachers follow-up on the tool at school. Some teachers organize intermediate tests at regular intervals, whereas others rarely do so or do not at all. Moreover, we expect students to value the tool more when they have a teacher with a positive attitude toward the tool, a second indicator of teacher involvement that we observe. Because Patall et al. (2010) suggest that both a low and high level of control may be counterproductive, we will test the moderating influence of the two teacher variables in a nonlinear way. It may indeed be that the most highly involved teachers are perceived as overly controlling (i.e., situated beyond the tipping point of optimal involvement).

Furthermore, we expect indicators of (math) ability to influence the behavior of the students. We use the general ability test taken at the end of primary school (Cito test) as an (imperfect)<sup>8</sup> indicator of the level of development of the student's metacognitive skills, and thus expect it to relate negatively to the use of the tool (Steenbergen-Hu & Cooper, 2013). Moreover, the starting test of the tool (taken in Week 37) functions as a specific signal to students about their math ability, which we expect to shape their expectancy about their potential in the tool. Students with relatively low results may be particularly demotivated to start using the tool at all.

However, the coupling of expectancy and value changes once students start using the tool. Positive experiences when using the tool (e.g., a relatively high number of sets of exercises finished) and good results on the intermediate tests (e.g., being offered new modules) are likely to motivate the student to continue using the tool. In other words, they raise the expectancy of the student. In contrast, continued negative experiences (e.g., doing poorly on the tests in relation to other students) can be expected to lower expectancy. Lower expectancy may lead to two reactions: students stop using the tool at all, or they focus on sets of exercises they know they have already mastered. The tool offers students various ways to act on the latter. They may repeat sets of exercises for as long as they have not taken an intermediate test at school (the latter is likely to remove the set from their menu of choices if they have

already mastered the associated level of exercises), and they may choose modules they perceive as easier. In this way, we expect the numbers module to be more attractive than the measurement module, and especially the proportions module, because the latter represent math skills that are developed later in primary school, whereas the numbers domain refers to basic numeracy skills acquired early on (Cowan & Powell, 2014; Geary, 2011).

However, the discussion about the likelihood of use does not imply that the tool in itself is ineffective. Although not all students may perceive the tool as valuable and/or some expect it to supersede their abilities, the tool is an up-to-date ITS, offering an adaptive, scaffolded, and interactive learning environment that is likely to enhance learning when used effectively, much as other ITS and CAI tools in the past (Coddling et al., 2007; Gatti & Petrochenkov, 2009; Li & Ma, 2010; McComb & Scott-Little, 2003; Steenbergen-Hu & Cooper, 2013).<sup>9</sup> Consequently, we will study both the likelihood of use of the tool and the learning effect among the students who decided to use the tool.

## The Setup of the Experiment

Figure 1 shows the timeline of the field experiment, which consisted of a pre-experiment period and the experiment period itself. In spring of their final year in primary education (sixth grade in the United States), students registered at their school of choice for secondary education. The secondary school used 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 (age equivalent to seventh grade in the United States) 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 two friends with whom to be placed in the same class. In the summer of 2012, Week 29, classes were randomly assigned to treatment and control groups. The experiment contains only two types of first-year classes (five prevocational classes and eight

<sup>8</sup> The Cito test is a general ability test with focus on math and reading, but as it is intended to predict a student's probability of study success in the different tracks in Dutch secondary education, it also links into metacognitive abilities.

<sup>9</sup> This is not to say that all evaluations of the past have reported positive effects; see, for example, Dynarski et al. (2007) and Campuzano, Dynarski, Agodini, and Rall (2009). A recent meta-analysis on supplementary CAI is the contribution of Cheung and Slavin (2013).

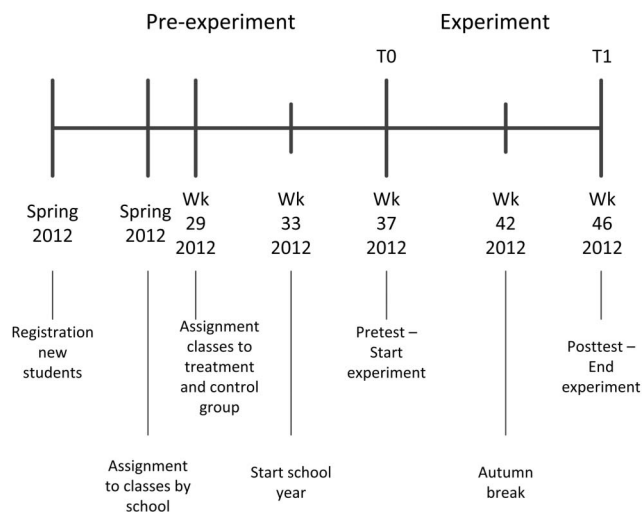


Figure 1. Time line and activities of the field experiment.

higher general/preuniversity classes). Two classes of each type were assigned to the control group (106 students), whereas the other nine classes formed the treatment group (230 students). In Week 33, 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 evening 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 and was completed by all students, regardless of whether they were part of the treatment group or control group. The experiment lasted 8 weeks and the posttest took place in Week 46, again for all children.

Only children that were assigned to the treatment group were granted access to the tool in the following 8. Specifically, they received a personal account with login information so that they had the opportunity to use the tool at home. Students were supposed to practice 30 min per week in the tool, at home. Math teachers were supposed to motivate students to practice with Mousework at home, and were supposed to check students' practicing behavior. A few teachers did use the possibility given by the tool to check on students practice behavior, but most did not.

The school used the interactive digital practice tool in addition to their math classes. All students, both in the treatment and control classes, were being taught mathematics using the math method that has been used over the previous years. For first-year secondary students (seventh grade) there are four math classes of 50 min each week, and students make an average of between 45 and 60 min of math homework per week.<sup>10</sup>

Control-group students did not use the tool. They had only their weekly math classes and their regular pen-and-paper homework.<sup>11</sup> During the course of the experiment, no other math-related activities were organized for control students. Furthermore, to avoid that control students would seek out-of-school compensation for not being allowed to use the tool, control students were promised

(and given) additional practice classes in the next semester, which were explicitly announced to compensate the loss of practice possibilities during the experiment.

In sum, we evaluated the effectiveness of an ITS tool offered to students as a supplementary tool that they can autonomously decide to make use of, but which is recommended to be used for 30 min a week and which can (but most often is not) be checked by their math teacher. Students use the tool at home, but take intermediate tests at school that redirect the content of the tool (at irregular intervals, to which we will return when we explain the operationalization of the practice data).

## Data Operationalization

### Operationalizing Performance

Several outcome measures are used in this study. In the analyses on the differential effectiveness of the tool, students' test-score growth between pretest and posttest for the overall test, as well as for the individual domains numbers, proportions, and measurement, is used as outcome measure. The pretest and the posttest, which is administered several weeks later, do not include any of the exercises treated students have worked on during the intervention period. In addition, all students have worked on exercises in the tool environment a week prior to the pretest. This is done to ensure that all students understand how the tool works and to minimize the possibility that any difference in performance between the treatment and control groups is attributable to the test administration procedure.

The pretest and posttest are standardized validated tests developed by the company of the tool, and these tests are based on other nationally validated tests. The reliability (Cronbach's alpha = .79 to .92) and validity of these tests is analyzed yearly by the tool developer, based on norm data of several participating schools (Schijf & Schijf, 2014). Although the pre- and posttest are digital tests that are developed by the same company as the tool, and are administered in the same digital environment as the tool, the tests themselves are external to the practice exercise tool and do not contain any of exercise questions.

The pre- and posttests contain 35 questions, both open and multiple choice, and the score is recalculated (by the tools' devel-

<sup>10</sup> Information on homework was gathered via a student questionnaire in which a multiple-choice question was used to ask students how much time they spend on homework for math, excluding the time they spend on "Mousework."

<sup>11</sup> Note that although, technically, the difference between treatment and control could also be defined as time spent on practicing versus no time spent on practicing, or as completing small intermediate tests versus not completing those tests, we are confident that the identified effect in this article is not purely because of practicing or testing, as is explained in Haelermans and Ghysels (2015). In this article, two experimental periods were compared, the one that is used in this article is compared with another practice period in which all exercises were available to all treatment students. In the latter case, the adaptation of the program to the current level of the student for each of the domains was completely nonexistent. The results show that there is no effect at all in this second experimental period, although students had practiced, on average, more minutes per week and were tested more often than in the first experiment. This implies that the effect is not simply because of additional practicing or testing, but merely to the adaptive and structured nature of the ICT-homework tool.



oper) to a score between 0 and 100. Table 2 shows the descriptive statistics of the pre- and posttest scores in total and per domain. Note that none of the pretest scores are significantly different between treated and control students. Correlation analyses among the pre- and posttest, as well as the correlations with the Cito math test (taken in sixth grade) reveal that pre- and posttest are moderately correlated ( $r$  between 0.40 and 0.68;  $p < .000$ ), and both are also moderately correlated with Cito math scores (see Table A1 of the online supplemental materials).

Table 2 also shows the descriptive statistics for the growth measures of both the overall score and the separate domains. In this article, we focus on growth as the outcome measure rather than the attainment of the national performance reference for a couple of reasons. First of all, although the pretest score of the treatment group is not significantly different from the pretest score of the control group for any of the outcome measures ( $t$  tests), the individual differences are still substantial, and in many cases the control group students scored higher on the pretest than the treatment group. To directly control for these differences in starting level, we use the growth score as output in our analysis.

Furthermore, we are concerned that an 8-week intervention period (which is the episode of Mousework use that we study here) is not sufficiently long to close the gap between low-achieving students' test scores and the national performance standard criterion, which might wrongfully lead to the conclusion that the tool is not effective in aiding students to achieve the required national performance standards. Lastly, we also wanted to investigate whether the tool has a positive effect on lifting the mathematics skills level of non-low-achieving students to a higher level, even if they have already passed the minimum skills level defined by the national performance standards. The advantage of mathematics skills growth as an index of mathematics skills development is that it is more likely to already be observed over a short period of time, and it shows whether students close the gap between their actual mathematics skills level and the national performance standards.

We are aware that the substantive choice for a growth, rather than an absolute, outcome measure is not without drawbacks (e.g., issues with statistical power). When presenting the results of the analyses, we will come back to several of the issues raised in earlier research to make sure that these do not interfere with the conclusions of our analyses.

### Operationalizing Prior Achievement Level

One of the hypotheses of the present study is that students from different achievement levels have a different practice behavior (Hypothesis 2), and another hypothesis (Hypothesis 4) is that the tool might have a different effect on students' of different achievement levels. In addition, an analysis of the effect of the intervention separately for these three groups shows that the regression slopes are significantly different for the three groups, pointing toward a violation of the homogeneity of regression slopes. Therefore, achievement-level subgroups are operationalized using the final sample of children participating in the experiment. First, three overall achievement-level subgroups are created based on students' overall mathematics pretest scores, which are the collapsed scores of their achievement on the three subdomain exercises. The sample is cut into subgroups so that the overall low-achieving mathematics students group includes the approximately 33% low-

est performing students. The overall high-achieving mathematics students group includes the approximately 33% highest performing students, and the overall middle-achieving mathematics students group includes the remaining students. The reason to choose the 33% lowest and highest, and not, for example, the 16th and 84th percentile, as cutoff points (see, e.g., Wanlass, 2012) is that we have a relatively small amount of children, and only few observations would make up the bottom and top 16%.<sup>12</sup> To avoid that we have too few children in a group (especially because we also split up those groups in treated and control students), we chose for the relatively standard division into three groups. Note that the groups are not exactly equal in size, because we used exact scores as cutoff points instead of ranking the students and using exactly one third as the lowest group, exactly one third as the middle group, and the remaining one third as the highest group (for the exact number of students per group, see Table A2 of the online supplemental materials).

Besides three overall mathematics achievement subgroups, three achievement-level subgroups are created for each of the three domains separately, because we gather from the literature that skills need not be distributed homogeneously across the domains (see literature section), and indeed observe in our data no more than medium correlations between the subscores ( $r$  ranges between 0.24 and 0.27; see Table A3 in the online supplemental materials). The same division guidelines are used as described in the previous paragraph, except that the groups are not defined based on students overall mathematics pretest scores, but on their aggregated pretest score on the relevant domain. As a result, in this study, 12 different mathematics-skills-level groups are distinguished (see Table A2 from the online supplemental materials), namely, three levels (low, middle, and high) per domain (numbers, proportions, and measurement) and overall (composite score of skills).

### Operationalizing Practice Data From the ITS

As explained in the description of the experiment, students are supposed to practice at home with the tool, and are supposed to regularly write small tests at school. The tool generates five indicators of student activity, separately for the practice moments at home and for the small tests that are taken at school, and separately for each domain. The tool registers (a) how many seconds the student has worked in the tool, (b) how many exercises/tests a student has completed, (c) what the score was on the exercises/tests, (d) how many explanation screens the student asked for during the exercise/test (note that during the test, explanation screens were not on content but on the instruction of the test), and (e) the time wasted, which counts the time that the student is logged onto the tool but no activity takes place, with the counting starting after 3 min of nonactivity by the student.

These five indicators possibly correlate with each other, and might give different signs with respect to the student. The number of exercises and the total time spent in the tool (note that this time measure does not include the time wasted) are both a measure of how much has been practiced, and most likely correlate with one

<sup>12</sup> However, note that results are very similar if we choose other cutoff points, such as the 16th and 84th percentiles, thereby making our results not dependent on the chosen cutoff points. Detailed estimates with other boundaries are available on request.

Table 2  
*Summary Statistics of Final Sample (N = 344) Outcome Measures*

	Observations	Mean	SD	Minimum	Maximum
Overall math score pretest	343	55.27	13.42	22	92
Overall math score posttest	336	61.02	14.22	15	90
Numbers score pretest	343	71.27	14.08	25	100
Numbers score posttest	336	72.30	13.56	18	100
Proportions score pretest	343	47.37	23.13	11	100
Proportions score posttest	336	56.72	24.75	11	100
Measurement score pretest	343	52.61	24.29	11	100
Measurement score posttest	336	55.47	23.56	11	100
Growth overall math score	336	5.71	11.65	-31	34
Growth numbers score	336	0.93	14.43	-53	53
Growth proportion score	336	9.40	23.95	-63	89
Growth measurement score	336	2.99	21.99	-73	74

another. However, it is likely that the score attained on the exercises also correlates with the amount of time spent in the tool, as students may have done the same exercise multiple times (especially when the score was low the first few times) or students may have practiced very similar exercises. Furthermore, the score could also be an indication of the dedication of the student. Students that take the exercises seriously most likely get a higher score because they think more thoroughly before they answer. Students that do not take practice seriously may give an answer too fast, and thereby make more mistakes. However, of course student ability is also reflected in the score, making it a complicated indicator. The same holds for the use of explanation screens. It is likely that the high-achieving student does not need the explanation screens, and therefore does not ask for them during the exercise, whereas the less-high-achieving student might need the screens and asks for them more often. However, the student that does not take the exercises seriously will most likely hardly ever, or never, ask for the explanation screens, making this a complicated indicator as well. Lastly, the time wasted is a more clear indicator, as it filters students who are doing many things at the same time and who do not fully focus on the tool (as they are doing something else for 3 min plus the recorded time wasted) from the students that seriously focus on the tool when they are working in it.

As mentioned before, for the differentiation element of the tool to be fully operative, students have to write small intermediate tests in school at regular intervals. There are two factors that determine whether students take these tests and how many. First, the teacher has to take the class to a computer room to enable students to take the test. Second, students need to have practiced as well, in order to have a test available to take. Whether a student has completed any small tests is therefore dependent on factors that the student can influence (practice), and also by factors the student cannot influence (teacher organizing a computer class).

Table 3 presents the descriptive statistics of the five practice indicators, separately for practice time and test time, and separately for the three achievement groups and the three domains. Table 3 shows the number and share of students that have practiced in the tool and completed tests (note that it is possible that more students have completed a test in a certain domain than have practiced in the tool, because the small tests are across domains). Especially for the domains proportions and measurement, only few

students actually practiced. Overall, the students from the middle-skills-level group practiced the least. However, the practice data for the domains separately show that the middle-group students practice the least in the measurement domain (although the difference in the share of students who have practiced is almost zero with the low-skills students), and that low-skills students practice the least in the proportions domain, and high-skills students practice the least in the numbers domain. Furthermore, Table 3 shows that practice and test behavior is different among skills-level groups and domains, but also within skills-level groups and domains, there are large differences between students. In addition, Table 3 shows that the number of students for the proportions and measurement domains are very different than for the numbers domain. This implies that only a select group of students has practiced in the latter two domains. With respect to Hypotheses 2 and 4, based on Table 3, we can conclude that students indeed practice more with the easier modules (numbers domain), and that low-skills students practice the least, for the proportions and measurement domains.<sup>13</sup> However, the fact that low-skills students have practiced more than other groups in the numbers domain both shows the exception for Hypothesis 2 (as low-skills students practice most) but also proves Hypothesis 3 correct (students practice more with easier modules).

Additional analysis on the reasons why students might practice more or less shows that there are significant nonlinear<sup>14</sup> correlations between whether a student has practiced and the attitude of the teacher ( $r = -0.35, p = .00$ ), the teachers' use of the tool in checking on students' practice behavior ( $r = .51, p = .00$ ), and the number of tests taken at school ( $r = .23, p = .00$ ). These findings both confirm and contradict Hypothesis 1. A teacher that uses the tool more, and a teacher that administers more small tests at school, indeed has students that practice more. However, a teacher

<sup>13</sup> When we test Hypothesis III with continuous initial skill indicators, like the Cito test score or the score on the pretest, similar results are found. The correlations of these indicators with the intensity of practice are, respectively,  $r = -0.23, p = .00$  (Cito score), and  $r = -0.16, p = .00$  (pretest).

<sup>14</sup> As discussed because Patall et al. (2010) suggest that both a low and high level of control may be counterproductive, we will test the moderating influence of the two teacher variables in a nonlinear way, by taking the logarithm of the variables.

Table 3

*Descriptive Statistics Activity in the Tool Split for Achievement Level (Overall and per Domain) for Students Who Have Finished at Least One Exercise*

	Low skills level						Middle skills level						High skills level					
	<i>n</i>	%	Mean	<i>SD</i>	Min	Max	<i>n</i>	%	Mean	<i>SD</i>	Min	Max	<i>n</i>	%	Mean	<i>SD</i>	Min	Max
Overall math																		
Practice in the tool																		
Total time in min	79	88	90.68	50.22	1.78	235.80	54	75	83.85	56.30	5.25	241.37	59	82	73.40	46.89	2.30	158.30
Average score	79	88	84.73	8.53	54.19	100.00	54	75	87.63	6.83	65.00	96.67	59	82	88.98	8.08	64.33	100.00
Explanation screen time in min	79	88	4.48	4.42	0.03	22.33	54	75	4.78	5.82	0.07	34.83	59	82	3.46	3.99	0.12	25.73
Time wasted in minutes	79	88	2.90	5.73	0.00	29.45	54	75	2.97	6.45	0.00	26.58	59	82	4.98	15.32	0.00	105.30
Number of exercises	79	88	19.68	12.07	1.00	60.00	54	75	18.17	11.04	1.00	47.00	59	82	16.81	11.34	1.00	50.00
Tests at school																		
Total time in minutes	83	92	25.09	17.51	1.60	68.95	56	78	25.93	14.34	3.32	65.10	64	89	26.17	17.66	2.30	77.37
Average score	83	92	75.94	12.92	32.00	96.00	56	78	81.84	11.82	48.80	100.00	64	89	83.95	10.48	55.50	100.00
Explanation screen time in min	83	92	0.60	0.67	0.03	4.60	56	78	0.82	0.88	0.03	4.78	64	89	0.60	0.48	0.03	2.12
Time wasted in minutes	83	92	0.47	1.22	0.00	6.25	56	78	0.26	1.27	0.00	9.33	64	89	0.15	0.51	0.00	2.83
Number of exercises	83	92	3.45	2.30	1.00	13.00	56	78	3.91	2.16	1.00	10.00	64	89	3.88	2.46	1.00	10.00
Domain numbers																		
Practice in the tool																		
Total time in minutes	78	86	66.17	47.13	1.78	235.80	49	78	47.79	39.52	2.30	179.32	56	70	60.56	40.70	1.57	156.32
Average score	78	86	87.21	8.25	54.00	100.00	49	78	90.06	6.64	71.00	100.00	56	70	93.20	4.50	77.00	100.00
Explanation screen time in min	78	86	2.84	3.14	0.03	14.83	49	78	2.13	2.99	0.05	19.02	56	70	3.84	5.10	0.05	27.33
Time wasted in minutes	78	86	1.97	4.58	0.00	24.85	49	78	1.90	5.70	0.00	32.83	56	70	3.94	14.22	0.00	96.13
Number of exercises	78	86	13.85	8.94	1.00	45.00	49	78	11.00	9.52	1.00	39.00	56	70	14.95	10.57	1.00	50.00
Tests at school																		
Total time in minutes	80	88	16.95	11.43	2.58	56.05	45	71	14.75	9.82	2.30	56.08	62	78	16.79	9.66	3.62	46.05
Average score	80	88	79.45	10.69	51.00	96.00	45	71	83.20	13.85	32.00	100.00	62	78	86.48	11.98	36.00	100.00
Explanation screen time in min	80	88	0.41	0.42	0.03	2.48	45	71	0.39	0.37	0.03	1.97	62	78	0.59	0.56	0.03	2.40
Time wasted in minutes	80	88	0.38	1.16	0.00	6.25	45	71	0.05	0.22	0.00	1.17	62	78	0.08	0.40	0.00	2.83
Number of exercises	80	88	2.31	1.35	1.00	6.00	45	71	2.44	1.57	1.00	9.00	62	78	2.74	1.56	1.00	7.00
Domain proportions																		
Practice in the tool																		
Total time in minutes	20	24	21.10	22.20	3.42	78.23	20	31	15.08	11.86	2.27	44.62	22	25	21.62	23.27	1.65	84.12
Average score	20	24	74.35	21.40	10.00	95.00	20	31	81.95	10.25	56.00	100.00	22	25	83.73	14.40	33.00	100.00
Explanation screen time in min	20	24	0.74	0.92	0.07	3.18	20	31	0.71	1.51	0.05	6.98	22	25	1.05	2.01	0.03	9.20
Time wasted in minutes	20	24	1.37	4.65	0.00	20.77	20	31	0.37	0.97	0.00	3.33	22	25	0.35	1.13	0.00	4.83
Number of exercises	20	24	3.35	3.47	1.00	13.00	20	31	2.85	1.73	1.00	7.00	22	25	3.14	2.51	1.00	8.00
Tests at school																		
Total time in minutes	24	29	7.48	2.84	3.58	16.08	19	30	6.59	3.05	0.73	12.15	25	29	8.89	5.65	2.82	29.77
Average score	24	29	72.50	16.38	44.00	100.0	19	30	64.74	19.93	20.00	94.00	25	29	79.52	10.81	53.00	100.00
Explanation screen time in min	24	29	0.15	0.09	0.03	0.33	19	30	0.14	0.12	0.03	0.40	25	29	0.15	0.14	0.05	0.53
Time wasted in minutes	24	29	0.08	0.25	0.00	0.93	19	30	0.00	0.00	0.00	0.00	25	29	0.12	0.40	0.00	1.85
Number of exercises	24	29	1.00	0.00	1.00	1.00	19	30	1.00	0.00	1.00	1.00	25	29	1.12	0.44	1.00	3.00
Domain measurement																		
Practice in the tool																		
Total time in minutes	21	25	34.91	40.38	1.08	161.7	12	24	35.18	37.95	2.08	129.08	25	25	17.62	15.56	1.43	53.17
Average score	21	25	65.19	19.96	0.00	93.00	12	24	71.50	17.77	25.00	89.00	25	25	74.96	15.10	33.00	100.00
Explanation screen time in min	21	25	1.61	2.23	0.03	8.95	12	24	2.84	4.60	0.17	15.80	25	25	1.00	1.34	0.03	5.27
Time wasted in minutes	21	25	2.91	6.54	0.00	27.28	12	24	2.77	4.68	0.00	12.82	25	25	0.55	1.88	0.00	9.17
Number of exercises	21	25	6.62	7.03	1.00	25.00	12	24	6.83	5.47	1.00	19.00	25	25	4.00	4.34	1.00	15.00
Tests at school																		
Total time in minutes	18	21	11.66	5.36	0.87	18.65	10	20	10.82	5.99	1.45	18.32	17	17	10.66	4.96	3.93	21.25
Average score	18	21	49.11	15.57	8.00	72.00	10	20	47.60	18.30	12.00	80.00	17	17	54.41	15.06	16.00	72.00
Explanation screen time in min	18	21	0.12	0.09	0.03	0.35	10	20	0.09	0.06	0.03	0.20	17	17	0.18	0.30	0.02	1.25
Time wasted in minutes	18	21	0.19	0.47	0.00	1.62	10	20	0.20	0.62	0.00	1.97	17	17	0.57	2.26	0.00	9.33
Number of exercises	18	21	1.00	0.00	1.00	1.00	10	20	1.00	0.00	1.00	1.00	17	17	1.00	0.00	1.00	1.00

that has a more positive attitude about the tool has students that practice significantly less. The latter may indicate that it may not be sufficient to have a positive attitude, but an active interest in the tool may be required to motivate and advise students on the use of the tool. Further research is required, however, to corroborate this idea. The full results of these additional analyses can be found in Table A4 of the online supplemental materials.

## Method

### Identification Strategy

The unbiased estimation of a treatment effect is a serious challenge, which many studies that investigate the effect of a tool on individual's achievement growth face. An unbiased estimate is

expected when students in the treatment and control groups are equal in expectation regarding the effect of the treatment. In practice, this equality in expectation derives from a research design that can reasonably be argued to produce equality in observable and unobservable student characteristics between the treated and nontreated student group. Two major challenges to causal inferences in studies can be identified. First, there is the possibility that schools preferably assign weaker students to the treatment. A second threat could be that particularly motivated teachers who are more willing to use the tool are assigned to the treatment group (Barrow, Markman, & Rouse, 2009). Both situations lead to an overestimation of the effect. To avoid these sources of bias, an experimental design is applied, in which classes are randomly divided over treatment and control group, while taking into account the track of the class.

Because the primary aim of the present article is the estimation of the variation in treatment effect for students' who exhibit different mathematics-skills starting levels, we first check the comparability of the treatment and control group for each defined mathematics-achievement-level variable separately. This implies that we compare the treatment and control groups for 12 different scenarios, namely, three performance groups multiplied with the overall math score and the three domains. The results of the *t* tests and Mann-Whitney analyses are presented in Table 4.

Although the groups are comparable on most observable characteristics, some groups are significantly different with respect to the variables oldest child in family, primary school advice, and primary school exit exam (Cito): math and total score. Being the oldest child in the family is included as a control variable in the regression analyses later on. Because of multicollinearity issues between the variables mathematics achievement level and primary school exit exam (Cito): math (which in turn is correlated with

primary school advice), we have decided not to include the latter two variables as control variables. We do not expect this to be a problem because the compared experimental and treatment groups score equally on the pretest, and this later measure of mathematics skills level has tested students on exactly the same math domains and content, which is not the case for the Cito test. Therefore, the pretest seems a better predictor.

### Estimation Strategy

The intention-to-treat effect (ITT) of having the opportunity to practice with the Mousework program on basic mathematics skills growth in secondary students is estimated using the following equation, first applied to observational data by Rosenbaum and Rubin (1983):

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

where  $d_{ij}$  is the dummy variable expressing the treatment, thus a students' access to the online tool, for a specific student,  $i$ , in class  $j$ , and  $Y_{ij}$  denotes the growth in mathematics skills test scores between pretest and posttest for each student in class  $j$ . The parameters  $y_{ij}(1)$  and  $y_{ij}(0)$  denote the growth in test scores for treated and untreated students in a class, respectively. As will be illustrated later in the article, not all children assigned to the treatment group actually used the tool. For that reason, we estimate the ITT and not the average treatment effect in this study. Moreover, for the policymaker and the school principal, the ITT is the relevant effect measure, as it reflects the effectiveness of the offer of the tool with the imperfections in use and teacher involvement that are typical of real life.

Table 4

Comparability Treatment and Control Group Split for Achievement Level (Overall and per Domain)

	Overall			Domain numbers			Domain proportions			Domain measurement		
	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
	Sample size											
Control group	33	34	43	41	25	44	34	38	38	40	17	53
Treatment group	90	72	72	91	63	80	83	64	87	84	50	100
	<i>t</i> value/ <i>z</i> value											
Variable	-0.16	0.63	-0.36	-1.03	1.50	-0.58	0.53	-0.66	-0.54	0.56	-1.05	-0.30
Age (in completed years)	-0.57	<b>-2.32</b>	-1.60	-1.22	-0.90	-0.81	-1.49	-0.11	-1.16	-1.08	-0.89	-1.15
Primary school exit exam (Cito): Math	-1.26	-0.77	-0.18	-0.60	0.34	-0.58	-0.99	0.40	-0.46	-0.56	-0.13	-0.70
Primary school exit exam (Cito): Language (Dutch)	-0.61	-0.98	0.31	0.18	0.33	-0.80	-0.38	0.62	-0.69	-0.83	-0.56	0.68
Primary school exit exam (Cito): Study ability	-0.73	<b>-2.25</b>	-0.57	-0.59	-0.25	-1.04	-1.24	0.23	-0.79	-0.89	-0.57	-0.75
Primary school exit exam (Cito): Total score	-0.65	-1.22	1.77	-1.13	1.39	-0.21	-1.68	-0.61	1.86	-0.21	0.05	0.23
Female students	0.34	-1.23	1.30	0.40	-1.11	-0.73	1.05	-1.71	-0.24	-0.56	-1.03	0.47
Students diagnosed with dyslexia	<b>-2.29</b>	0.41	0.03	-1.51	-0.51	0.28	<b>-2.84</b>	0.57	0.43	-0.47	1.22	-1.64
Oldest child in his/her household (1 = yes)	1.18	-0.67	0.73	-0.08	-0.40	1.58	1.52	-0.19	-0.34	-0.62	1.21	0.93
Situation at home (1 = both parents home)	-0.40	0.47	-0.06	0.58	0.14	-0.66	-0.39	-0.07	0.38	1.22	-0.48	-0.88
Living area	-0.86	1.46	-0.37	0.11	0.63	-1.35	0.48	-1.30	x	x	-1.61	0.80
Country of birth	-0.86	1.46	-0.77	0.07	-0.63	x	-0.21	x	x	x	-0.58	-0.06
Nationality	0.78	0.88	-0.73	0.33	1.46	-0.48	1.23	-0.87	0.92	1.35	0.22	-0.63
Religion	-1.65	0.48	0.79	-1.95	0.31	1.43	-0.08	-0.27	0.34	-1.77	-0.63	1.94
Primary school	-1.95	-1.69	0.34	<b>-1.97</b>	0.40	0.14	<b>-2.29</b>	1.03	-0.09	-1.43	-1.06	0.19
Primary school advice												

Note. Statistics presented are *t* statistics and *z* statistics. Bold values are significant at the 5% level. x = no difference between groups or no observations.

Given that, in our study, the randomization ensures the independence between the treatment and potential outcomes, the ITT can be derived from Equation 1 as

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

The ITT effect can be estimated using either simple  $t$  statistics or a linear regression, and the interaction effect between the treatment and basic mathematics skills level of students can only be estimated using the latter. Therefore, two linear regression specifications are estimated in this study:

$$Y_i = \alpha_i + \tau_2 d_j + (\varepsilon_i + u_j) \quad (3)$$

and

$$Y_i = \alpha_i + \tau_2 d_j + \beta_{1i} \delta_c + \beta_{2i} d_j * \delta_c + \beta_3 X_i + (\varepsilon_i + u_j) \quad (4)$$

where  $d_j$  is the treatment status of class  $j$ , the parameter  $\varepsilon_i$  are the residuals at the student level and  $u_j$  are the residuals at the class level. Equation 3 is an ordinary least squares (OLS) regression used to estimate the first-order effect of treatment. In Equation 4, the interaction effect between treatment and students pretest mathematics-skills-level subgroup, denoted by  $d_j * \delta_c$ , is included next to multiple other students' observable characteristics,  $X$ , such as gender. Because of multicollinearity, only observable characteristics that are consistently weakly correlated with students' pretest mathematics skills level, captured by a correlation coefficient of 0.3 or lower, are included in the last equation. For that reason, the following variables are not included as covariates: mathematics skills level at pretest, primary school advice, primary school exit exam scores (Cito; math, language, study ability, total score; see correlation in Table of the online supplemental materials).

Lastly, it is important to mention that all estimated equations contain standard errors clustered at the class level, which is a statistical way of correcting for the multiple levels in the experiment (Angrist & Pischke, 2009; Wooldridge, 2010). By doing this, the equations are corrected for internal correlations between variables within a class. This is necessary because classes define common characteristics to all pupils making part of it (such as peer group, teacher, and instructional environment). Classes are also grouped based on ability, but this should not be a problem in our experiment, given that we have two classes of each level (track) in the control group and the remaining classes of the two levels in the treatment group. Nonetheless, any remaining nonrandom differences are corrected for by clustering the standard errors at the class level.

## Results

### Differential Effect of Mathematics Skills Level Overall

The effect of the tool on students' mathematics skill development is first determined for students' overall mathematics skills growth, by regression analysis. Three regression specifications are estimated. In Specification 1, the ITT is estimated with simple OLS with standard errors clustered at class level. Two additional specifications are estimated to shed light on the question whether growth in mathematics skills resulting from the opportunity to use the tool varies between students with different mathematics skills

levels. Also in these specifications, a regression analysis with standard errors clustered at class level is used. In Specification 2, the effect for the low-achieving mathematics students is contrasted with the effect for the other mathematics achievement subgroups (through interaction effects). In Specification 3, the student-specific characteristics gender, age, home situation (a proxy for SES), oldest child in household, dyslexia diagnosis, and religion are added as covariates.

Recall that there are three mathematics achievement subgroups based on students' pretest scores. The low-achieving subgroup, which includes the children performing at the 33rd percentile and below, form the reference group in the analyses displayed in this article. The reason for this is twofold. First, analyzing the effect for low achievers is most interesting because these children exhibit the lowest mathematics skills, and consequently the discrepancy between the skills they have and the skills they should possess, as defined by the national performance standards, is largest. Second, they are expected to have the highest growth rate in response to the intervention. Low-achievers' mathematics skills level is lower than middle achievers or high achievers at the start of the intervention, and therefore they have more room to grow. Moreover, their skills are still rather undeveloped and not automatized, allowing students to still develop different aspects (e.g., understanding, procedure, and automation) of a skill.

The essential results of the regression analyses estimating the effect of the tool on students' overall mathematics skills growth are displayed in Table 5 (the complete results can be found in Table A5 of the online supplemental materials). In this analysis, we test the first part of Hypothesis 4, in which we state that the effect of the tool differs by prior performance level. The treatment effect is found to be significant in the simple OLS specification (Specification 1). Students who have access to the tool experience, on average, more growth.

Treated students grow, on average, 4 points more than their peers who do not receive the treatment (in which the latter grow, on average, 2 points on the 0-to-100 scale of the test). This effect corresponds with a small standardized effect of 0.18 of a standard deviation. A significant treatment effect is also observed in Specification 2. However, because this specification includes interaction effects (reference group = low-achieving students), the effect of the treatment needs to be interpreted differently. In particular, the treatment effect in Specification 2 captures the effectiveness of the treatment for the low-achieving students (the reference group to the middle- and high-achieving students), but not its effectiveness overall. Therefore, the significant treatment effect in these specifications implies that the treated low-achieving students experience significantly more growth than the low-achieving students in the control group. Specifically, the treated low-achieving students grow about 8 points more than the low achievers who do not have access to the tool, which is a standardized effect of 0.33 of a standard deviation. The effect is robust to the addition of student ability variables and student-specific characteristics (Specification 3).

Moreover, Table 5 shows, for Specifications 2 and 3, that the effect for middle-achieving students does not differ significantly from the effect of the low-achieving students, but that the effect is considerably smaller for high-achieving students. Supplementary regression analyses using middle-achieving and high-achieving students as a reference group have been conducted as well (the

Table 5

*The Effect of Having Access to the Tool and Its Interaction With Mathematics Skills Level Overall: Regression Analyses With Standard Errors Clustered at Class Level*

Outcome variable = Absolute growth in test score T0–T1	Specification 1	Standardized coefficient	Specification 2	Standardized coefficient	Specification 3	Standardized coefficient
Treatment						
Coefficient	4.46**	0.18	8.18***	0.33	8.49***	0.34
Standard error	–1.67		2.72		2.83	
Treatment × Mathematics Skills Level (Middle)						
Coefficient			–3.06	0.11	–3.26	0.12
Standard error			2.05		1.90	
Treatment × Mathematics Skills Level (High)						
Coefficient			–8.93***	0.31	–9.32***	0.33
Standard error			–2.60		–2.72	
Covariates ( $X_{ji}$ )					Student ability; student characteristics	
Observations ( $n$ )	336		336		336	
Groups (classes)	13		13		13	
$R^2$	0.03		0.10		0.11	

\*  $p < .10$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

detailed results can be found in Tables A6 and A7 of the online supplemental materials). In line with the results shown in Table 5, a significant positive treatment effect is observed for middle-achieving students. However, a negative but insignificant effect is found for high achievers, whereby both treated and untreated high achievers experience a negative growth across the 8-week experiment period. The negative sign might be a consequence of the fact that high-achievers test scores are already relatively high at the pretest (about 90%), and consequently they might have not been motivated to work seriously on the posttest. Although students can unlimitedly increase in practice level in the online environment of the tool, the pre- and posttest only measure the level they are supposed to have achieved at the end of sixth grade. Unfortunately, this means that a potential positive effect of the tool for high achievers is not visible in the test score, because students score high on both tests, and might even decrease the test score by a point or so because of sloppiness. Given these issues, we believe that we cannot draw conclusions about the treatment effects for high achievers based on the data used in the present study. For the same reason, one should also be cautious in formulating conclusions based on the reported interaction effects of the high achievers in Table 5 and Table 6.

To conclude, the results indicate that the positive, overall effect of treatment highlighted by Specification 1 is attributable to the positive effect for the low-achieving students and middle-achieving students. This finding supports Hypothesis 4, namely, that there may exist differential effects between different achievement groups. Yet it also raises an interesting (and well-known) point regarding test construction in the light of national reference levels. It may well be that the norms of the test (linked to the national reference levels) were inappropriate for the high-performing students. We previously showed that they did not practice less than other students, which was in line with our expectations regarding the value and expectancy that high-performing students attribute to the ITS. However, the test did not allow them to elicit skills growth, which means that practicing with the ITS may or may not have been effective for this group, a result we will elaborate on in the robustness analyses.

### Effect of Treatment Across Achievement Subgroups per Domain

In this subsection, the impact of the tool on students' mathematics skills growth across achievement subgroups is reported per domain. In this part of the analysis, we test the second part of Hypothesis 4, in which we assume that the effect of the tool differs by math domain. We estimate the same three specifications as before, but instead of using overall growth as the outcome measure, the growth children exhibited on the individual domains (numbers, proportions, or measurement) is included as an outcome measure. Like before, the low-achieving students form the reference group. Yet remember that for the domain-specific analyses, the variable mathematics skills level (low, middle, and high) has been operationalized by dividing the student sample into low-achieving students, middle-achieving students, and high-achieving students based on their pretest score on the respective domain.

The domain-specific results are presented in Table 6 (the detailed results can be found in Tables A8 through A10 of the online supplemental materials). The main effect of the treatment for all treated students, which is estimated with Specification 1, is only significant for the proportions domain. Specification 2 shows that this is different for low-achieving students. They experience significantly more growth than their middle-achieving peers in all three of the domains as well. The treatment coefficients imply that they exhibit about 7 points (numbers), 22 points (proportions), and 14 points (measurement) more growth than the low-achieving students in the control group. This corresponds to standardized effect sizes of 0.25, 0.44, and 0.31, respectively.

Examining the interaction effect patterns in Table 6 shows that the treatment effect for middle-achieving students is consistently lower than for low-achieving students, and even negative for the numbers domain. All of the interaction effects are robust to the addition of covariates (Specification 3). This implies that, in general, middle-achieving students benefit significantly less from having access to the tool than low-achieving students for two of the three domains assessed, although still positive effects are found, especially for proportions. This finding also confirms Hypothesis

Table 6

*Per Domain, the Effect of Having Access to the Tool and Its Interaction With Mathematics Skills Level: Regression Analyses With Standard Errors Clustered at Class Level*

Outcome variable = Absolute growth in test score T0–T1	Specification 1	Standardized coefficient	Specification 2	Standardized coefficient	Specification 3	Standardized coefficient
<b>Domain numbers</b>						
Treatment						
Coefficient	0.55	0.02	7.84***	0.25	8.30***	0.26
Standard error	2.36		3.07		3.05	
Treatment × Mathematics Skills Level (Middle)						
Coefficient			−8.95***	0.01	−9.29***	0.02
Standard error			1.61		1.57	
Treatment × Mathematics Skills Level (High)						
Coefficient			−14.04***	0.16	−14.58***	0.17
Standard error			2.40		2.21	
Covariates (X <sub>ji</sub> )					Student ability; student characteristics	
Observations ( <i>n</i> )	336		336		336	
Groups (classes)	13		13		13	
R <sup>2</sup>	0.00		0.12		0.14	
<b>Domain proportions</b>						
Treatment						
Coefficient	10.57**	0.20	22.81***	0.44	23.42***	0.45
Standard error	4.36		4.45		4.69	
Treatment × Mathematics Skills Level (Middle)						
Coefficient			−16.08***	0.18	−16.44***	0.18
Standard error			2.64		2.62	
Treatment × Mathematics Skills Level (High)						
Coefficient			−21.01***	0.07	−21.85***	0.06
Standard error			2.45		2.42	
Covariates (X <sub>ji</sub> )					Student ability; student characteristics	
Observations ( <i>n</i> )	336		336		336	
Groups (classes)	13		13		13	
R <sup>2</sup>	0.04		0.14		0.15	
<b>Domain measurement</b>						
Treatment						
Coefficient	4.12	0.09	14.72***	0.31	15.11***	0.32
Standard error	3.46		3.44		3.83	
Treatment × Mathematics Skills Level (Middle)						
Coefficient			−7.71**	0.19	−7.58**	0.20
Standard error			3.41		3.41	
Treatment × Mathematics Skills Level (High)						
Coefficient			−20.94***	0.16	−22.01***	0.14
Standard error			2.32		2.15	
Covariates (X <sub>ji</sub> )					Student ability; student characteristics	
Observations ( <i>n</i> )	336		336		336	
Groups (classes)	13		13		13	
R <sup>2</sup>	0.01		0.13		0.15	

\*  $p < .10$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

4, namely, that there are differential effects between mathematics domains. As an explanation for this, note that low and middle achievers were struggling more with proportions in the pretest than (a) with the other domains, and (b) than high-achieving students, who struggled more with measurement. High achieving students also benefit from access to the tool in the proportions domain, but show a negative effect in the other two domains. Potential reasons for this negative effect have already been given Remember that it is students' access to the tool which is the intervention in our study. However, not all students having access actually practiced,

a point we will return to in the section on additional analyses on students' practice behavior.

### Robustness Check

Given that we decided to set the cutoff points at the 33rd and 67th percentiles to create groups of roughly equal size, we should also check whether an approach more common in the literature would lead to similar results. This would mean using the 16th and 84th percentiles, as discussed in the section where we operation-

alized the prior achievement level. However, the risk of this approach is that the low- and high-achievers subgroups are relatively small as a result of the operationalization criteria used, potentially leading to a failure to reject the false null hypothesis (lack of statistical power). However, the actual estimates show that this is not the case.

The regression analyses of overall performance, with standard errors clustered at class, similar to the analysis in Table 5, largely confirm the results that have been found when including three mathematics-skills-level subgroups that are roughly equal in size (the detailed results can be found in Table A11 of the online supplemental materials).

A second robustness analysis that we perform is an analysis for which we do not use the growth score as outcome measure, but the absolute score at Time 1, where we control for the pretest score at Time 0 in the regression analysis. Performing this analysis requires us to model the nonlinear relation of previous performance with the intervention differently than using the three groups with respect to performance level, as pretest and three performance-level groups are highly correlated. However, as we assume that there may exist a nonlinear relationship, we cannot simply include pretest and interact pretest with the treatment. Therefore, we have included both pretest and pretest-squared, as well as the interaction of these two with the treatment (instead of interacting the treatment with the performance group dummy variables, as we have done before).

The results of this second robustness analysis produced very similar results to those in Table 5, in the sense that it confirms that pretest performance is linked in a nonlinear way to posttest performance (a positively significant coefficient of the quadratic effect; the detailed results can be found in Table A12 of the online supplemental materials).

### Additional Analyses on Students' Practice Behavior

The results we have studied thus far revealed that granting students access to the supplementary tool can be an effective way to improve (some of) their math skills. Yet it does not tell much about the mechanisms at work. In effect, the literature suggests that differences in practice behavior account for the growth differences between mathematics skills subgroups and individual students (Biagi & Loi, 2013; Spiezia, 2010). To explore this idea, we ran supplementary regression analyses, with growth in score as the outcome measure, in which we include four aspects of practice behavior: whether a student has practiced (at least once in the observation period), how many minutes, how many explanation screens per minute the student asked for, and how many exercises per minute the student completed. Furthermore, we again allow for variation according to the starting level and include interaction terms of these four practice indicators with the middle- and high-achievement-group dummy variables. Finally, student characteristics and teacher dummy variables are added as general control variables.

Note that, compared with our earlier descriptive analyses focused on Table 3, for the regression analyses, we have replaced the practice variables of all students who did not practice (and who were excluded from Table 3) with a "0" because students that are not registered in the tool activity data have not been active in the

tool, which is an indicator in itself. Moreover, we express the use of explanation screens and the number of exercises finished as a proportion of the time spent practicing, to make sure that these variables represent characteristics of the practice behavior apart from the decision to practice and the time spent on it. Finally, also note that we have not incorporated a fifth characteristic of practice behavior (time wasted), because we do not see it to add a clear additional dimension apart from the four characteristics already presented.

The main results of the regression analyses are presented in Table 7 (the detailed results can be found in Table A13 of the online supplemental materials). Note that, similar to before, the coefficient of the interaction with the middle- and high-achievement group should be interpreted as additional to the overall coefficient, which represents the low-achieving group. Also note that we are not comparing treatment with control students anymore, but students who have practiced with students who have not practiced.<sup>15</sup>

Table 7 shows the regression outcomes for the overall measure and each of the three domains separately. The first two columns show the unstandardized coefficients and standard errors in the first column and the standardized coefficient (in terms of the standard deviation of the dependent variable) in the second column. Overall, we see that there is a significant effect (of 0.64 standard deviation) of whether a student practices on the absolute growth in the test score for the low-skills group, and this effect is not significantly different for the other two groups. This implies that practicing is useful for all three types of students.

For the numbers domain, Table 7 shows no significant effect of practicing for the low-skills group, and no significant difference from that for the middle-skills group. In other words, practicing with the tool does not help for these groups. This could be attributed to the fact that numbers is a domain that shows up in many other subjects at school, not in the least in the regular math class. Especially at the numbers level that these seventh grade students are working in, it seems that these skills are acquired either way, regardless of additional practice in the tool. On the other hand, we also see that the high-skills group is significantly different from the low- and middle-achieving group, and given the coefficient, the effect of practicing seems negative for this group on the numbers domain. However, given the positive significant coefficient of the number of exercises and the number of explanations screens, the estimates suggest that high-achieving students only benefit from practicing with the tool for numbers when they write a sufficient number of exercises and ask for explanation screens once in a while. This does not need to be contradictory with our earlier conjecture regarding the generally ineffective nature of the use of the tool in the numbers domain. Although low and middle achievers tend to repeat skills already acquired elsewhere or before, it may well be that those high-achieving students who do enjoy growth in the numbers domain make use of the tool

<sup>15</sup> We have used the entire sample of students in the analysis of practice effects, to analyze the causal effect of practicing in the tool. In order to do so, we also have to include the students that (unwillingly) have not practiced. However, if we run the analyses for only the students in the treatment group, who have practiced, the results are very similar and we draw the same conclusions.



**Table 7**  
*Correlates of Growth in Math Skills Level and Practice Behavior*

Outcome variable = Absolute growth in test score T0-T1	Overall	Standardized coefficient	Numbers	Standardized coefficient	Proportions	Standardized coefficient	Measurement	Standardized coefficient
Practiced dummy variable	15.11**	0.64	9.95*	0.34	27.42**	0.44	12.21*	0.21
Coefficient	4.46		5.46		9.33		5.96	
Standard error								
Practiced Dummy variable × Mathematics								
Skills Level (Middle)								
Coefficient	-5.67	0.47	-9.55	0.11	-40.0**	0.05	-0.05	0.21
Standard error	7.14		6.22		11.03		7.14	
Practiced Dummy variable × Mathematics								
Skills Level (High)								
Coefficient	-17.10*	0.09	-32.8***	0.50	-38.13**	0.05	-13.26	0.05
Standard error	8.40		6.00		15.32		11.84	
Minutes practiced								
Coefficient	-0.03	0.13	-0.05	0.15	-0.08	0.04	-0.06	0.04
Standard error	0.03		0.06		0.07		0.08	
Minutes Practiced × Mathematics Skills								
Level (Middle)								
Coefficient	0.00	0.12	0.04	0.14	0.76***	0.18	0.08	0.01
Standard error	0.03		0.07		0.12		0.15	
Minutes Practiced × Mathematics Skills								
Level (High)								
Coefficient	-0.03	0.21	0.10	0.05	0.12	0.00	-0.21	0.10
Standard error	0.04		0.07		0.25		0.28	
Explanation screens								
Coefficient	-0.62*	0.16	-0.95	0.26	1.14***	0.08	0.33	0.05
Standard error	0.32		0.58		0.30		1.64	
Explanation Screens × Mathematics Skills								
Level (Middle)								
Coefficient	0.47	0.08	0.44	0.22	-1.61	0.01	0.09	0.06
Standard error	0.54		1.09		1.19		1.60	
Explanation Screens × Mathematics Skills								
Level (High)								
Coefficient	0.58	0.07	1.23***	0.04	-1.67	0.01	-0.24	0.03
Standard error	0.48		0.58		1.17		1.70	
Nr of exercises								
Coefficient	-19.79	0.24	-0.32	0.00	-22.48	0.10	-8.72**	0.05
Standard error	12.68		9.32		50.00		3.05	
Nr of Exercises × Mathematics Skills								
Level (Middle)								
Coefficient	-3.27	0.27	-22.03	0.15	30.38	0.02	-43.23*	0.15
Standard error	16.91		23.60		59.73		21.28	
Nr of Exercises × Mathematics Skills								
Level (High)								
Coefficient	9.61	0.16	19.84**	0.15	5.47	0.08	-5.40	0.07
Standard error	18.81		7.91		61.62		20.32	
Covariates (Xj)								
Student ability; student characteristics; teacher dummy variables								
Student ability; student characteristics; teacher dummy variables								
Observations (n)	336		336		336		336	
Groups (classes)	13		13		13		13	
R <sup>2</sup>	0.23		0.22		0.18		0.08	

\*  $p < .10$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

for numbers skills that have not yet been covered in class (and hence make good use of the explanation screens).

The analysis for proportions shows a significant and positive effect of practicing for the low-skills group of students. However, the effect of middle- and high-skills students is negative, and significantly different from the low-skills group. Low-skills students that use explanation screens while practicing even have a higher absolute growth in test score. Middle-group students that spent enough minutes on the proportions domain do have a positive effect of practicing, although they would have to spend more than 16 min on proportions alone to achieve this effect (the difference in effect with low-achieving students is 40 min 27 s, which is about 12-points difference in score growth; because the coefficient of minutes practice is around .75, you need to practice  $.75 \times 12$ , 16 min, to offset the former negative coefficient). So middle-achieving students that make enough minutes while practicing also benefit from practicing for proportions. For high-achieving students, it does not seem to contribute at all. The latter is in line with the fact that we hardly find an effect for these students (see Table 6) in the first place.

For the measurement domain, we see a significant positive effect for low-achieving students, and no significant difference for the other two groups. This implies that practice is beneficial for all students. But, the negative coefficients for the number of exercises for the low and middle group imply that students should not practice too many exercises for this domain (note from Table 3 that the average is six exercises) or this will offset the positive effect. This can possibly be explained by the fact that this is a rather difficult domain for students (early in seventh grade), and that low- and middle-achieving students that have done many exercises may not have taken them very seriously, by just clicking through the exercises without really trying to answer the questions and learn from the feedback. This would also explain why there is no effect of minutes practiced (additional to the practice dummy variable of course). Note that the explained variation ( $R^2$ ) for the measurement domain is also very low compared with the other domains. This implies that, to a larger extent than for the other domains, there are other important factors at play to explain growth in score, that is, other than practice and student characteristics and teacher dummy variables.

Last, the teacher dummy variables that are included in Table 7 are highly significant, in most cases (see Table A13 of the online supplemental materials for coefficients and significance levels). The latter reiterates the importance of the teacher in the absolute growth in the test score of a student. However, including teacher dummy variables in the regression analysis, compared with not including these dummy variables, does *not* impact the size of the practice coefficients, although in some cases, the significance increases. This implies that the tool, and practicing with the tool, works independently of the contribution of the teacher.

All in all, from the analysis of practice behavior, we conclude that, in most cases, practice is effective to increase the growth in test scores, although these effects differ slightly by domain and skills group, which suggests that future research may want to address a more restrictive orientation of the practice behavior than the one currently implemented in the tool studied. Regardless of this, the analyses also show that the teacher has an important influence on a student's growth in test score.

## Discussion and Conclusion

In this study, we investigated an educational experiment with a unique combination of three elements: homework, the use of ICT, and a large degree of freedom of choice (student autonomy). More particularly, we studied the effectiveness of a web-based intelligent tutoring system that a school offers to its students as optional homework, in a sample of first-year secondary students, using an experimental design. Based on the literature and theory on the use of ICT tools, the function of homework, and the value of freedom of choice in the learning process, we formulated the following four hypotheses:

1. A teacher with a positive attitude and who uses the tool more actively than other teachers (to check on students' practice behavior) will inspire his or her students to practice more (adding to the perceived "value" of homework).
2. Students with a relatively weak initial performance will tend to work less with the tool than the other students (even though they can be expected to benefit more from practice; expectancy trumps rational choice).
3. Students engage disproportionately in the easier modules (i.e., numbers) at the detriment of the more difficult modules (i.e., measurement and proportions), irrespective of their level of expertise (i.e., they do not optimize their exercises portfolio from a learning perspective, and hence Mousework does not offer sufficient choice structure).
4. The effectiveness of the Mousework tool depends on the type of student (with respect to initial performance) and the module in which students have practiced: Practicing helps, and helps students with an initially low performance relatively more, but its effect is limited to the domains practiced.

In testing the first three hypotheses, the results show that whether students make this noncompulsory homework in the tool is dependent on their prior achievement and their teacher, and if they practice, they tend to choose easier modules. All these findings concur with our theoretical framework, which relates homework behavior to the expectancy and value students attribute to the task. Based on our study, we can add to the theory that the Trautwein framework on homework behavior extends to online homework, even if it is framed as optional (i.e., has a lower starting value than is commonly true for homework assignments).

Concurrently, we referred to the role of freedom of choice when optimizing homework effort and completion rates. Previous research stressed that freedom may be beneficial if well-structured. We assumed that the intrinsic structure of an ITS would comply with this requirement, but observe the latter to only be true to some extent. Liberty of choice meant, for example, that students could choose the exercises they felt most sure about, and could repeat these instead of engaging in more demanding exercises. Whether this practice led to suboptimal learning is a matter we cannot assess in this experiment and

leave for future research, but it definitely raises a question about the optimal balance between freedom of choice and optimal “adaptiveness” in ITS tools. As such, we can add to the theory on freedom of choice that structure in itself does not necessarily lead to the correct choices. Future research will have to clarify whether a slightly less free, differentiated practicing structure (e.g., requiring students to practice in all domains) would foster learning even more.

With respect to the fourth hypothesis, we conclude that there is a differential effect both for achievement groups and for the mathematics domains. Although the implementation of a non-compulsory supplementary tool leads to improvements in the overall mathematics skills of both low-achieving and middle-achieving students, it is significantly less effective for middle achievers than for low achievers. This finding is especially true for the analyses split for mathematics skills domains (numbers, proportions, and measurement). We also find that having access to the tool has, on average, a significant positive effect on students’ proportions skills, but not on their numbers and measurement skills. Thus, on average, we observe a differential effect of the tool across distinct mathematics domains. Yet further analyses indicate that this conclusion should be nuanced. Low-achieving students benefit significantly from the intervention in all mathematics domains, whereas middle-achieving students benefit significantly from the intervention in the domains proportions and measurement.

The main implication of our results is that schools may want to target the implementation of a noncompulsory homework tool. Although implementing all parts of the tool is a rational choice for low-achieving students, schools should focus on providing additional exercise opportunity in the proportions domain and measurement to middle-achieving students. Yet we should be careful drawing definitive conclusions. Our study includes only one school, which does mean that we have an external validity issue. In addition, the assessed treatment period is rather short, and therefore we cannot exclude the possibility that the program has a positive effect on high-achieving students’ achievement after a more extended implementation period. Furthermore, the fact that there could be a ceiling effect in the test, but not in the practice environment, means that we also cannot exclude the possibility that there is an effect for high-achieving students that currently remains unobserved. Future research may show that a common targeting practice of schools, limiting ITS to a remediation context, may be misguided and should be replaced by a wider use of ITS, with a tailored choice structure rather than a limited access.

Lastly, we combine the four hypotheses and test whether students’ achievement level is related to both students’ practice behavior and teacher characteristics. The results show that although a vast number of treated students have worked on exercises in the numbers domain, relatively few have practiced (even once) their proportions or measurement skills. Although the teacher significantly influences the growth in test score by the student, and is correlated with practice behavior, we find that practice in the tool *does* matter independently, although some differences are found between domains and achievement groups as well.

All in all, we would suggest interpreting the teacher as a type of external motivation for students, both for practicing in the

tool and for motivation to perform well in general. Students sensitive to this type of motivation will practice more, but are still likely to practice mainly in the numbers domain, because there is a general conception that numbers skills are easier than measurement and proportions skills, given that they are the first type of mathematics skills introduced in primary school (“expectancy”). Supporting this interpretation is our finding that many of the children do practice numbers skills, whereas only few children have accessed exercises falling into the measurement and proportions domains. The children who have decided to practice these types of mathematics skills might be different in, for example, internal motivation than the vast group of children that have practiced only number skills.

The results emphasize that for most students, the mere implementation of a noncompulsory, supplementary ITS tool does not by itself make students practice more. The finding on the importance of the teachers is in line with previous studies, which have concluded that teacher explanations and involvement are indispensable for a positive effect of a supplementary ITS tool on students’ learning (Angrist & Lavy, 2002; Cheung & Slavin, 2012; Koedinger, McLaughlin, & Heffernan, 2010; Kroesbergen & van Luit, 2003; Leh & Jitendra, 2013; Roschelle et al., 2010). We therefore would recommend that schools invest more in stimulating teachers to explicitly motivate their students to use the tool in order to work on their weaknesses, and to check on their students’ use of the tool (both with a relatively modest workload).

As a general conclusion, it can be stated that we observe a differential effect of having access to a noncompulsory supplementary ITS tool across achievement subgroups and mathematics domains. The most promising impact of the tool in the short run is found for low- and middle-achieving students and for the proportions domain. Lastly, the results that are reported in the present article show that the overall treatment effect found in the previous study by Haelermans and Ghysels (2015) is largely driven by the positive effect for low- and middle-achieving students and the growth achieved in the proportions domain.

However, there are a few aspects that one should bear in mind with respect to our study that should be taken into account in future research on this topic. First of all, the experiment only lasted 8 weeks. This is quite short for an experiment like this, and it is possible that the time span of the experiment influenced the results. However, it is most likely that the results are downward biased because of the short time span, because students have had less time to practice. Furthermore, because the experiment took place at the beginning of the school year, it is likely that the first few weeks were more startup weeks, and less time was spent on the experiment. Second, the students were tested at the level they were supposed to have achieved at the end of sixth grade. Therefore, it is very likely that there was a ceiling effect in the test results of the high-achieving students. Although they could practice at their own level, the tests were not able to show whether there is a positive effect of practicing for these students, simply because they already performed very high in the pretest, and therefore literally had nothing to gain in the posttest. Lastly, overall, students have not practiced that much. It is very likely that the effect of more minutes of practice is not linear, but has an optimum at a certain amount of practice time. Unfortunately, in the current study, this could not

be tested because of a lack of variation in the practice data (especially at the upper half of the distribution).

Despite these aspects, the results of this study are very promising, and future research should try to extend these findings.

## References

- Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal*, *112*, 735–765. <http://dx.doi.org/10.1111/1468-0297.00068>
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal Economic Policy*, *1*, 52–74. <http://dx.doi.org/10.1257/pol.1.1.52>
- Biagi, F., & Loi, M. (2013). Measuring ICT use and learning outcomes: Evidence from recent econometric studies. *European Journal of Education*, *48*, 28–42. <http://dx.doi.org/10.1111/ejed.12016>
- 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*, 184–191. <http://dx.doi.org/10.1177/0741932510381652>
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts (NCEE 2009–4042)*. Washington, DC: Institute of Education Sciences.
- Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*, 198–215. <http://dx.doi.org/10.1016/j.edurev.2012.05.002>
- Cheung, A. C. K., & 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. <http://dx.doi.org/10.1016/j.edurev.2013.01.001>
- Cleary, T. J. (2006). The development and validation of the self-regulation strategy inventory-self report. *Journal of School Psychology*, *44*, 307–322. <http://dx.doi.org/10.1016/j.jsp.2006.05.002>
- Codding, R. S., Shiyko, M., Russo, M., Birch, S., Fanning, E., & Jaspen, D. (2007). Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? *Journal of School Psychology*, *45*, 603–617. <http://dx.doi.org/10.1016/j.jsp.2007.06.005>
- Commissie Meijerink. (2008). *Over de drempels met taal en rekenen - Hoofdrapport van de expertgroep doorlopende leerlijnen taal en rekenen* [Over the threshold with math and language - Main report of the expert group on continuous learning for math and language]. Enschede, Netherlands. Retrieved from <http://www.slo.nl/nieuws/dll/Hoofdrapport.pdf>
- Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology*, *106*, 214–229. <http://dx.doi.org/10.1037/a0034097>
- Crawford, L., & Ketterlin-Geller, L. R. (2008). Improving math programming for students at risk: Introduction to the special topic issue. *Remedial and Special Education*, *29*, 5–8. <http://dx.doi.org/10.1177/0741932507309685>
- Dettmers, S., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2010). Homework works if homework quality is high: Using multilevel modeling to predict the development of the achievement in mathematics. *Journal of Educational Psychology*, *102*, 467–482. <http://dx.doi.org/10.1037/a0018453>
- Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 324–332. <http://dx.doi.org/10.1177/00222194050380040801>
- 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 (NCEE 2007–4005)*. Washington, DC: Institute of Education Sciences.
- Gatti, G. G., & Petrochenkov, K. (2009). *Pearson Successmaker Math efficacy study (2009–10 Final Report)*. Pittsburgh, PA: Gatti Evaluation.
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552. <http://dx.doi.org/10.1037/a0025510>
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293–304. <http://dx.doi.org/10.1177/00222194050380040301>
- Gu, J., Fu, J., & Tong, Y. (2010). Potential uses of computer to enhance reform of mathematics teaching. In *Proceeding of the 5th International Conference on Computer Science & Education*, August 24–27, 2010. Hefei, China: IEEE.
- Haelermans, C., & Ghysels, J. (2015). *The effect of individualized digital practice on math skills at home—Evidence from a randomized experiment on whether and why it works*. (TIER Working Paper Series 15/01). Maastricht, The Netherlands: Top Institute for Evidence-Based Education Research (TIER), Maastricht University.
- Inspectie van het Onderwijs. (2007). *Aansluiting voortgezet onderwijs op het basisonderwijs* [From primary school to secondary school]. (Tech. Rep. No. 20). Retrieved from <http://www.onderwijsinspectie.nl/>
- Käser, T., Baschera, G.-M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., . . . von Aster, M. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology*, *4*, 489.
- Ketterlin-Geller, L. R., Chard, D. J., & Fien, H. (2008). Making connections in mathematics: Conceptual mathematics intervention for low-performing students. *Remedial and Special Education*, *29*, 33–45. <http://dx.doi.org/10.1177/0741932507309711>
- Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, *43*, 489–510. <http://dx.doi.org/10.2190/EC.43.4.d>
- Koninklijke Nederlandse Akademie van Wetenschappen. (2009). *Rekenonderwijs op de basisschool: Analyse en sleutels tot verbetering* [Math education at primary school: Analysis and the keys for improvement]. Retrieved from <https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/20091080.pdf>
- Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education*, *24*, 97–114. <http://dx.doi.org/10.1177/07419325030240020501>
- Leh, J. M., & Jitendra, A. K. (2013). Effects of computer-mediated versus teacher-mediated instruction on the mathematical word problem-solving performance of third-grade students with mathematical difficulties. *Learning Disability Quarterly*, *36*, 68–79. <http://dx.doi.org/10.1177/0731948712461447>
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, *22*, 215–243. <http://dx.doi.org/10.1007/s10648-010-9125-8>
- McComb, E. M., & Scott-Little, C. (2003). *A review of research on participant outcomes in after-school programs: Implications for school counselors (ED482765)*. Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services.
- Muiswerk Educatief. (2013). *Welkom bij Muiswerk Educatief* [Welcome to Mousework]. Retrieved from [www.muiswerk.nl](http://www.muiswerk.nl)
- Núñez, J. C., González-Pienda, J. A., González-Pumariega, S., Rocés, C., Alvarez, L., González, P., . . . Rodríguez, S. (2005). Subgroups of attributional profiles in students with learning difficulties and their

- relation to self-concept and academic goals. *Learning Disabilities Research & Practice*, 20, 86–97. <http://dx.doi.org/10.1111/j.1540-5826.2005.00124.x>
- Organisation for Economic Cooperation and Development. (2014). *PISA 2012 Results in Focus, What 15-year-olds know and what they can do with what they know*. Paris, France: OECD.
- Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102, 896–915. <http://dx.doi.org/10.1037/a0019545>
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012109890-2/50043-3>
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., . . . Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, 47, 833–878. <http://dx.doi.org/10.3102/0002831210367426>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55. <http://dx.doi.org/10.1093/biomet/70.1.41>
- Scheltens, F., Hemker, B., & Vermeulen, J. (2013). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 5: Uitkomsten van de vijfde peiling in 2011 (PPON-reeks nummer 51)* [Drawing up the balance sheet of math language at the end of primary school: Results from the fifth round]. Retrieved from <https://zoek.officielebekendmakingen.nl/blg-219337.html>
- Schijf, G. M., & Schijf, T. J. (2014). *Handleiding Online Testprogramma's* [Guidebook to the online testing programs]. Uithoorn, Netherlands: Muiswerk Educatief.
- Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79, 839–911. <http://dx.doi.org/10.3102/0034654308330968>
- Spiezia, V. (2010). Does computer use increase educational achievements? Student-level evidence from PISA. *OECD Journal: Economic Studies*, 2010, 1–22.
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. *Journal of Educational Psychology*, 105, 970–987. <http://dx.doi.org/10.1037/a0032447>
- Trautwein, U., Lüdtke, O., & Schnyder, I. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, 98, 438–456. <http://dx.doi.org/10.1037/0022-0663.98.2.438>
- van Groenestijn, M. (2007). Rekenvaardigheid in de brugklas. *Remedial*, 7, 5–9.
- Wanlass, R. L. (2012). *The neuropsychology toolkit: Guidelines, formats, and language*. New York, NY: Springer. <http://dx.doi.org/10.1007/978-1-4614-1882-5>
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20, 429–444. <http://dx.doi.org/10.1007/s10648-008-9080-9>
- Wooldridge, J. (2010). *Econometric analysis of cross-section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

Received June 26, 2014

Revision received April 16, 2015

Accepted April 18, 2015 ■

## ORDER FORM

Start my 2016 subscription to the *Journal of Educational Psychology*® ISSN: 0022-0663

_____ \$120.00	APA MEMBER/AFFILIATE	_____
_____ \$240.00	INDIVIDUAL NONMEMBER	_____
_____ \$899.00	INSTITUTION	_____
<i>Sales Tax: 5.75% in DC and 6% in MD</i>		
<b>TOTAL AMOUNT DUE</b>		<b>\$ _____</b>

**Subscription orders must be prepaid.** Subscriptions are on a calendar year basis only. Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION

### SEND THIS ORDER FORM TO

American Psychological Association  
Subscriptions  
750 First Street, NE  
Washington, DC 20002-4242

Call **800-374-2721** or 202-336-5600  
Fax **202-336-5568** : TDD/TTY **202-336-6123**  
For subscription information,  
e-mail: [subscriptions@apa.org](mailto:subscriptions@apa.org)

**Check enclosed** (make payable to APA)

**Charge my:**  Visa  MasterCard  American Express

Cardholder Name \_\_\_\_\_

Card No. \_\_\_\_\_ Exp. Date \_\_\_\_\_

\_\_\_\_\_  
Signature (Required for Charge)

### Billing Address

Street \_\_\_\_\_

City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_

Daytime Phone \_\_\_\_\_

E-mail \_\_\_\_\_

### Mail To

Name \_\_\_\_\_

Address \_\_\_\_\_

City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_

APA Member # \_\_\_\_\_

EDUA16