Exploring the Summer Slide in the Netherlands

Francette Broekman Roger Smeets Eric Bouwers Jessica Piotrowski

Abstract

This study considers the existence of a decline of academic skills in the summer – the so-called "summer slide" – in the context of a relatively short, six-week summer holiday. It further proposes that the extent of this decline will be moderated by home-based practice, and that this practice itself is a function of student and context-level factors. To investigate these issues, a pre-post (summer) study on mental arithmetic was conducted in the summer of 2018 among 932 Dutch grade 2 children. The results demonstrate a sizeable summer learning loss, finding a 0.26 s.d. decline in performance post vs pre-summer. The effect is significantly ameliorated through *observed* practice on an online educational platform, but not through *self-reported* practice. In contrast, perceived competence, gender, and socio-economic status correlate with self-reported practice, but not with observed practice. We discuss the implications of these results for existing and future research in the literature.

Introduction

Since the standardization of grade-leveled curricula in primary schools, the need to standardize the amount of time children spend at school also arose. In the United States, this norm was initially set to a 9-month school calendar with a 3-month summer holiday to align with agricultural production (Cooper et al. 1996). Although ties to agriculture have dropped significantly over the years (from 85% to 3%), this school calendar has never changed, despite concerns that a 3-month break from education may lead to a loss of children's knowledge and skills (i.e. summer learning loss) (Cooper et al. 1996). Indeed, extant research confirms the existence of knowledge and skill loss during the summer (K. L. Alexander, Entwisle, and Olson 2007b; Becker et al. 2008; Coelen and Siewert 2008; Cooper et al. 1996; Lindahl 2001; Lynch and Kim 2017; Shinwell and Defeyter 2017; Zaromb et al. 2014). Summer learning loss is particularly strong in mathematics, and it increases with grade level (K. L. Alexander, Entwisle, and Olson 2007b; Cooper et al. 1996).

These studies raise concern about (the consequences of) summer learning loss. Not only are children ultimately forced to invest extra energy at the beginning of the school year, but the effect is also acutely experienced by teachers. When children have to invest extra effort, this can have a cumulative effect on the classroom whereby teachers are forced to spend time reiterating (part of) the curriculum of the previous year. Scholars have therefore argued that repeated summer learning loss may set some children behind over time, ultimately expanding the achievement gap and impacting these children's future academic careers (K. L. Alexander, Entwisle, and Olson 2007a).

In the United States, these findings have fanned a discussion on the (non)sense of shortening the summer break, thereby increasing the amount of time students spend at school (McCombs et al. 2011). It has been argued that simply increasing "school time" by itself will not help; instead, what matters is increasing "academic learning time", which is time that students spend actually learning material that is relevant to them at that time (Aronson, Zimmerman, and Carlos 1998). The current study addresses this issue more directly, by investigating the absence or presence of learning loss after a much shorter summer break: six instead of twelve weeks.

Further, the present study improves the theoretical foundation of the summer learning loss literature, by incorporating insights from the literature on homework behavior (Trautwein et al. 2006; Trautwein and Ludtke 2009). Specifically, we argue that home-based practice during the summer holiday can be viewed through the same lens as homework behavior. This has two implications: First, home-based practice is a potential moderator of summer learning loss. Second, student and context-level factors might influence the extent of summer learning loss, through their impact on the extent of home-based practice. As a result, incorporating the literature on homework behavior allows us to hypothesize on which children are more or less likely to suffer from summer learning loss.

We address these issues by investigating the extent to which grade 2 children (age 7-9) in the Netherlands experience summer learning loss when they have a short (six-week) summer holiday. We first estimate models of homework behavior as a function of child and context-characteristics. As part of the research design, we have offered the participants free access to an online educational platform ([BLINDED]) during the summer holiday. Importantly, and unlike previous studies in this domain, this enables us to construct an *observed* (in addition to a *self-reported*) measure of homework effort. Second, we estimate the difference between pre and post-summer holiday math performance in second-stage models, while including homework effort as a moderator of difference.

Our results indicate that, even in the context of a six-week summer break, substantial learning loss occurs, with performance dropping by 0.26 s.d. post versus pre-summer. Yet we also find that *observed* – but not *self-reported* – home-based practice ameliorates this loss by approximately 40%. However, *observed* homework effort does not correlate with any of the student and context-level characteristics, whereas *self-reported* homework effort does. We reflect on the implications of these findings for the discussion on the length of summer breaks, as well as research practices in the literature on homework behavior.[1]

[1] The pre-registration of this study can be found online at [BLINDED]

Previous literature and theory

Math Development

Research on the development of numerical cognition has demonstrated that infants already learn number names and use them in counting games (Bruer 1997). By the age of kindergarten, this develops into an ability to compare small numbers for size and devise counting strategies. Even before entering the formal school system, many children use these abilities to learn about basic number facts. All of these skills are thus mostly developed informally, outside the school system.

According to Griffin, Case, and Siegler (1994), this informal knowledge development is key to learning formal arithmetic in school. It includes knowledge of the number names in the child's native language, being aware of the necessity to assign one number name to each object when counting, and understanding that number names occur in a fixed order (Bruer 1997). Eventually, children will develop a sense of cardinality. Mastering these basic concepts lays the foundation on which most complex mathematics skills, such as automaticity, can be built (Entwisle and Alexander 1990; Nesher 1986).

Automaticity means that one is able to proficiently execute on a specific skill without the need for explicit attention or monitoring (Goldman and Pellegrino 1987). It is typically achieved through extensive practice. It is also deemed essential to successfully master a variety of domains in higher mathematics, since it reduces the cognitive load experienced when executing tasks (Woodward 2006). Proper automatization of basic number facts (i.e. mental arithmetic) thus frees up resources which can be spent on solving more complex tasks (Gersten and Chard 1999).

Given its importance for children's further mathematics development, it is not surprising that mental arithmetic is a key skill in the primary school curriculum. In the classroom, significant time and energy are spent on achieving automaticity of basic arithmetic facts (addition, subtraction, multiplication and division). Both through direct teaching of *strategies* and *drill and practice* activities children are encouraged to climb their way up the chain of math development towards automaticity. Generally, these skills are practiced multiple times a week, if not daily throughout the academic year.

Math facts, when introduced to children, are problems to solve that require certain strategies (Siegler 1987). Earlier studies have shown that teaching strategies to solve such problems enhance automaticity (Nunes et al. 2007; Morin and Miller 1998). These may comprise of using visual displays - e.g. ten frames and number lines - and having classroom discussions in which students can share and discuss their own math fact strategies with their classmates (Van De Walle et al. 2007; Siegler 1987; Thornton 1990). The assumption is that when such strategies are taught, children are provided with *procedural knowledge*, which means children learn methods that they can use to derive answers for problems that do not (yet) have pre-stored answers (Nesher 1986). Ultimately, basic math facts turn into *declarative knowledge*, which can be thought of as a more easily and readily accessible mapping between problems and their answers. Teaching children these strategies facilitates long-term retention and direct recall, not only of basic addition facts but also of higher order concepts and problems (Isaacs and Carroll 1999).

Frequently timed *drill and practice* activities are essential to mastering automaticity of math facts (Joy Cumming and Elkins 1999). Over the years, research has demonstrated the effectiveness of teaching children mathematics through drill and practice in a regular classroom setting (Ashcraft 1987; Gagne 1983) as well as through the use of software (Alcalde et al. 1998). As such, both teaching children math strategies as well as performing drill and practice activities are, together, essential for children to master mental arithmetic, which in turn is essential for more complicated math tasks (Gersten and Chard 1999).

Summer learning loss

An interruption in formal instruction during the summer months interrupts the learning process, which has potential consequences for children's mastery of mental arithmetic (Alexander, Entwisle, and Olson 2001). It is likely that children who have not fully mastered mental arithmetic before the summer holiday and do not engage in drill and practice activities during the summer, will fall back to their more basic procedural knowledge. Consequently, by the end of the holiday period, their mental arithmetic development will have deteriorated to some extent - the so-called "summer slide".

Indeed, a body of research dating from 1906 (White 1906) up until today (Paechter et al. 2015) has shown that the summer holiday disrupts the daily rhythm of children's learning and can lead to losses in mental arithmetic skills. This research primarily focuses on lengthy (12 week) summer holidays, typically in the United States. It has sparked an ongoing debate on the necessity of shortening the length of the US summer break, or alternatively, doubling down on summer school programs (McCombs et al. 2011). The aim is to increase the amount of time children spend in school, thereby reducing the extent of the summer slide.

However, it is not at all clear that reducing the length of the summer holiday will help in this regard. As discussed in Aronson, Zimmerman, and Carlos (1998), a distinction should be made between different *types* of time spent on school work. In particular, simply increasing so-called “allocated time” – i.e. the time spent attending school – is insufficient to support learning. Instead, effort should be put in increasing “academic learning time” – which is itself a subset of “time-on-task” – in which students are engaged in academic and/or instructional activities that align with their needs at that point in time.

These considerations raise the question whether a short(er) summer holiday still results in a decrease in academic performance in general, and mental arithmetic in particular. The existing literature does not shed much light on this issue. However, the debate on allocated time vs. academic learning time suggests that even in the presence of a shorter summer holiday, summer learning loss might still arise. As such, we hypothesize the following:

> **H1: Primary school children (7-9 years) will show decreased mental arithmetic performance after a six-week summer holiday compared to their performance before the summer holiday**

Not all children are expected to suffer a summer learning loss to the same extent, however. After all, what induces this effect is a general interruption in (drill and) practice activities, or time-on-task. Yet some children might continue to practice to some degree during the summer holiday. For these children, the interruption of formal instruction or time-on-task will be less severe, and as such we should expect no, or a more limited summer learning loss (Resnick and Ford 2012; Li 1999). Therefore, we derive the following additional (moderating) hypothesis:

> **H2: Home-based math practice during the summer holiday will ameliorate the summer learning loss**

Home-based practice

To understand the potential drivers of home-based practice, we turn to the literature on homework behavior. Although the context of the current study is slightly different – typically, Dutch students do not get explicit homework instructions during school holidays – the underlying construct – i.e. home-based practice – is similar.

Trautwein et al. (2006) develop a conceptual mode of homework behavior.[2] First, they define homework behavior as a combination of three components: homework effort, time spent on homework, and learning strategies. They then provide an elaborate discussion of all the drivers of homework behavior. It is primarily driven by homework motivation, which is split up into an expectancy component - how competent do students feel doing (home)work in a particular subject - and a value component - how do they gauge the cost-benefit calculation of doing (home)work. More specifically, students with high self-competence (in a specific domain) and/or high perceived net utility of doing homework will, *ceteris paribus*, demonstrate more favorable homework behavior.

Perceived (math) competence has already been demonstrated to be positively correlated with math performance directly (Guay, Marsh, and Boivin 2003; Marsh and Craven 2006; Marsh and Martin 2011). Based on the theory of Trautwein et al. (2006), we should further expect a positive correlation between (math) homework behavior and perceived (math) competence. As discussed below, our measures of homework behavior mostly capture self-reported or observed home-based practice. As such, we will frame our hypotheses accordingly, leading to our third hypothesis:

[2]See in particular Figure 1 on p.440 of the paper.

H3: Perceived mathematics competence is positively related to the extent of home-based math practice during the summer holiday

Trautwein et al. (2006) further describe a number of factors that drive homework motivation (i.e. expectancy and value). They split these factors into the learning environment, student characteristics, and the role of parents. One of the student characteristics that they include concerns gender.

The impact of gender on homework behavior might be understood through its role in achievement motivation (Meece, Glienke, and Burg 2006). Gender differences in different achievement domains have been widely established, and they have been explained from various angles, such as attribution theory or expectancy-value theory (Weiner 1986; Eccles et al. 1983). Research shows that boys generally attribute success more to natural ability, whereas girls mention hard work more often (Meece, Glienke, and Burg 2006).

Along similar lines, Trautwein et al. (2006) argue that gender has a *direct* effect on homework behavior, with girls generally exercising more homework *effort*. However, there is also a *mediating* effect of gender via homework *motivation* (i.e. the expectancy-value part of the framework). Whether or not this reinforces or opposes the direct effect depends on the specific subject, according to the authors.

This raises the question if boys are more or less motivated than girls to do *math* homework. In a variety of different studies, Trautwein and his co-authors find that boys have higher self-competence in mathematics, but not in other areas (such as English, German, or biology) (Trautwein et al. 2006; Trautwein and Ludtke 2009; Dettmers et al. 2010). This result is consistent with other studies, which further show that this effect is most notable at a younger age, and that it is not present (for math) in the value component of the framework (Meece, Glienke, and Burg 2006).

In sum, although girls display more favorable homework behavior in general, their expected lower self-competence in mathematics specifically works against this effect in the specific domain of *math* homework. The net effect is thus an empirical question, and as such we formulate a neutral hypothesis:

H4: Gender is related to the extent of home-based math practice during the summer holiday

The role of parents is another factor in the model of Trautwein et al. (2006). The authors break this down into a number of aspects, such as academic expectations, parent-child communication, parent attitudes regarding homework, and the quantity and quality of homework help.

There is a more generic and well-established literature on the impact of parental involvement on academic achievement (Hoover-Dempsey and Sandler 1995; Hoover-Dempsey et al. 2005). Although some forms of parent engagement are more effective than others - with apparent contingencies on e.g. child age and ethnicity - the general finding is that more parent engagement positively affects academic achievement. The channels through which this factor operates include providing support, providing resources, and providing a role model for children. According to Trautwein et al. (2006), these factors thus also contribute to more homework effort.

As discussed more elaborately below, capturing parental engagement via child-based surveys proved very challenging, forcing us to look for another metric. A typical (albeit coarse) proxy for parental engagement is socio-economic status (SES) (Thomson 2018). It aims to capture a variety of traits that correlate with the transmission channels just mentioned, such as the level of education of parents, their financial resources, and the quality of the home environment. In a broad meta-review of the literature, Sirin (2005) confirms the positive relationship between SES and academic achievement, and its use as a construct of parental engagement. We thus arrive at our fifth and final hypothesis:

H5: Children from high-SES communities are more likely to engage in home-based math practice during the summer holiday than children from low-SES communities

Taken together, the previous sections suggest a two-step empirical approach. First, we will model home-based math practice during the summer holiday, based on the drivers discussed in H3-H5. Second, we will model summer learning loss, and consider the moderating effect of home-based math practice during the summer holiday, as discussed in H1 and H2.

Method

Design and Procedures

After receiving ethical approval from the Ethical Committee of [BLINDED], a pretest-posttest survey was designed to test the hypotheses. Specifically, a mental arithmetic speed test was conducted with primary school children before (grade 2) and after (grade 3) the summer holiday. The test was administered in a classroom-setting within the participating school. All tests were conducted by three researchers that followed a pre-defined procedure. The pre- and posttest were planned as close to the summer holiday as logistically possible, which turned out as the penultimate week before the holiday for the pretest, and the first week after the summer holiday for the posttest. Planning tests closer to the start of the holiday was deemed not feasible due to extra curricular activities planned. However, as the last week of the school year is regularly filled with more recreational activities, we can assume that very little instruction was given.

The administration procedures for the pretest and posttest were largely similar. After entering the classroom, one of the researchers gave a brief introduction. Thereafter, all participants were seated according to the test-taking seating arrangement in that particular classroom, which most often involved separating all tables from each other. Prior to administering the mental arithmetic speed test, the participants received a survey-package consisting of the pretest-survey, the posttest-survey, a name form, and a letter for the parents. Each survey-package had a unique number linking the forms together. After handing out a package to each participant, the researcher dictated each question of the pretest-survey and overlooked whether all participants filled in the correct question (together with the second researcher). If something was unclear to the students, they were allowed to ask questions.

Following completion of the individual difference items, the researcher explained the mental arithmetic speed test according to the instructions manual of the validated test. The test included four sheets of 50 sums each; addition, subtraction, multiplication, and division. Participants had two minutes per sheet to solve as many problems as possible without using additional aids. Afterward, the test including the pretest-survey was collected by the researchers. As an incentive, all participants were given a set of stickers and the researcher explained that the letter to their parents contained a code with which they could claim free access to an online educational platform during the summer holiday.

In the posttest, after the summer holiday, the survey-package was handed out (with help of the teacher) based on the name written on the name form. The procedure of the posttest was similar to the pretest with the exception that no individual difference variables were collected. As an incentive, the participants were given stickers as well as a board-game for the whole classroom. The pretest survey took 30-45 minutes while the posttest took 20-30 minutes.[3]

Recruitment of Schools

Schools were recruited through a digital newsletter for teachers that was sent via email as well as through social media. This call for participation explicitly stated the desired target group of Grade 2 students (7-9 years). A total of 124 teachers responded to the call. When asked for further contact information, teachers from 90 schools replied and were called for informed consent. They were explained the survey design and the contents of the mental arithmetic speed test and survey, as well as the expected effort from their side (logistics). From the schools that gave active consent (N = 88) a (semi)random selection was made of 43 schools to better represent the Netherlands at large. Specifically, sampling was based on an equal distribution of participants across the five Dutch regions defined for market research (the "Nielsen Districts", see ("Gouden Standaard: Een Uniek Ijkingsinstrument Voor Nationale En Regionale Steekproeven" 2018)). For logistic reasons, schools were selected based on the possibility of visiting all locations within the weeks of the pretest (weeks 26, 27, and 28 of 2018) and the posttest (weeks 34, 35, and 36 of 2018).[4]

Teachers of the selected schools were asked to send a passive consent letter to parents whereby the parent could opt out of the study (which a total of four chose to do). This letter stated that researchers

[3]The first survey took longer because the researchers and research objective had to be introduced to the children, and a number of personal traits (such as age and gender) had to be surveyed.

[4]In terms of school holidays, the Netherlands is divided in three regions (North, Middle and South), with each region having a different start date (and end date) of the summer holiday. In 2018, the summer holiday ran during weeks 28-33 in region South, weeks 29-34 in region Middle, and weeks 30-35 in region North.

would come by and perform a short survey as well as a math test with the children in the classroom. Data was excluded if 1) the arithmetic performance test was not fully completed, either before or after the summer holiday or 2) the arithmetic performance test was only filled out once, either before or after the summer holiday. In the analyses, only students with complete data were included.

Participants

A total of 43 Dutch schools participated in the study. In the pretest, 1034 children participated in the study. Of these children, a total of 984 children participated in the posttest.[5] Hence, 984 children between 7-9 years of age (M = 7.7, SD = 0.58) participated in both the pre- and posttest. After testing the assumptions for the multilevel analysis and dropping students with incomplete data, a total of 932 children were included in the study. The distribution of children over the five different regions in the Netherlands was largely proportional to population-level statistics (three big cities Amsterdam, Rotterdam, The Hague 18%; North 16%; East 19%; South 19%; West 27%).

Measures

Mental Arithmetic Performance (DV - student-time variable) The mastery of automaticity (mental arithmetic) was measured through the use of a mental arithmetic speed test ('Tempo-Test Automatiseren' (TTA)) (De Vos 2010). The TTA is the successor of the 'Tempo-Test Rekenen' (De Vos 1992) and is among the most frequently used tests to timely detect dyscalculia and other mathematics disorders (Desoete 2004). It has been used in previous studies on children's arithmetic performance (Jansen et al. 2013; Van Mier, Schleepen, and Van den Berg 2018). It gives an indication of the level of mental arithmetic at which the child performs. The TTA is a standardized test which can be administered in both a one-on-one as well as a group setting. The choice for this test over a simpler school skill test mental calculation ('Schoolvaardigheidstoets Hoofdrekenen' (SVA)) is that the TTA provides more detail on the different domains of mental arithmetic. It measures the level of mental arithmetic on four sub-domains; addition, subtraction, multiplication and division. Every domain includes 50 mathematical problems (sums) on one sheet. The participant has two minutes per sub-domain to solve as many sums as possible, thus eight minutes of total testing time. The raw scores of the test were converted into the norm-scores (percentiles) provided by De Vos (2010). In the analyses, mental arithmetic performance was measured through the total norm-score in percentiles (0 - 100) over all four domains ($M_{\text{total\_pretest}} = 1.3$, $SD_{\text{total\_pretest}} = 1$; $M_{\text{total\_posttest}} = 1$, $SD_{\text{total\_posttest}} = 0.96$). A normed-score between 40 and 60 is marked as 'average score' which shows that the pretest arithmetic performance in the current sample is below-average. This may be due to the inclusion of division, which Dutch children start learning in Grade 3. However, given the SD, the sample ranges from below-average performers to above average performers.

Hypothesis 1 will be tested by including a *post-summer indicator* variable that takes the value 1 for mental arithmetic performance measured after the summer, and 0 for that measured before the summer. Based on our first hypothesis, we expect that its estimated coefficient will be negative.

Home-based practice (DV and IV - student variable) As discussed above, Trautwein et al. (2006) suggest that homework behavior is ideally conceptualized and measured as a multidimensional construct, involving effort, time and learning strategies. However, due to the young age of the children in our sample (7-8), asking complex and nuanced survey questions was unfeasible. Instead, we rely on three simple self-reported measures of home-based practice effort, as well as one observed measure.

Children were asked in the plenary administered posttest to what extent they perceived to have practiced for school during the summer months. Specifically, children were asked if, during the summer holiday, they had (a) practiced for school with their parents, (b) practiced specifically on the online educational platform ([BLINDED]), and (c) practiced math in particular during the summer. The first two items were measured on a 5-point Likert scale, ranging from 1 (not at all) to 5 (very often). The latter question was measured on a binary scale (i.e. yes/no).

[5] Reasons for dropout include students that got left back, students that were sick on the day of the posttest, and students that had moved to a different school. Regarding left back students, we made an effort to get them included in the posttest as well, but this was not always successful.

To measure *observed* practice, we collected online play data for those participants that activated their free access to the online educational platform that was given as an incentive for participation.[6] To distinguish the participants that only explored the platform from those who used the platform more frequently for practicing mathematics, observed practice was defined following the typical usage patterns of the educational platform. In this study, length (minutes) and frequency (number of weeks) of use are considered as the two most prominent indicators that can determine children's practice behavior. Observed practice was therefore defined as visiting the educational platform in at least two weeks during the summer holiday while practicing 15 minutes of math on average per active week. In total, 282 students activated their account, while 83 students used it according to the definition provided above.

Based on Hypothesis 2, we expect that these measures of home-based practice show a significant interaction with the post-summer indicator discussed above. Specifically, we expect that the estimated coefficient on the interaction term will be positive.

Perceived Competence (IV - student variable) Children's perceived math competence was measured by a selection of five items from the Intrinsic Motivation Inventory (IMI) (Vos, Van Der Meijden, and Denessen 2011). The original questionnaire (measuring intrinsic motivation) consists of seven sub-scales. For this study, the sub-scale 'perceived competence' was selected since this scale has been used in earlier research with young children (Vos, Van Der Meijden, and Denessen 2011). The scale consists of five (slightly altered) items such as "I think I am good at math" and "I think I am pretty good at math, compared to others". Children were asked to rate to what extent they agreed with the statements on a five-point Smiley-O-Meter. The Smiley-O-Meter (Read 2008) was used, because the statements were offered to children aged 7-9 years and research has shown that this age group can have difficulties indicating their level of agreement on a regular five-point Likert scale. The perceived competence scale was administered in the pretest as well as in the posttest. There were no significant mean differences between pretest and posttest scores, $t(939) = -0.1$, $p = 0.92$. Confirmatory factor analysis shows that all five variables load above .40 on one component and reliability analysis shows that the scale is reliable ($\alpha = 0.87$). Perceived competence was computed by calculating the mean scores of all items. Because this variable was skewed, for the purpose of the multilevel models, perceived competence was mean centered on the grand mean.

Based on Hypotheses 3, we expect that this variable will have a positive *direct* effect on home-based practice. Following previous literature, we also include it as a control variable in the summer learning loss model, to capture its impact on general math performance.

Gender (IV - student variable) In this study, gender was measured by asking the participants to fill out their gender in the administered pretest-survey. The sample includes 493 boys and 462 girls. Based on Hypothesis 4, we expect that this variable will have an impact on the extent of home-based practice during the summer holiday, although it is unclear from the theoretical framework what the (net) effect will be. Further, we also include this variable as a control feature in the summer learning loss model.

Socio-economic status (IV - school variable)

We initially attempted to collect suitable data to capture SES at the individual level. However, this did not work well for the children in our sample. Specifically, we attempted to measure individual SES by asking how many cars where present in the household. Even though Torney-Purta and colleagues (2001) showed that such questions provide reasonable proxies for SES, the current study revealed inadequate estimates by the children, which made this measure very unreliable.[7] Instead, we use the mean average household income per postal code (of the school) as a SES indicator, which was retrieved from Statistics Netherlands (CBS) (M = 35198; SD = 6820). Based on Hypothesis 5, we expect a positive effect on the extent of home-based practice

[6]Upon registration for use of the platform, parents have to agree to the terms of use and the privacy policy. Chapter 13 in the privacy policy states that automatically generated play data can be used by designated employees for research and analysis, with the objective of evaluating and improving the performance of the platform. It guarantees that this data will only be used in aggregated form, and will never be tied back to individual users. The collection and use of practice data in the present study conforms with this policy.

[7]Asking children to count the number of cars in a household created unforeseen problems, such as how to (correctly) count when parents live separately, or when (a) car(s) is provided by a parent's employer (which is fairly common in the Netherlands). Children whose parents owned a garage or taxi company also were confused by the question.

during the summer holiday. We also include this variable as a control feature in the summer learning loss model.

Statistical Model

As discussed above, we will estimate two separate models. First, we estimate the impact of perceived math competence, gender and SES on home-based practice (H3-H5). Second, we estimate the extent of summer learning loss, as well as the moderating impact of home-based practice on this loss (H1-H2), while controlling for the direct impact of the explanatory variables from step one.

Note that the collected data form a natural hierarchy. In particular, in the second model we have two (i.e. repeated) observations (level 1) of students (level 2) in schools (level 3). The implication is that observations at lower levels are not independently distributed, violating a common assumption of regular statistical analyses (e.g. t-tests or multiple regression analysis). Applying such methods will lead to biased estimates of standard errors. Moreover, cross-level interdependencies typically create heteroskedastic error terms, which violates another key assumption of such analyses. Further, we are interested in estimating the direct and moderating effects of covariates on the dependent variable at various levels of observation. This requires an estimation method that can properly handle the different dependencies between these levels (Karakolidis, Pitsia, and Emvalotis 2016).

For these reasons, we use multilevel analysis in the second model (Hox 2010).[8] Specifically, we estimate a *random coefficient* model, where we allow lower level regression coefficients to vary at higher levels. Specifically, we allow for *random intercepts* at both the student and school level, as well as a random slope for the post-summer indicator variable at the school level.[9]

Results

Descriptive statistics

Table 1 presents average pre and post-summer performance scores per math domain, split by gender. Girls on average score lower than boys on all domains of mental arithmetic (addition, subtraction, multiplication, division). This seems in line with the steeper growth in mathematics development of boys found in earlier research (Leahey and Guo 2001). However, the (unconditional) pre-post drop does not show significant gender differences. Second, the performance drop is most pronounced in multiplication (12.1%-points for girls, and 9.4%-points for boys). This is likely driven by the fact that multiplication is taught from grade two onward, whereas addition and subtraction are taught from grade one onward. Division is taught starting in grade three, which explains the substantially lower scores, as well as the comparatively limited drop.

Table 2 shows the school-level variation in average pre and post-summer norm score percentiles - which, as the table shows, is quite large. On the low end, some schools fall in the 5th/7th percentile of the norm score distribution (below average), whereas on the high end, some schools fall in the 71st/78th percentile (above average). In some schools, we visited multiple classes, which allows us to compare these numbers for within-school, across-classroom variation. Although the difference between best and worst performing classes is notably lower in this case, the median difference is still between 16 percentiles (pre) and 12 percentiles (post). This suggests that teacher differences are a significant driver of performance differences between students.

Table 3 presents performance correlations across math domains, both before and after the summer. All of them correlate positively, yet the correlation is notably higher between addition and subtraction than between the other domains. This corroborates the observation above, i.e. that addition and subtraction have been taught starting in grade one, unlike the other two domains. Further, correlations are very similar pre and post-summer holiday.

[8]The first (homework behavior) model is also characterized by a natural hierarchy, although here we only have one observation per student, further reducing the available degrees of freedom. Moreover, we are not primarily interested in cross-level interaction or moderating effects here. Therefore, we rely on ‘regular’ regression techniques in this case.

[9]Since we only have two observations per student, we do not have enough degrees of freedom in the data to meaningfully estimate random slopes at the student level as well. Therefore, we restrict the random component at the student level to intercepts only.

Table 4 shows the empirical relationships between the four metrics of home-based practice. To facilitate comparisons, self-reported practice for school and self-reported online practice were transformed to binary variables, with 0 for students who indicated no practice at all, and 1 for the rest. Further, for *observed* online practice, we scored students as 1 if they played on the online platform *at all*, i.e. not yet adopting the definition of minimum/typical usage discussed above.

The table contains a number of quadrants. Each cell c_{ij} of a quadrant is computed as follows:

$$c_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \tag{1}$$

with N_{ij} being the number of respondents in row i and column j. For example, the top left quadrant shows that of all participants indicating that they practiced for school during the summer holiday, 65% indicates they practiced on the online platform and 35% indicates they did not. For those indicating that they did *not* practice for school, these percentages are reversed.

The key insight from Table 4 arises by comparing the quadrants in the bottom row with the others. In all quadrants above the bottom row, the diagonal percentages (i.e. the combinations yes-yes and no-no) are >= 50%. However, in the bottom row, this is not the case. In particular, the off-diagonal percentages in the no observed online practice row are always >= 50%. This implies that the majority of respondents that report to have practiced during the summer, did not do so on the online platform.

Notably, this deviation is most pronounced for the combination self-reported online practice versus observed online practice. In this case, 85% of participants who report to have practiced online are not observed to have done so. These results highlight a disconnect between what the children in our study say versus what they do. We will return to this issue below.

Finally, Table 5 presents pairwise correlations across variables included in the models.

Home-based practice

Table 6 presents the results of four home-based practice models, one with each of the four home-based practice variables as a dependent variable. The first two models are based on linear regression (as the dependent variables are measured on a Likert scale), whereas the last two models use logistic regression (since the dependent variables are binary).[10] All beta coefficients are expressed as effect sizes, i.e. they indicate the standard deviation change in y following a one standard deviation increase in x. For the binomial models, odds ratios are reported in addition to the beta coefficients.

The first two models capture *general* home-based practice, i.e. not necessarily math-related. Apart from a significant positive impact of being a girl on generic practice for school, none of the explanatory variables have a significant impact.

This changes notably in the third model, which captures whether or not children engaged in any self-reported *math* home-based practice during the summer holiday. In this case, all the variables have a positive and significant impact, corroborating H3 (math competence) and H5 (SES). Regarding H4 (gender), we see that girls are more likely to report to have engaged in math home-based practice than boys, similar to the results in model (1). These results broadly resonate with those reported by Trautwein et al. (2006).

However, when considering *observed* (as opposed to self-reported) math home-based practice based on the online play data in model (4), none of the explanatory variables have a significant impact. This suggests either that most children practiced math outside the online platform, or that the correlations in model (3) are capturing an underlying, unobserved relationship, such as self-confidence or overconfidence.

Summer learning loss

Table 7 presents the results for the multilevel summer learning loss models. The model is gradually built up across columns. All coefficients report effect sizes. The model deviance statistic tests are always with the most complete nested model as a baseline. E.g. the model in column (3) uses the model in column (2) as its baseline, whereas the models in columns (4)-(8) use the model in column (3) as their baseline.

[10] We also experimented with transforming the dependent variables in the first two models into binary indicators and estimating logistic regression models. The results were unchanged.

Column (1) estimates a model without predictors (i.e. a null model). The intercept represents the overall performance on mental arithmetic.[11] The intraclass correlation (ICC), which represents the proportion of variance explained due to the grouping structure (two observations per student, and multiple students per school), shows that for overall mental arithmetic performance 66% of the variance could be explained at the student-level and 21% at the school-level.[12] The remaining 13% of variance is generated across time, i.e. between pre and post summer variation.

The model in column (2) adds a post-summer indicator. The coefficient is negative and statistically significant, supporting H1. Specifically, after the summer holiday mental arithmetic performance declines by approximately 26% of a standard deviation relative to before the summer holiday.[13] Recall from the discussion above that we allow this effect to vary across schools. The breadth of this variation is computed by adding/subtracting twice the school-level post-summer standard deviation (also reported in the table) from the post-summer coefficient. This yields a range of (-0.5,-0.022) between which the summer learning loss varies across schools. There is thus significant variation across schools in the extent to which children experience summer learning loss.

The model in column (3) adds the explanatory variables from Table 6 as covariates in the summer learning loss model. Consistent with earlier literature, all three variables have the expected statistically significant effects on overall math performance. Reported math competence is positive, boys perform on average better than girls, and an increase in SES increases math performance. The latter two effects are comparable in (absolute) size.

Models (4)-(7) add each of the home-based practice variables considered in Table 6 individually, as well as their interaction with the post-summer indicator. Self-reported school and online practice demonstrate no effect whatsoever. In contrast, self-reported math practice in column (6) shows a positive and significant impact on overall math performance, yet no moderating effect on summer learning loss. The opposite is true for observed online math practice in column (7), which shows no direct impact on overall math performance, yet a positive moderating effect on the extent of summer learning loss, consistent with H2. In terms of effect sizes, having practiced math online during the summer holiday roughly reduces the summer learning loss on average by 40% (0.103/-0.270)

Finally, the model in column (8) combines all the features from columns (2)-(7) in one big model. All the previous results carry over to this setting as well, except for one variable: The negative direct impact of self-reported school practice now becomes statistically significant. This likely is driven by multicollinearity with some of the other home-based practice variables (cf. Table 5). Based on the results in column (8), Figure 1 plots the development in performance before and after the summer for the students that did vs did not have any observed online math practice. It is clear that the decline in the former group is less steep.

Conclusion and discussion

Our results show that Dutch primary school children's mental arithmetic performance significantly decreases after a six-week summer holiday when compared to their performance before the summer holiday, supporting H1 of this study. Specifically, the drop constitutes +/- a 0.26 standard deviation reduction in performance (approximately a seven percentile point decrease on the test curve). Moreover, the multilevel model demonstrates considerable school-level variation in summer learning loss.

These findings provide additional support for the claim that simply reducing the length of the summer break (e.g. in the United States) by itself is no guarantee that summer learning loss will be prevented (Aronson, Zimmerman, and Carlos 1998; McCombs et al. 2011). However, it might be the case that the summer learning loss is decreased following a shorter summer break. Unfortunately, differences in measurements of performance loss across studies impede such a comparison.

[11] The fact that the coefficients express effect sizes makes this somewhat difficult to interpret. The unstandardized coefficient is 29.9, implying that on average (pre and post summer combined) the students in our sample score around the 30th percentile of the test curve.

[12] The ICC for a particular hierarchical level is computed as the sample variation on that level, divided by the sum of sample variation at all levels (Hox 2010). Sample variations at each level are computed by taking the square of the standard deviations reported in Table 7. The time-level variation is the remainder of total variation, i.e. which is not captured by student and school variation.

[13] In absolute terms, the students in the sample drop +/- 7 percentile points on average on the test curve after the summer holiday.

The results further suggest a negative impact of interrupting an incomplete learning process. In other words, as grade 2 children have not yet reached the final stage of automaticity before the summer holiday, continued practice seems necessary. Indeed, consistent with H2, we find that *observed* home-based practice during the summer on the online platform [BLINDED] helps to limit the extent of summer learning loss by almost 40%. This result aligns with the distinction between "school time" or "allocated time" versus "academic learning time", and suggests that the latter can also occur outside the school and classroom (Aronson, Zimmerman, and Carlos 1998).

In the models that capture homework behavior, we found support for H3-H5 when considering *self-reported* home-based practice in math, but not when using *observed* practice. One explanation for this discrepancy is that online practice is just a (small) part of general math practice, and as such the latter metric is incomplete. Indeed, 65% of children report to have practiced some math during the summer holiday, whereas only 30% activated their online platform account.

Another explanation is that what kids say is different from what they do, as illustrated in Table 4. In part, this can be driven by providing socially-desirable answers. Table 4 shows that only 15% of participants that report to have practiced online have actually done so (i.c. 82 out of 533). It is not unlikely that many children experienced pressure to answer in accordance with what they perceived the researchers (some of whom were affiliated with the online platform) would want to hear. In other part, it can be driven by misinterpreting the question, or conflating (perceived) competence with home-based practice effort. A t-test comparing perceived math competence between students that did versus did not report any math home-based practice shows a significantly higher value for the former group. Further, the results in Table 4 highlight the much stronger correlations between all self-reported measures of correlation *vis-a-vis* the observed measure. A shared underlying bias, such as self-confidence or overconfidence, may be behind these results.

Ironically then, the metric of home-based practice that *does* correlate significantly with its proposed drivers in Table 6 (i.e. self-reported math practice) does *not* ameliorate summer learning loss. Conversely, the metric that *does* limit summer learning loss in Table 7 (i.e. observed online math practice) does *not* correlate with its proposed drivers. Stated differently, self-reported effort does nothing to moderate math skill deterioration over time, and has a weak relation with actual (i.e. observed) home-based practice effort, which *does* ameliorate summer learning loss.

This observation also sheds a critical light on the homework behavior literature. The majority of studies in this field completely rely on self-reported measurements of the various model components. This is a logical consequence of the fact that homework behavior by definition occurs outside the classroom, and hence cannot be easily observed or objectively monitored. Yet this approach runs the risk of generating conflated or misinterpreted measures, with correlations between model constructs being driven (in part) by unobserved underlying constructs. With the advent of online learning tools for use at home, it would be useful to extend the homework behavior literature with observed measurements of homework behavior as well.

References

Alcalde, C, JI Navarro, E Marchena, and G Ruiz. 1998. "Acquisition of Basic Concepts by Children with Intellectual Disabilities Using a Computer-Assisted Learning Approach." *Psychological Reports* 82 (3): 1051–6.

Alexander, Karl L., D. Entwisle, and L. Olson. 2001. "Schools, Achievement, and Inequality: A Seasonal Perspective." *Educational Evaluation and Policy Analysis* 23 (2): 171–91.

———. 2007a. "Lasting Consequences of the Summer Learning Gap." *American Sociological Review* 72 (2): 167–80. https://doi.org/10.1177/000312240707200202.

———. 2007b. "Summer Learning and Its Implications: Insights from the Beginning School Study." *New Directions for Youth Development* 2007 (114): 11–32.

Aronson, J., J. Zimmerman, and L. Carlos. 1998. "Improving Student Achievement by Extending School: Is It Just a Matter of Time?" *WestEd*, 1–9.

Ashcraft, M. 1987. "Children's Knowledge of Simple Arithmetic: A Developmental Model and Simulation." In *Formal Methods in Developmental Psychology*, 302–38. Springer.

Becker, Michael, Petra Stanat, Jurgen Baumert, and Rainer Lehmann. 2008. "Lernen Ohne Schule: Differenzielle Entwicklung Der Leseleistungen von Kindern Mit Und Ohne Migrationshintergrund Wahrend Der Sommerferien." In *Migration Und Integration*, 252–76. VS Verlag fur Sozialwissenschaften.

Bruer, John T. 1997. "Education and the Brain: A Bridge Too Far." *Educational Researcher* 26 (8): 4–16.

Coelen, Hendrik, and Jörg Siewert. 2008. "Der Ferieneffekt—Auch in Deutschland Schichtspezifisch?" In *Chancenungleichheit in Der Grundschule*, 87–90. Springer.

Cooper, H., B. Nye, K. Charlton, J. Lindsay, and S. Greathouse. 1996. "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review." *Review of Educational Research* 66 (3): 227–68.

Desoete, A. 2004. "Diagnostische Protocollen Bij Dyscalculie: Zin of Onzin?" *Significant* 3.

Dettmers, Swantje, Ulrich Trautwein, Oliver Ludtke, Mareike Kunter, and Jurgen Baumert. 2010. "Homework Works If Homework Quality Is High: Using Multilevel Modeling to Predict the Development of Achievement in Mathematics." *Journal of Educational Psychology* 102 (2): 467–82.

De Vos, Teije. 1992. "Tempo-Test-Rekenen." *Handleiding.[Tempo Test Arithmetic. Manual]. Nijmegen: Berkhout*.

———. 2010. *TempoTest Automatiseren*. Boom Test Uitgevers.

Eccles, J., T. Adler, R. Futterman, S. Goff, C. Kaczala, and J. Meece. 1983. "Achievement and Achievement Motives." In, edited by J. Spence. San Francisco: Freeman.

Entwisle, Doris R, and Karl L Alexander. 1990. "Beginning School Math Competence: Minority and Majority Comparisons." *Child Development* 61 (2): 454–71.

Gagne, Robert M. 1983. "Some Issues in the Psychology of Mathematics Instruction." *Journal for Research in Mathematics Education*, 7–18.

Gersten, Russell, and David Chard. 1999. "Number Sense: Rethinking Arithmetic Instruction for Students with Mathematical Disabilities." *The Journal of Special Education* 33 (1): 18–28.

Goldman, Susan R, and James W Pellegrino. 1987. "Information Processing and Educational Microcomputer Technology: Where Do We Go from Here?" *Journal of Learning Disabilities* 20 (3): 144–54.

"Gouden Standaard: Een Uniek Ijkingsinstrument Voor Nationale En Regionale Steekproeven." 2018. https://www.moaweb.nl/services/services/gouden-standaard.html.

Griffin, Sharon A, Robbie Case, and Robert S Siegler. 1994. *Rightstart: Providing the Central Conceptual Prerequisites for First Formal Learning of Arithmetic to Students at Risk for School Failure.* The MIT Press.

Guay, Frederic, Herb Marsh, and Michel Boivin. 2003. "Academic Self-Concept and Academic Achievement: Developmental Perspectives on Their Causal Ordering." *Journal of Educational Psychology* 95 (1): 124–36.

Hoover-Dempsey, Kathleen, and Howard Sandler. 1995. "Parental Involvement in Children's Education: Why Does It Make a Difference?" *Teachers College Record* 95: 310–31.

Hoover-Dempsey, Kathleen, Joan Walker, Howard Sandler, Darlene Whetsel, Christa Green, Andrew Wilkinson, and Kristen Closson. 2005. "Why Do Parents Become Involved? Research Findings and Implications." *The Elementary School Journal* 106 (2): 105–30.

Hox, Joop J. 2010. *Multilevel Analysis. Techniques and Applications.* Routledge: New York.

Isaacs, Andrew C, and William M Carroll. 1999. "Strategies for Basic-Facts Instruction." *Teaching Children Mathematics* 5 (9): 508.

Jansen, Brenda RJ, Jolien Louwerse, Marthe Straatemeier, Sanne HG Van der Ven, Sharon Klinkenberg, and Han LJ Van der Maas. 2013. "The Influence of Experiencing Success in Math on Math Anxiety, Perceived Math Competence, and Math Performance." *Learning and Individual Differences* 24: 190–97.

Joy Cumming, J, and John Elkins. 1999. "Lack of Automaticity in the Basic Addition Facts as a Characteristic of Arithmetic Learning Problems and Instructional Needs." *Mathematical Cognition* 5 (2): 149–80.

Karakolidis, Anastasios, Vasiliki Pitsia, and Anastassios Emvalotis. 2016. "Examining Students' Achievement in Mathematics: A Multilevel Analysis of the Programme for International Student Asessment (Pisa) 2012 Data for Greece." *International Journal of Educational Research* 79: 106–15.

Leahey, Erin, and Guang Guo. 2001. "Gender Differences in Mathematical Trajectories." *Social Forces* 80 (2): 713–32.

Li, Shiqi. 1999. "Does Practice Make Perfect?" *For the Learning of Mathematics* 19 (3): 33–35.

Lindahl, M. 2001. "Summer Learning and the Effect of Schooling: Evidence from Sweden."

Lynch, K., and J. S. Kim. 2017. "Effects of a Summer Mathematics Intervention for Low-Income

Children: A Randomized Experiment." *Educational Evaluation and Policy Analysis* 39 (1): 31–53.

Marsh, Herb, and Rhonda Craven. 2006. "Reciprocal Effects of Self-Concept and Performance from a Multidimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspectives." *Perspectives on Psychological Science* 1 (2): 133–1163.

Marsh, Herb, and Andrew Martin. 2011. "Academic Self-Concept and Academic Achievement: Relations and Causal Ordering." *British Journal of Educational Psychology* 81 (1): 59–77.

McCombs, J., C. Augustine, H. Schwartz, S. Bodilly, B. McInnis, D. Lichter, and A. Cross. 2011. *Making Summer Count: How Summer Programs Can Boost Children's Learning.* RAND Corporation.

Meece, J., B. Glienke, and S. Burg. 2006. "Gender and Motivation." *Journal of School Psychology* 44 (5): 351–73.

Morin, Victoria A, and Susan Peterson Miller. 1998. "Teaching Multiplication to Middle School Students with Mental Retardation." *Education and Treatment of Children*, 22–36.

Nesher, Pearla. 1986. "Learning Mathematics: A Cognitive Perspective." *American Psychologist* 41 (10): 1114.

Nunes, Terezinha, Peter Bryant, Deborah Evans, Daniel Bell, Selina Gardner, Adelina Gardner, and Julia Carraher. 2007. "The Contribution of Logical Reasoning to the Learning of Mathematics in Primary School." *British Journal of Developmental Psychology* 25 (1): 147–66.

Paechter, M., S. Luttenberger, D. Macher, F. Berding, I. Papousek, E. M. Weiss, and A. Fink. 2015. "The Effects of Nine-Week Summer Vacation: Losses in Mathematics and Gains in Reading." *Eurasia Journal of Mathematics, Science & Technology Education* 11 (6).

Read, J. C. 2008. "Validating the Fun Toolkit: An Instrument for Measuring Children's Opinions of Technology." *Cognition, Technology, and Work* 10 (2): 119–28. https://doi.org/10.1007/s10111-007-0069-9.

Resnick, Lauren B, and Wendy W Ford. 2012. *Psychology of Mathematics for Instruction.* Routledge.

Shinwell, Jackie, and Margaret Anne Defeyter. 2017. "Investigation of Summer Learning Loss in the Uk?Implications for Holiday Club Provision." *Frontiers in Public Health* 5: 270.

Siegler, Robert S. 1987. "The Perils of Averaging Data over Strategies: An Example from Children's Addition." *Journal of Experimental Psychology: General* 116 (3): 250.

Sirin, Selcuk R. 2005. "Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research." *Review of Educational Research* 75 (3): 417–53.

Thomson, Sue. 2018. "Achievement at School and Socioeconomic Background – an Educational Perspective." *Npj Science of Learning* 3 (5): 1–2.

Thornton, Carol A. 1990. "Solution Strategies: Subtraction Number Facts." *Educational Studies in Mathematics* 21 (3): 241–63.

Torney-Purta, Judith, Rainer Lehmann, Hans Oswald, and Wolfram Schulz. 2001. *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen.* ERIC.

Trautwein, Ulrich, and Oliver Ludtke. 2009. "Predicting Homework Motivation and Effort in Six School Subjects: The Role of Person and Family Characteristics, Classroom Factors, and School Track." *Learning and Instruction* 19 (3): 243–58.

Trautwein, Ulrich, Oliver Ludtke, Inge Schnyder, and Alois Niggli. 2006. "Predicting Homework Effort: Support for a Domain-Specific, Multilevel Homework Model." *Journal of Educational Psychology* 98 (2): 438–56.

Van De Walle, J. A., K. S. Karp, J. M. Bay-Williams, J. A. Wray, and E. T. Brown. 2007. "Elementary and Middle School Mathematics: Teaching Developmentally."

Van Mier, Hanneke I, Tamara MJ Schleepen, and Fabian CG Van den Berg. 2018. "Gender Differences Regarding the Impact of Math Anxiety on Arithmetic Performance in Second and Fourth Graders." *Frontiers in Psychology* 9.

Vos, Nienke, Henny Van Der Meijden, and Eddie Denessen. 2011. "Effects of Constructing Versus Playing an Educational Game on Student Motivation and Deep Learning Strategy Use." *Computers & Education* 56 (1): 127–37.

Weiner, B. 1986. *An Attributional Theory of Motivation and Emotion.* New York: Springer Verlag.

White, W. 1906. "Reviews Before and After Vacation." *American Education* 10 (3): 185–88.

Woodward, John. 2006. "Developing Automaticity in Multiplication Facts: Integrating Strategy Instruction with Timed Practice Drills." *Learning Disability Quarterly* 29 (4): 269–89.

Zaromb, F., R. M. Adler, K. Bruce, Y. Attali, and J. Rock. 2014. "Using No-Stakes Educational

Testing to Mitigate Summer Learning Loss: A Pilot Study." *ETS Research Report Series* 2014 (2): 1–17.

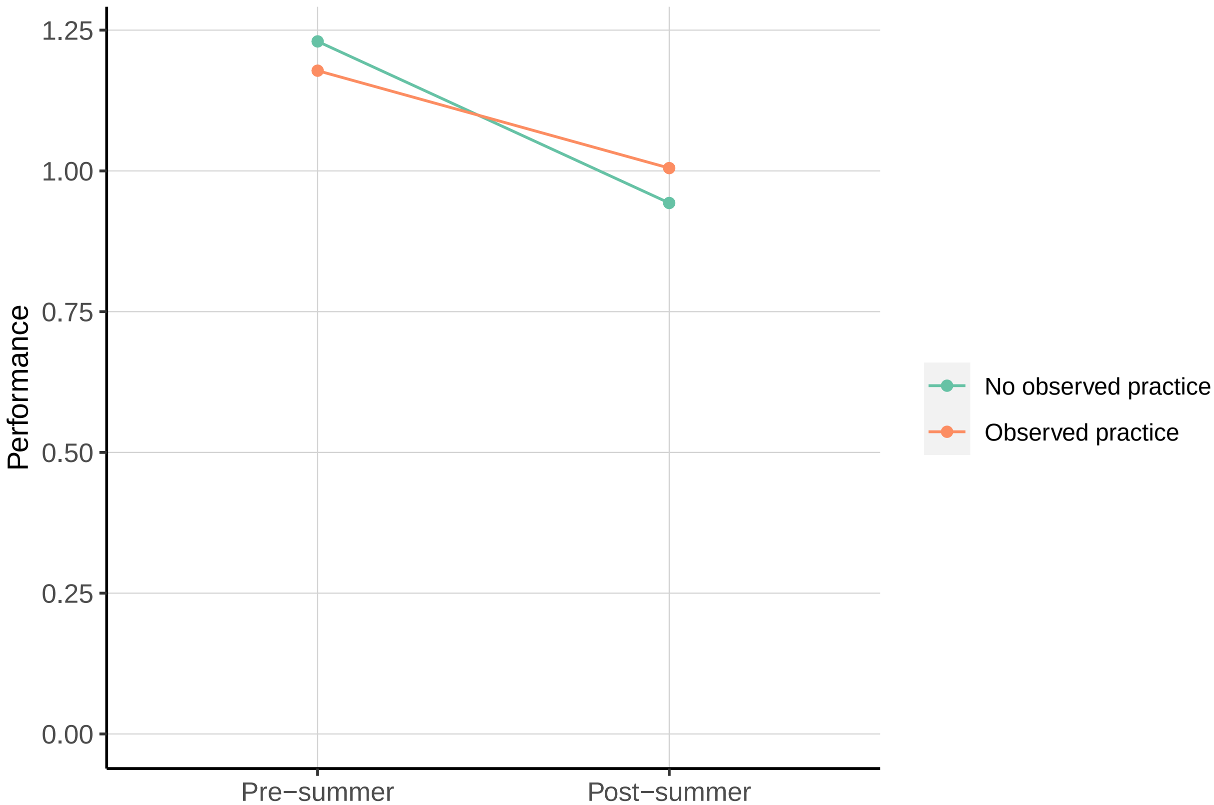

Figure 1: Summer learning loss and observed practice

Table 1: Pre-post math performance by gender

Gender	Domain	Pre	Post
Female	Addition	40.6	36.3
	Subtraction	41.2	35.6
	Multiplication	36.8	24.7
	Division	25.2	20.1
	Total	31.5	24.1
Male	Addition	49.4	46.2
	Subtraction	53.2	46.4
	Multiplication	37.3	27.9
	Division	28.7	24.4
	Total	38.4	31.3

Table 2: Pre-post total math performance by school

Period	Minimum	Q1	Median	Mean	Q3	Max
Pre summer-holiday	7.0	23.0	33.6	33.0	42.3	78.0
Post summer-holiday	5.5	15.4	25.4	26.0	34.2	70.5

Table 3: Pre and post math correlations across domains

		Addition	Subtraction	Multiplication
Pre	Addition	1.00		
	Subtraction	0.84	1.00	
	Multiplication	0.67	0.65	1.00
	Division	0.57	0.58	0.61
Post	Addition	1.00		
	Subtraction	0.83	100	
	Multiplication	0.64	0.63	1.00
	Division	0.60	0.61	0.63

Table 4: Home-based practice metric relations

		Self-reported school practice		Self-reported online practice		Self-reported math practice	
		Yes	No	Yes	No	Yes	No
Self-reported online practice	Yes	65	34				
	No	35	66				
Self-reported math practice	Yes	77	30	75	50		
	No	23	70	25	50		
Observed online practice	Yes	35	16	15	0	36	20
	No	65	84	85	100	64	80

Note: Columns report percentages. Percentages add up to 100% within a quadrant and column.

Table 5: Correlation table

	1	2	3	4	5	6	7	8	9
1. Performance	1.00								
2. Time	-0.13	1.00							
3. Perceived practice school	-0.00	0.00	1.00						
4. Perceived practice online	0.02	0.00	0.42	1.00					
5. Perceived practice math	0.14	0.00	0.45	0.27	1.00				
6. Observed practice online	-0.00	-0.00	0.15	0.31	0.09	1.00			
7. Perceived competence	0.44	0.01	0.01	0.03	0.07	-0.01	1.00		
8. Gender	-0.13	-0.00	0.07	0.00	0.11	0.03	-0.14	1.00	
9. SES	0.21	0.00	0.04	0.02	0.09	0.03	0.02	-0.07	

Table 6: Home-based practice models

	(1) Reported school practice		(2) Reported online practice		(3) Reported math practice			(4) Observed online math practice		
	Beta	S.E.	Beta	S.E.	Beta	S.E.	OR	Beta	S.E.	OR
Intercept	-0.08*	0.05	-0.01	0.05	0.36***	0.09	1.43	-2.44***	0.17	0.09
Reported math competence	0.02	0.03	0.02	0.03	0.17**	0.07	1.19	-0.04	0.11	0.96
Gender	0.17**	0.03	0.02	0.07	0.55***	0.14	1.73	0.22	0.23	1.25
SES	0.05	0.03	0.02	0.03	0.20***	0.07	1.23	0.13	0.11	1.14
R^2-adj	0.01		0.00							
AIC					1193.9			565.61		
Observations	934		934		934			934		

Note: Reported betas expressed as effect sizes. Models (1)-(2) estimated using OLS, models (3)-(4) estimated using logistic regression
OR = odds ratio. *** p<0.01; ** p<0.05; * p<0.1

Table 7: Summer learning loss

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Intercept	1.11***	0.077	1.23***	0.083	1.38***	0.078	1.38***	0.079	1.38***	0.078	1.28***	0.084	1.38***	0.079	1.23***	0.078
Post-summer			-0.260***	0.024	-0.261***	0.025	-0.261***	0.025	-0.261***	0.025	-0.281***	0.032	-0.270***	0.025	-0.287***	0.034
Reported school practice							-0.042	0.027							-0.096***	0.032
... × post-summer							0.014	0.015							0.012	0.017
Reported online practice									-0.003	0.027					0.008	0.031
... × post-summer									-0.001	0.015					-0.018	0.017
Reported math practice											0.171***	0.058			0.260***	0.065
... × post-summer											0.031	0.031			0.025	0.035
Observed online math practice													-0.056	0.096	-0.052	0.100
... × post-summer													0.103**	0.051	0.114**	0.053
Reported math competence					0.159***	0.016	0.159***	0.016	0.159***	0.016	0.157***	0.016	0.158***	0.016	0.156***	0.016
Gender					-0.213***	0.053	-0.208***	0.053	-0.213***	0.053	-0.234***	0.053	-0.213***	0.053	-0.230***	0.053
SES					0.216***	0.057	0.217***	0.057	0.216***	0.057	0.205***	0.055	0.215***	0.057	0.203***	0.056
Time level SD	0.362***	0.008	0.298***	0.007	0.308***	0.008	0.308***	0.008	0.308***	0.008	0.308***	0.008	0.307***	0.008	0.307***	0.008
Student level SD	0.823***	0.021	0.836***	0.021	0.758***	0.020	0.757***	0.020	0.758***	0.020	0.755***	0.020	0.758***	0.020	0.751***	0.020
School level SD	0.459***	0.060	0.497***	0.090	0.440***	0.007	0.444***	0.015	0.440***	0.007	0.430***	0.028	0.440***	0.025	0.435***	0.026
School-level post-summer SD			0.119**	0.054	0.125**	0.051	0.126**	0.049	0.125**	0.050	0.126**	0.051	0.126**	0.051	0.127**	0.049
Deviance	3883.6		3557.3***		3439.6***		3437.1		3439.6		3427.6***		3432.4**		3412.6***	
Observations	1868		1868		1868		1868		1868		1868		1868		1868	

Note: Reported betas expressed as effect sizes. Deviance statistics test relative to most complete nested model.

*** $p<0.01$; ** $p<0.05$; * $p<0.1$

Time, student, and school-level SDs capture random intercepts are these respective levels. The school-level post-summer SD. captures the random slope coefficient of the the post-summer indicator variable at the school-level.