Primary school mathematics during the COVID-19 pandemic: No evidence of learning gaps in adaptive practicing results

Martijn Meeter

LEARN! Research Institute, Vrije Universiteit Amsterdam, V.d. Boechorststraat 7, 1081 BT Amsterdam, the Netherlands

A B S T R A C T

Background: The COVID-19 pandemic induced many governments to close schools for months. Evidence so far suggests that learning has suffered as a result. Here, it is investigated whether forms of computer-assisted learning mitigated the decrements in learning observed during the lockdown.
Method: Performance of 53,656 primary school students who used adaptive practicing software for mathematics was compared to performance of similar students in the preceding year.
Results: During the lockdown progress was faster than it had been the year before, contradicting results reported so far. These enhanced gains were correlated with increased use, and remained after the lockdown ended. This was the case for all grades but more so for lower grades and for weak students, but less so for students in schools with disadvantaged populations.
Conclusions: These results suggest that adaptive practicing software may mitigate, or even reverse, the negative effects of school closures on mathematics learning.

1. Introduction

While the COVID-19 pandemic is in first instance a medical emergency, it has had vast consequences in many sectors. For education, lockdowns included the closure of schools in many countries [1,2]. While schools have typically moved to forms of distance learning, it has now become clear that the school closures have led to decrements in learning [3,4]. Moreover, these decrements were projected to be unevenly distributed, with students from less privileged backgrounds hit harder than others [5,6].

Most studies of learning during periods of school closures have analyzed standardized test data of primary school students. These studies have generally found that, on average, students had lower achievement at the end of a semester including the school closures than their peers did in the same period in previous years. A very large American study of 4.4 million grade 3–8 students found performance on the MAP Growth assessments at normal levels in October 2020 (i.e., after school closures), but math scores in the same assessments some 5–10 percentile scores lower than in previous years [7]. Decrements were larger for mathematics scores than for English scores, and larger for lower grades (i.e., 3–5) -although also higher grades (6–8) did not perform as well in 2020 than they did in earlier years. Two studies done on Dutch students compared standardized tests administered just before

a lockdown period (February 2020) to performance after an eight-week school closure period (June 2020). One study, using data from 350,000 students, concluded that the gap in achievement compared to earlier years was of a size (3 percentile points) consistent with that the weeks of online education had been a vacation in which no learning had occurred [6]. Another study, using data from 110.000 students [8], showed that decrements were seen for all grades and all levels of prior proficiency, most strongly for reading comprehension but also for mathematics and spelling. A study of grade-6 exam results from 402 of Belgium's Dutch-language schools showed a drop equal to 0.2 standard deviations compared to previous years in mathematics, and a drop equal to 0.3 standard deviations in language scores [9].

School closures were also feared to lead to an increase in educational inequality [5]. Findings in this regard were somewhat mixed, depending on what data was available to correlate with test scores. Three studies [6–8] noted that decrements in achievement were not correlated with previous attainment – in other words, student who had lower scores preceding the lockdown did not suffer more during it. At the school level, two studies [6,9] observed strong variation between schools in how much their students had suffered from the lockdown, which Maldonado and De Witte [9] could link to the student population: schools with more disadvantaged students in their population were more likely to show decrements in results after the school closures. Engzell et al. [6]

did not have such school-level data, but they could couple results to parental educational attainment. They showed that in their sample, students with parents with less educational attainment had learning decrements that were up to 55% larger than students with highly educated parents. Kuhlfeld et al. [7] could not confirm any of these patterns, but they did see worrying signs of mounting inequality in their attrition analyses: Compared to other years, many more students with a disadvantaged background were not tested at all.

All studies discussed above rely on a comparison of results from standardized tests from 2020 with previous years. The use of such data from standardized tests has several advantages. They tend to be mandatory for schools and students, allaying fears about selection. Moreover comparability with previous years is usually guaranteed by the rigorous testing procedures. However, relying on such tests also has disadvantages. First, they lack temporal precision, with typically one or two measurements a year. This means that what is observed is a decrement in achievement *after* a lockdown, not a decrement in learning *during* the lockdown. While it seems a small step to infer lower learning *during* a lockdown from lower achievement *after* one, such an inference is not entirely justified. In particular, any effect from the period of school closures cannot be disentangled from the period when schools had reopened, which is typically also part of the interval in between two testing occasions. This is problematic because after reopening, education was in all likelihood not equal to how it would have been in a normal year.

Second, and related, the critical test in the comparison was administered in June or October 2020, when schools were still heavily affected by the COVID-19 pandemic. In all studies it was noted that not all schools had administered the test, or not to all students, and that nonresponse was not randomly distributed [6,7,9]. Moreover, the setting of the test may have been abnormal, schools may have prepared less or differently for it than in other years, and students may have worked differently than they would otherwise have (e.g., because emotional upheaval as a results of the pandemic).

Here, we present results from a different source of data, adaptive practicing software. Such software typically computes proficiency scores on a fine-grained time scale (up to real-time, updated after every exercise), which allows us to plot learning throughout the whole school closure period and the period thereafter. Two other studies have used similar data sets, one looking at online mathematics learning in German primary school students [10], and at online foreign vocabulary learning in secondary school children [11]. Both found no evidence of learning decrements during the lockdown, which seem to be at odds with the findings reported above. We used similar data to study two questions:

- If learning decrements can be detected, how do they build up during the school closures, and the period immediately thereafter?
- Is there evidence of increasing inequality in scores, either as a function of prior learning or of background, during the school closures?

2. Methods

2.1. Context

Snappet is a digital learning environment that is primarily used for teaching and adaptive practicing in mathematics, language and spelling within classroom contexts. It is aimed at primary schools, and is used by a sizeable number of schools in both the Netherlands and Spain. Only Dutch data was used, and only data related to mathematics.

Snappet comes pre-installed on tablets that schools can hand out to their students. It is an environment in which students can practice mathematical skills, mostly with exercises that require them to give numeric answers but also some multiple-choice questions. Students receive immediate feedback on each exercise. Data from each student is then collected on a dashboard for the teacher, who can use real-time

data to adjust their instruction or give personal feedback to individual students [12]. Internally, Snappet computes estimates of student achievement using item response theory, using the so-called Rasch model [13]. There are computed both at the level of individual skills and of general mathematical achievement. Item response theory is a framework where the difficulty of each item (usually referred to, for item i, as δ_i) and the 'trait score' of each respondent (referred to, for respondent n, as β_n) are estimated on the basis of the responses given [14]. Within Snappet, the 'trait score' is not a fixed trait, but instead achievement as can be derived from recent performance, with the estimate updated after each item. Each Sunday morning these estimates were stored in a database, which is the raw data used in this study.

Snappet contains pools of exercises for each mathematics learning objective taught in primary school (item pools were relatively stable for the two years covered in this study). Teachers select a learning objective for a lesson, and then typically select a few exercises that are performed by all students in class, and discussed collectively. After that, students work at their own level: exercises are selected so that, given the student's level of achievement, the likelihood of answering them correctly is within a fixed bandwidth.

2.2. Participants

Of the 6333 primary schools in the Netherlands, around a third is currently using Snappet, of these 810 (36%) consented to using their data on a pseudonymous basis for scientific research.

To have some handle on population in the schools, we used the 'disadvantaged population' scores computed by the Dutch national bureau of statistics (CBS) for the purpose of disbursing remediation funds to schools with many students from disadvantaged households. The score is derived from a regression, predicting standardized test scores from demographic characteristics. Variables that predict lower results are then added to the score of a school, such as students stemming from household with low parental educational attainment, low income, migration background, single-parent household, etc. These scores were subdivided into deciles (for 12% of consenting schools, this index was not available). Fig. 1 shows the proportion of schools that fall within each decile, both of all Snappet schools and the schools that consented to data use. The distribution of consenting schools over deciles did not differ from all Snappet schools (χ^2(9)=6.35, p=.70), but it did from all Dutch schools (χ^2(9)=23.39, p=.005) due to a slight overrepresentation of schools with a more disadvantaged population.

Within the participating schools, 100,471 students used Snappet sufficiently (i.e., for 16 weeks of the year, see below) to have achievement estimates for the periods under study. No clear estimate can be given of the number of students who did not meet the inclusion criteria, but consideration of the number of classes included (between 2200 and 2500 in both years) and the average number of students per class in Dutch primary schools (23.4) suggest that only around 2% of students were excluded. Due to schools using Snappet in some grades but not others, or starting and stopping to use it for different cohorts of students, there was sizeable variation in student numbers per grade (see Table 1). Few schools used Snappet in first grade (over the two years, 446 students) – this grade was therefore dropped in the analyses. There was some increase in student numbers from school year 2018–19 to 2019–20, probably resulting from the fact that some consenting schools started using Snappet only in 2019–20 (while schools that had used it in 2018–19 but not in 2019–20 were not asked for their consent in using the data).

2.3. Data analysis

Data analysis was run on Snappet servers with a script provided by the researcher and a student assistant. They did not have access to raw data, out of privacy concerns. Data was used from the schools that consented to pseudonymous use for research.
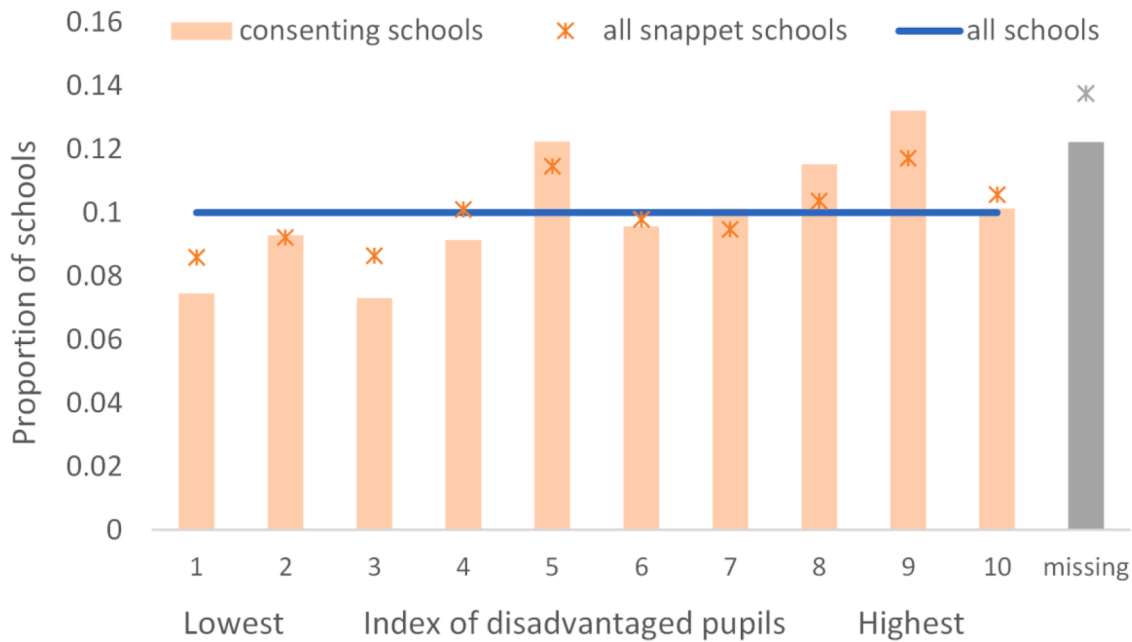
Fig. 1. Distribution of schools consenting to data use as a function of the index of disadvantaged students in their population. All Dutch primary schools were subdivided into deciles on the basis of this index (with the first decile having very few disadvantaged students, and the 10th many). Missing refers to the proportion of schools for which this index was missing – all other proportions were computed with reference to the number of schools where the index was not missing.

Table 1
Included students using Snappet in each grade, in school years 2018–19 and 2019–2020, and the same periods in school year 2018–19.

Grade	2018–19	2019–20
2	6845	7755
3	9633	11,975
4	12,722	12,614
5	6338	14,979
6	11,277	6333
Total	46,815	53,656

The weekly estimates of mathematics achievement were divided, for the school years 2018–19 (control year) and 2019–20 (year of COVID-19), into three periods: the pre-lockdown period (i.e., from the start of the school year up to March 16), lockdown period (March 14 – May 11th), and post-lockdown period (from May 11th up to the end of the school year). Students were included in the analyses when they had at least 16 weekly estimates (i.e., who had used Snappet for at least 16 weeks, or 40% of the year), were in classes with at least 10 users (to exclude classes in which Snappet was used either only for remediation or as optional material), and had at least two data points in each of the three periods.

In our main analysis, we compared the achievement estimates for students who used Snappet in 2019–20 with those who used it in 2018–19, for each of the three periods. We averaged the weekly achievement estimates (i.e., parameter β_n of the Rash model) per period resulting in three averages per student. Since the data has a hierarchical structure, with students nested within schools, we performed a linear mixed model (LMM) analysis with random intercepts included for both schools and students, and as fixed factor the three periods (before, during and after lockdown). This analysis was repeated for each grade.

Exercises within Snappet are related to learning objectives. It might be possible that students, during the lockdown, worked on a few learning objectives that they would then master to a high degree, artificially inflating their achievement estimates. To analyze whether students covered the same breath of material during the lockdown as they would have in previous years, we analysed the number of learning objectives broached by the students during the lockdown, and during the

same period in the control year (2018–19). A learning objective was counted as broached when the student performed one or more items related to it. Since the accessibility of learning objectives is under teacher control, this would indicate that the teacher had made the learning objective available, and thus deemed suitable for the student. A *t*-test was done per grade comparing the number of sets done by the students in the two years. We repeated the LMMs for only the lockdown period with the number of sets finished as a fixed-effect covariate to control for any differences in the materials covered in the two years.

To separate learning from achievement, a linear regression line was fit, for each student and each period separately, on the weekly achievement estimates within that period. This led to, for each student and period, two coefficients: an intercept that indicated the starting level of the student, and a slope coefficient that indicated the average increase in achievement per week (i.e., learning). To measure the attainment at the closing of a period, we used the regression coefficients to compute an estimated endpoint of the regression line for each individual student.

Students were then split into three equal bins on the basis of their results in the first half of the year. This was done on a per-class basis, to ensure that any effect of bin was not confounded with school- or class-level differences (binning the whole cohort in one go yielded only numerically different results). For each period, ANOVAs were performed with either learning or attainment at the end of the period as the dependent measure, and year and student bin as fixed factors.[1]

Both grade and school bin, defined by the decile of 'disadvantaged population' scores shown in Fig. 1, were also investigated. No omnibus analysis was attempted since the large data set would have resulted in significant effects that would have been difficult to interpret given the many levels of both variables. Instead, for each school bin and grade, Cohen's d was computed as an effect size (as the difference in mean learning or achievement between 2018–19 and 2019–20, divided by the pooled standard deviation), Then, 99% confidence intervals were

---

[1] LMMs using these coefficients failed to converge, which is why the main analysis was done with simple averages of the weekly estimates and analyses with student and school bins with ANOVAs.

computed around each effect size. To pinpoint results to specific bins or grades, confidence intervals for the different school bins and grades were compared.

2.4. Ethics

The study was conducted under a protocol approved by the local Institutional Review Board. The data cannot be made publicly available due to privacy restrictions.

3. Results

Fig. 2 shows average estimated attainment as a function of school week and grade, separately for the year of COVID-19 and the control year 2018–19. For most grades there was a slight advantage of 2018–19 students over their 2019–20 peers before the lockdown. However, once the lockdown started, the two years started to diverge. Learning was stronger for the lockdown year than the year before, and this effect remained when the lockdown ended, although the lines of the two years converged towards the end of the year.

Table 2 shows the results from the LMM analysis, performed separately for each grade, of the mean weekly estimated attainment for the three periods. The after-lockdown period was used as reference level, and for each grade except the sixth there was a positive effect of year, meaning higher attainment in the year of COVID-19 than in the control year. Coefficients for the "before lockdown" and the "year-before lockdown" interaction were negative, reflecting that attainment was lower at the start of the school year than at the end, and that students had started out at a slightly lower level in 2019–20 than in 2018–19. For the lockdown period, the "year:lockdown" interaction was still negative in each grade (indicating greater advantage for the COVID-19 year in the after-lockdown period than in the lockdown period), but much smaller than that of the before-lockdown periods, reflecting the stronger learning during the lockdown in 2019–20.

The average number of learning objectives broached during the lockdown varied between 6.6 (grade 2 in 2019–20) and 8.5 (grade 5, 2019–20). It was between 6% and 10% lower in 2019–20 than in 2018–19 for all grades except grade 5, where it was 5% higher ($t>6$; $p<.001$ for all grades). LMMs for only lockdown period with year and number of learning objectives broached as variables showed that the number of learning objectives broached was positively related to the

weekly attainment estimates ($z>18$, $p<.001$ for all grades). However, difference in learning objectives broached did not explain stronger learning in the COVID-19 year – in the LMM, coefficients that expressed the effect of year somewhat higher than those reported in Table 2.

Moving to the learning and attainment coefficients that resulted from linear regressions per participant, Fig. 3 shows the difference in learning and attainment coefficients between 201819 and 2019–20 expressed as an effect size (Cohen's d). Panels a and c show average learning and attainment as a function of student bin and period. Pre-lockdown, there was slightly stronger learning in 2019 than 2018 (positive effect size for all three bins), in which 2019–20 students were making up for a lower level of attainment at the start of the year (negative effect size for pre-lockdown period in panel c, of attainment). During the lockdown learning was much stronger in 2019–20 (positive effect sizes in panel a), leading to higher attainment at the end of the period (positive effect sizes in panel c). Higher attainment was still visible at the end of the post-lockdown period, even though weaker learning in this period (negative effect sizes for the post-lockdown period in panel a) mitigated some of it.

These patterns were confirmed with a set of ANOVAs per period, with student bin and year as fixed factors. For the pre-lockdown period, there was a main effect of year on learning, in favor of 2019–20, $F(1, 100,456)=94,05$, $p<.001$. There was also a main effect of student bin, reflecting somewhat stronger learning for the lowest bin, $F(2, 100,456)=91.13$, $p<.001$, but no interaction between the two factors, $F(2, 100,456)=2.09$, $p=.13$. For the lockdown period, there was a main effect of year in favor of 2019–20, $F(1, 100,456)=6844$, $p<.001$, of student bin, again in favor of the lower bin, $F(2, 75,459)=1052$, which was qualified by an interaction between year and student bin, $F(2, 100,456)=368$, $p<.001$): stronger learning during the lockdown was particularly pronounced for the lowest bin, with the higher bins benefiting progressively less (see Fig. 3, panel a). In the post-lockdown period, effects were reversed. The main effect of year, $F(1, 100,456)=1128$, $p<.001$, now favoured 2018–19, while the interaction between student bin and year, $F(2, 100,456)=53.19$, $p<.001$, showed that the higher bin progressed more strongly after the 2020 lockdown that the lower bins. Only the main effect of student bin, $F(2, 100,456)=56.54$, $p<.001$, continued to favor the lower bin.

Moving to attainment, an ANOVA with student bin and year as fixed factors for the pre-lockdown period showed a main effect of year, $F(1, 100,456)=34.38$, $p<.001$, reflecting somewhat better attainment in
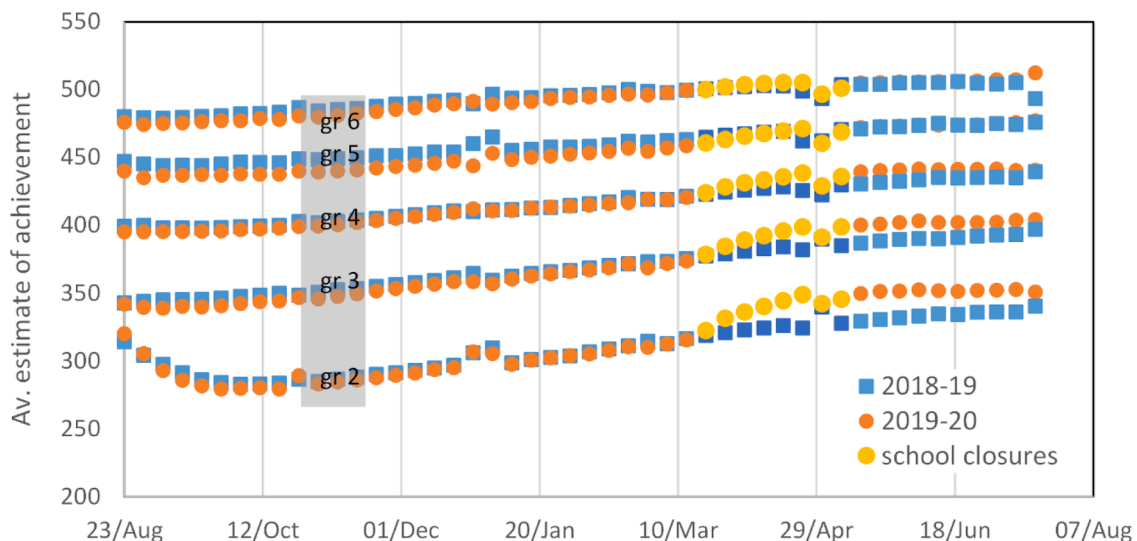
Fig. 2. Weekly estimates of achievement for 2018–19 and 2019–20, separately for grades 2 to 6. The lockdown period is marked in yellow. Most students start using the software in grade 2, and it takes one to two weeks for the estimates to go from their default starting level to the true level of the student. Since the start of the school year is staggered in the Netherlands, this effect is smeared out over time. Weeks with sudden jumps up or down (e.g., the weeks underneath the gray box, the first May week and the last week) are vacation weeks in which only few students use the software.

Table 2

LMM coefficient, standard errors (SE) and associated z value for the year (2019–20=1), and periods. Since the "after lockdown" period was used as reference level (which has highest attainment since it came last), coefficients for the other two periods and for year/period interactions are negative. *** indicates p<.001.

	grade 2			grade 3			grade 4			grade 5			grade 6		
	coeff	SE	z	coeff	SE	z	coeff	SE	z	coeff	SE	z	coeff	SE	z
Constant	337.5	15.5	21.7***	389.2	22.7	17.1***	428.6	53.6	8***	465.9	235.9	2***	499.6	344.8	1.40
Year	15.3	0.70	22***	10.0	0.57	17.5***	4.4	0.52	8.4***	1.6	0.69	2.3***	1.4	0.79	1.80
before lockdown	−132.3	2.53	−52.2***	−137.8	2.49	−55.2***	−106.1	2.64	−40.2***	−107.9	3.17	−34***	−77.8	4.05	−19.2***
lockdown period	−14.0	2.47	−5.7***	−22.6	2.46	−9.2***	−16.0	2.61	−6.1***	−22.7	3.15	−7.2***	−14.0	4.01	−3.5***
year:before lockdown	−18.0	0.34	−52.2***	−14.1	0.26	−55.3***	−8.3	0.21	−40.2***	−8.2	0.24	−34.1***	−4.2	0.22	−19.3***
year:lockdown period	−1.9	0.34	−5.7***	−2.3	0.25	−9.2***	−1.3	0.20	−6.1***	−1.7	0.24	−7.2***	−0.8	0.22	−3.5***
Var(school)	4.7			10.1			6.9			1.2			3.9		
Var(student)	2339.6			2259.1			2352.5			2524.3			3047.6		

2018–19 than in 2019–20, with a main effect of student bin, $F(1, 100,456)=21,151$, $p<.001$ (obvious since bins were defined on the basis of attainment), and no interaction between these two factors, $F(1, 100,456)<1$. After the lockdown, the main effect of year had reversed, $F(1, 100,456)= 102$, $p<.001$ with better attainment, now, for 2019–20. The main effect of student bin was still there, $F(1, 100,456)= 19,158$, $p<.001$, and now there was an interaction between student bin and year, $F(1, 100,456)= 23.79$, $p<.001$: attainment was especially higher in 2019–20 for the students in the lowest bin (see Fig. 3, panel c). Due to missing information no ANOVA of attainment at the end of the post-lockdown period was possible, but Fig. 3, panel c shows that for both the lowest and the middle student bin, the confidence intervals around the effect size show that attainment was still higher for these groups than for their peers in the preceding year. This was not the case, however, for the highest bin, where attainment at the end of the year was not different from that of peers in the preceding year.

Fig. 3, panel b shows the effect on learning during lockdown and post-lockdown split out per grade. Stronger learning during lockdown was especially prominent for the lower grades, and less so for the higher grades (this can also be seen in Fig. 2). However, the slowing of learning post-lockdown was also especially prominent for those grades. The suggestion of a trade-off between strong learning during the lockdown and weak learning after it was further strengthened by the small, negative correlation between learning estimates during and after the lockdown ($r=−0.25$, $p<.001$). No such negative correlation was present between learning estimates for the pre-lockdown and lockdown periods ($r = 0.06$, $p<.001$). We come back to this in the discussion.

Fig. 3, *panel d* shows attainment at the end of the school year as a function of school bin, where schools were binned on the basis of having a disadvantaged student population. Although there is some variation, stronger learning in the COVID-19 year was especially evident for the school bins with less disadvantaged populations. The two bins with the most disadvantaged populations were also two of the ones where the mean effect was not significantly different from 0 (as derived from the confidence intervals shown in the figure).

One possible explanation for the stronger progress during the lockdown in 2019–20 is simply more practicing. Fig. 3, *panel e* shows that while practice in the two years was about equal for the pre-lockdown period, students finished more exercises on Snappet during and after the lockdown in 2019–20 than their peers had done in 2018–19. However, the different was larger in the post-lockdown as opposed to the lockdown period. In the lockdown period, practicing was some 15% higher than it had been in the same period in 2018–19. In the post-lockdown period the difference was 30%, mostly because usage dropped in that period in 2018–19 but not in 2019–20. High use correlated with stronger learning, though weakly so ($r = 0.15$, $p<.001$). This correlation was about the same strength for the pre-lockdown period ($r = 0.145$, $p<.001$ between pre-lockdown practice and learning), but was absent for the post-lockdown period ($r=−0.001$, $p=.07$).

4. Discussion

Several studies have shown, also for Dutch primary education [8,15], that, after the school closures, students had fallen behind compared to previous years. Here, using data from adaptive practicing software, we found no evidence of any learning decrements. During the period of school closures, students progressed more strongly than their peers in previous years had. This was especially true for younger students (grades 4 and 5), students who had weaker results before the lockdown, and students in schools with less disadvantaged student populations. These gains diminished in the weeks after the closures, when education returned to normal, but were still present. Although unexpected, the gains replicate those found in recent papers analyzing results from on-line learning in German primary school students [10], and of foreign vocabulary learning in secondary school children [11].

Neither of the latter two studied the time after the lockdown. Why
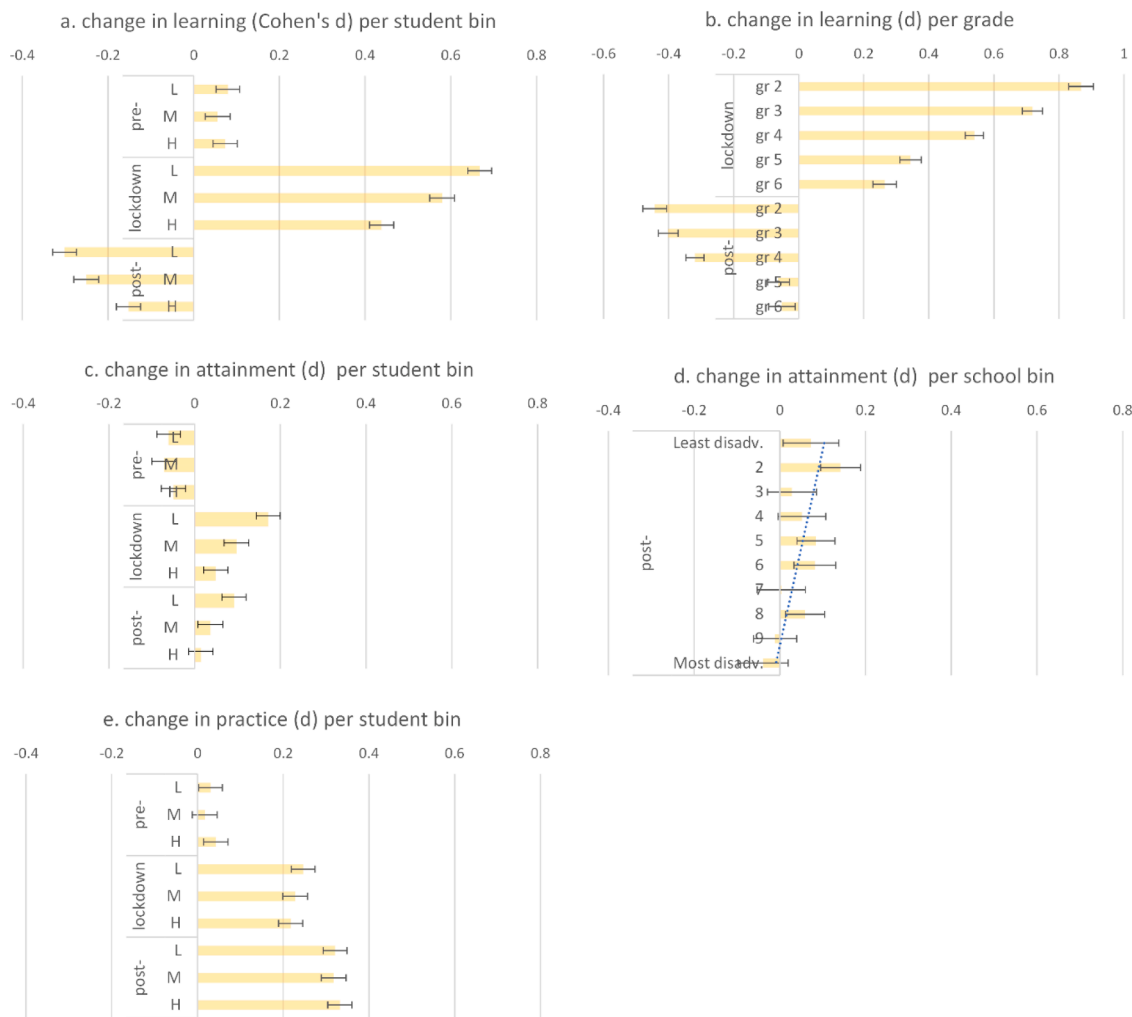
Fig. 3. Comparison of the COVID-19 year with the preceding control year (2018–19). Each bar shows an effect size (Cohen's d) computed from the comparison. The error bars denote the 99% confidence interval around the estimate. (a). Change in average weekly learning for the pre-, lockdown, and post-lockdown periods, split up for the lowest- (L), middle- (M) and highest-achieving students, as defined by their average performance in the pre-lockdown period. (b) Same as preceding panel, but now split up per grade. (c) Change in estimated attainment at the end of each period for each of the three student bins. (d) Same, but only for the post-lockdown period (i.e., end of the year) and binning based on school disadvantaged student population scores. In blue the line of best fit through the ten estimates. (e). Change in practice, as measured by the number of exercises finished in a week by students, again separately for each period and student bin.

the advantage for 2020 students dissipated during the return to normal education is unclear. It was not a result of less practice – in fact, number of completed exercises was higher in the COVID-19 year than the control year for this period. Education was messy in these weeks, with students first being taught in half classes for half of the week. This may have led to lower gains. Perhaps schools made a conscious choice to concentrate on student welfare or neglected skills.

Nevertheless, for lower grades and weaker students, some gains were still present at the end of the school year. There are several possible explanations for the contrast between these results and those of other studies:

4.1. Altered testing circumstances

it is possible that the decrements in standardized test scores observed in other studies do not truly reflect learning deficits. Instead, they may reflect altered testing circumstances, such less preparation for the tests than in other years. The test analyzed by Engzell et al. [6] and Lek et al. [8] is formative, which may mean that schools may have not put too much emphasis on optimal preparation, optimal administration or optimal concentration by their students. However, this would not explain why the results on a high stakes test from a neighbouring

country would yield very similar results [9].

4.2. Concentrating on the core

it may have been that the school closures induced teachers to focus on core skills that, they felt, had to be covered or could be covered at a distance better than other topics. This might have led to strong learning on those skills (the learning objectives targeted by the adaptive practicing) at the detriment of other skills that are not emphasized during online practice but were part of the curriculum. As it is known that practice is essential to learning [16,17], more practice may have led to more learning. Indeed, practice was more intense during the 2020 school closures than in the same period a year earlier (although the difference was not large, and did not align well with where the gains were seen). While this may explain the surprising finding of stronger learning in the lockdown than in the preceding year, it does not provide a good explanation for the mismatch between current findings and previous ones [6,8], since Snappet covers the same topics as standardized mathematics tests.

4.3. Reading comprehension

One difference that does exist between Snappet exercises and those in standardized tests is that those in standardized tests tend to include many word problems, for which good reading skills are needed [18]. Both Engzell et al. [6] and Lek et al. [8] showed stronger decrements in reading comprehension than in mathematics after the school closures (this was also found in Belgium but not the US). Perhaps those decrements in mathematics scores were at least partly a result of deficits in reading comprehension sustained by the school closures, and less of deficits in mathematical skills.

4.4. More effective teaching

The difference between the current results and those of Engzell et al. [6] and Lek et al. [8] may also reflect a true difference between schools relying on adaptive practicing software for mathematics teaching, and other schools. In particular, immediate feedback as offered by Snappet and similar software is known to be very effective in steering learning [19]. Indeed, the effectiveness of Snappet at boosting mathematics scores was shown in two large quasi-experiments [12,20]. Faber et al. [20] quantified it as 1.5 month of additional gain for primary school students using Snappet for four months. Schools in the current sample all relied at least partly on Snappet for mathematics teaching, while little is known about how schools in the sample of Engzell et al. [6] taught mathematics – presumably through of mix of methods that may not all have been as effective. Schools using Snappet may also have had the advantage that they could rely on methods of teaching that were well-adapted for distance education, while for schools relying on paper-and-pencil methods, the transition might have been more difficult.

4.5. More effective distance education

Benefits of adaptive practicing software and learning environments like it may be accentuated by school closures. The breakdown of routine and the lack of social contacts during school closures may lead to demotivation and problems keeping oneself at work, even in college students [21]. The current results show that students using adaptive practicing software increased the time spent practicing with it. It is unclear whether other forms of mathematics distance education lead to the same amount of practice, but one study suggests that they are not always very effective [22]. If home practice is more effective with adaptive practicing software like that of Snappet than in other forms, this may have been because of the rewarding nature of practice and seeing oneself improve, or the fact that education at home could rely on routines with Snappet tablets established in school.

4.6. Inequality

With regard to educational inequality, the current results replicate previous ones [6–8] in that students with lower prior results did not suffer more from the lockdown than other students. To the contrary, weak learners seemed to catch up to their more advanced peers during the school closures, to a larger extent than did peers in the preceding year. This replicates such findings of Spitzer and Musslik [10].

Part of this catching up was reversed as students returned to schools. An explanation for this may reside in differential effects of adaptive practice within the classroom. When Snappet is used in the classroom, more proficient students tend to benefit relative to students in classrooms where no adaptive practicing software was used [20]. In normal classrooms, those at the highest level of proficiency barely made any progress during a five-month study period. This relative lack of progress was much less pronounced in classes using adaptive practicing, purportedly because adaptive practicing allowed proficient students to practice at their own level [12,20]. However, what then explains the finding of strong progress for weaker students during the school closures? This remains to be explored, but a reason may be that adaptive practicing software mostly incorporates mastery learning [23], which may be more effective if unmoored from classroom routines.

While the results thus contain some good news, there are also indications for increased inequality as a function of student background. Replicating Maldonado and De Witte [9], we found that schools with more disadvantaged student populations were characterized by less strong learning gains than schools with more advantaged populations. This adds to strong evidence that unequal outcomes are more a function of student background than of prior attainment [6]. In a study performed in Great Britain, it was found that students from more disadvantaged backgrounds spent fewer hours learning at home during school closures than did their more fortunate peers [24], while a Dutch survey of parents found that parents with more education were more likely to state that they could help their children with distance education than parents with less education, even though both valued it equally [25]. Meanwhile, in the United States, it was found that schools with poorer student populations were more likely to close than schools with less poor student populations [26].

4.7. Limitations

The study has several strengths, such as that it shows the time course of learning during the school closures, but also several limitations. The analyses were concentrated on mathematics skills, while other studies (e.g., [8,9]) found larger decrements in reading comprehension than in mathematics in primary school students. Moreover, privacy concerns led to data analysis being less extensive than it would otherwise have been, since it had to be done via intermediaries. This resulted in, for example, not all analyses being done with linear mixed models. However, the multilevel linear mixed models that could be fit confirmed the findings from other analyses, suggesting that the main results are not dependent on the particular statistical analysis used.

Moreover, the schools who made their data available are not a random sample of all schools using adaptive practice, while those are in turn not a random sample of all primary schools in the Netherlands. A strong bias towards schools with privileged populations could be excluded (see Fig. 1), but less visible biases may exist and limit generalizability. Moreover, generalizability to other countries may be problematic.

5. Conclusion

Here, learning in mathematics was analysed during the pandemic-induced school closures of spring 2020 using adaptive practicing software. To our surprise, stronger learning was found during the school closures than in the year before. These gains were stronger for lower grades, and for students with weaker previous learning. However, while students in schools with more disadvantaged populations benefited, they benefited less. The study thus adds to those to suggest increased educational inequality as a result of the COVID-19 pandemic. However, more positively, it suggests that adaptive practicing software may be a way to attenuate learning losses due to school closures, or even reverse them.

Declaration of Competing Interest

None

Ethics

The study was conducted under protocol #2019–147, approved by the institutional review board of the Faculty of Behavioural and Movement Sciences of Vrije Universiteit Amsterdam.

References

[1] C. Hirsch, Europ's coronavirus lockdown compared Politico Europe (2020) 31–33, 2020.

[2] S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L.Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, T. Wu, The effect of large-scale anti-contagion policies on the COVID-19 pandemic, Nature 584 (7820) (2020) 262–267.

[3] K. Zierer, Effects of Pandemic-Related School Closures on Pupils' Performance and Learning in Selected Countries: a Rapid Review, Edu. Sci. 11 (6) (2021) 252.

[4] S. Hammerstein, C. König, T. Dreisoerner, A. Frey, Effects of COVID-19-Related School Closures on Student Achievement—A Systematic Review, PsyArxiv Preprints (2021).

[5] M. Kuhfeld, J. Soland, B. Tarasawa, A. Johnson, E. Ruzek, J. Liu, Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement, Educ. Res. 49 (8) (2020) 549–565.

[6] P. Engzell, A. Frey, M.D. Verhagen, Learning loss due to school closures during the COVID-19 pandemic, in: Proceedings of the National Academy of Sciences of the United States of America 118, 2021.

[7] M. Kuhfeld, B. Tarasawa, A. Johnson, E. Ruzek, K. Lewis, Learning during COVID-19: initial findings on students' reading and math achievement and growth, NWEA Research Brief, Northwest Evaluation Association (2020). Portland, OR.

[8] K. Lek, R. Feskens, J. Keuning, Het Effect Van Afstandsonderwijs Op Leerresultaten in Het PO, 2020 [The effect of distance education on learning results in primary education]Cito, Enschede, the Netherlands.

[9] J. Maldonado, K. De Witte, The Effect of School Closures On Standardised Student Test, KU Leuven Department of Economics, 2020. FEB Research ReportLeuven (Belgium.

[10] M.W.H. Spitzer, S. Musslick, Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic, PLoS ONE 16 (8) (2021), e0255629.

[11] M. van der Velde, F. Sense, R. Spijkers, M. Meeter, H. van Rijn, Lockdown Learning: changes in Online Foreign-Language Study Activity and Performance of Dutch Secondary School Students During the COVID-19 Pandemic, Front. Edu. 6 (294) (2021).

[12] I. Molenaar, C.A.N. van Campen, K. van Gorp, Onderzoek Naar Snappet; Gebruik En Effectiviteit, 2016. Kennisnet, Zoetermeer (Netherlands).

[13] G. Rasch, Probabilistic Models For Some Intelligence and Attainment Tests, The University of Chicago Press, Chicago, 1980.

[14] S.E. Embretson, S.P. Reise, Item Response Theory For psychologists, L. Erlbaum Associates, 2000. Mahwah, N.J.

[15] P. Engzell, A. Frey, M.D. Verhagen, Learning Inequality During the COVID-19 Pandemic, 2020. OSF preprint.

[16] M.W.H. Spitzer, Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study, Eur. J. Psychol. Educ. (2021).

[17] K.A. Ericsson, R.T. Krampe, C. Teschromer, The Role of Deliberate Practice in the Acquisition of Expert Performance, Psychol. Rev. 100 (3) (1993) 363–406.

[18] H. Korpershoek, H. Kuyper, G. van der Werf, The Relation between Students' Math and Reading Ability and Their Mathematics, Physics, and Chemistry Examination Grades in Secondary Education, Int. J. Sci. Math. Educ. 13 (5) (2015) 1013–1037.

[19] K. VanLehn, The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, Educ. Psychol.-Us 46 (4) (2011) 197–221.

[20] J.M. Faber, H. Luyten, A.J. Visscher, The effects of a digital formative assessment tool on mathematics achievement and student motivation: results of a randomized experiment, Comput. Educ. 106 (2017) 83–96.

[21] M. Meeter, T. Bele, C. den Hartogh, T. Bakker, R.E. de Vries, S. Plak, College students' motivation and study results after COVID-19 stay-at-home orders, PsyArxiv (2020), https://doi.org/10.31234/osf.io/kn6v9 p. https://doi.org/10.31234/osf.io/kn6v9.

[22] J.L. Woodworth, M.E. Raymond, K. Chirbas, M. Gonzalez, Y. Negassi, W. Snow, Y. Van Donge, Online Charter School Study, Center for Research on Educational Outcomes, 2015. https://credo.stanford.edu/sites/g/files/sbiybj6481/f/online_charter_study_final.pdf. Stanford, CA.

[23] S. Ritter, M. Yudelson, S.E. Fancsali, S.R. Berman, How mastery learning works at scale, in: Proceedings of 3rd ACM Conference on Learning@ Scale, 2016, pp. 71–79, https://doi.org/10.1145/2876034.2876039.

[24] N. Pensiero, A. Kelly, C. Bokhove, Learning Inequalities During the Covid-19 pandemic: How Families Cope With Home-Schooling University of Southampton, 2020. Southampton, UK.

[25] T. Bol, Inequality in homeschooling during the Corona crisis in the Netherlands. First results from the LISS Panel, SocArxiv (2020), https://doi.org/10.31235/osf.io/hf32q.

[26] Z. Parolin, E. Lee, Large Socio-Economic, Geographic, and Demographic Disparities Exist in Exposure to School Closures and Distance Learning, OSF Preprints (2020), https://doi.org/10.31219/osf.io/cr6gq.