Computers and Education Open 6 (2024) 100162

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

# Performance increases in mathematics within an intelligent tutoring system during COVID-19 related school closures: a large-scale longitudinal evaluation

Markus Wolfgang Hermann Spitzer [a,*], Korbinian Moeller [b,c,d]

[a] Department of Psychology, Martin-Luther University Halle-Wittenberg, Halle, Germany
[b] Centre for Mathematical Cognition, School of Science, Loughborough University, Loughborough, United Kingdom
[c] Leibniz-Institut fuer Wissensmedien, Tuebingen, Germany
[d] LEAD Graduate School and Research Network, University of Tuebingen, Germany

A R T I C L E I N F O

*Keywords:*
Mathematics
Online learning
COVID-19
Intelligent tutoring system
Large-scale data analysis
Longitudinal study

A B S T R A C T

The closure of schools due to COVID-19 disrupted learning routines of thousands of students, resulting in reported performance decreases, especially among lower-performing students. However, some studies on students' performance within intelligent tutoring systems (ITS) also found significant performance increases for times of school closures as compared to before. However, little is known about students' longitudinal performance trajectories within ITS. Accordingly, we evaluated longitudinal data from German students ($n \approx$ 2,700 students; $n \approx$ 5 million problems) enrolled in an ITS for learning mathematics from January 2017 until the end of May 2021 to investigate the effect of periods of school closures (first and second) on students' performance within the ITS during, between, and after school closures. We observed significant performance increases for both lower- and higher-performing students during, between, and after COVID-19 related school closures. Importantly, these improvements were more pronounced for lower-performing students compared to higher-performing students. Together, these results suggest that ITS may have helped to maintain mathematics learning, particularly for lower-performing students during COVID-19 related school closures and that these beneficial effects persisted at least for the following months when schools opened again. As such, the use of ITS for learning mathematics seems an appropriate approach for distance learning during times of crisis.

## Introduction

In 2020, the COVID-19 pandemic caused abrupt school closures worldwide [54]. In Germany, schools closed in mid March 2020, which necessitated prompt adaptations to teaching modes to distance learning (e.g., [13,25,28]). One approach to this was the increased reliance on digital resources such as intelligent tutoring systems (henceforth ITS), as such systems provide students with rapid learning aids when working through mathematical problems, such as feedback, hints, or additional problem sets adaptively assigned to students by the system in response to specific errors students committed to fill their knowledge gaps (e.g., [9,26,42,46]). As such, ITS served as a promising software for times during school closures and the demand for such ITS enabling distance learning led to a surge in their usage starting in mid March 2020 [35,47,53].

While numerous studies have examined the impact of COVID-19 related school closures on students' performance in general (for a meta-analysis and review see: [4]), only a few studies investigated the effect of school closures on students' performance within ITS in particular [35,48–50,53]. Most of the studies that utilized data from ITS compared different cohorts of students using an ITS either before or during school closures [35,49,53]. However, to the best of our knowledge, studies about longitudinal performance trajectories following up the same students using these platforms from before to during and after periods of school closure to investigate the effect of school closures on performance outcomes as well as the persisting effects of school closures are missing.

Therefore, we investigated the effects of school closures on longitudinal performance trajectories following a cohort of German students who already studied mathematics using an ITS before school closures

* Corresponding author.
*E-mail address:* markus.spitzer@psych.uni-halle.de (M.W.H. Spitzer).

https://doi.org/10.1016/j.caeo.2024.100162
Received 31 July 2023; Received in revised form 17 October 2023; Accepted 17 January 2024
Available online 19 January 2024

and kept studying with this ITS during and after school closures. Our analysis particularly focused on potential selective effects on lower- vs. higher-performing students as previous research from large-scale panel studies indicated selective performance decreases for lower-performing students (e.g., [44]), while research that considered data from ITS suggested performance increases—especially for lower-performing students (e.g., [50]).

In the following, we will first briefly review the existing literature on the effects of school closures from large-scale panel studies. We will then review further research on large-scale panel studies on potential longitudinal consequences of school closures, before we elaborate on research that utilized data from ITS to examine the effect of school closures on students' performance within ITS.

*Panel studies on students' performance changes during school closures*

A wealth of research has documented significant performance decreases across different countries resulting from school closures ([4, 10–12,14,17–21,24,28]; Kuhfeld et al., 2022; [31,32,36,39,40,44,45, 56,57]). Additionally, research stated that these performance decreases were particularly severe for mathematics ([4]; Kuhfeld et al., 2022; [44], 2022) with devastating performance decreases observed particularly for lower-performing [16,44]. For instance, Grewenig and colleagues (2020) reported that during school closures, lower-performing students in Germany spent only half the time learning compared to higher-performing students. In another study, Schult et al. [44] evaluated data from a German panel study ($n \approx 330{,}000$; grade 5; age range 10 to 11) and observed more pronounced performance decreases in mathematics for lower-performing students compared to higher-performing students. Similar results were also reported by Maldonado et al. [32] who reported more pronounced performance decreases in mathematics particularly for lower-performing students in a Flemish sample ($n \approx 10{,}000$; grades 4 and 6; age range 10 to 12). However, it should also be noted that studies on data from Australia [15], Finnland [30], and Denmark observed [5] no performance decreases during COVID-19 related school closures which may be attributed to varying COVID-19 policies between countries.

*Panel studies on longitudinal performance changes due to school closures*

Although a large body of research considered the impact of school closures on students' performance, thus far only a few studies thus far considered students' performance trajectories over a longer time period from before the pandemic to after school closures. For instance, Kuhfeld and colleagues (2022) considered a US sample of close to 5 million students and found that students' performance, in general, was still below pre-pandemic levels by spring 2021. Additionally, these authors found that students particularly struggled to learn mathematics since COVID-19 related school closures. Importantly, however, Kuhfeld et al. [29] compared different cohorts of students in a between-student design and contrasted the performance of two cohorts (pre- vs. post pandemic) with each other.

In another study, Schult et al. [43] evaluated the performance of all incoming students of fifth grade in September for pre-pandemic years (2017–2019), for September 2020 to examine performance 6 months after the first school closures, and also for September 2021 to evaluate potential effects persisting after 18 months since first school closures in the following cohort. The authors reported selective influences on students' performance in mathematics. Six months after school closures, they observed performance decreases in students' number knowledge (e.g., understanding, interpreting, and representing whole numbers) and basic arithmetic skills (e.g., applying and combining arithmetic processes). However, 18 months after school closures, the performance of the following cohort of incoming fifth grade students in number knowledge recovered to pre-pandemic levels. In contrast, basic arithmetic skills did not fully recover to pre-pandemic levels. Moreover, they found that the lowest 5 % (considering performance) of students showed severe performance decreases due to COVID-19 related school closures in September 2020 as well as in September 2021.

In summary, the reviewed studies provide evidence suggesting performance decreases due to COVID-19 related school closures and especially pronounced performance decreases for lower-performing students. While these studies considered data from large-scale panel studies, other research that utilized data from ITS seem to differ from those observed from in-person testing by indicating performance increases (e.g., [35,49,50,53,55]).

*Performance changes within intelligent tutoring systems during the first period of school closures*

Another set of research investigated the effects of school closures on students' performance outcomes within ITS [35,49,50,53,55]. While one study involving Swiss students ($n \approx 350{,}000$; grades 3 to 9: age range: 8 to 16) observed mixed results [53] other studies found performance increases instead of performance decreases during periods of school closures [35,49,50,55]. For instance, in a sample of Dutch students ($n \approx 130{,}000$; grades 7 to 10; age range: 12 to 16) who used an ITS for learning French, van der Velde (2021) found significantly more pronounced performance increases during the period of school closures as compared to before. Similarly, Meeter [35] conducted another study on Dutch students ($n \approx 100{,}000$; grades 2 to 6: age range: 7 to 11) who studied mathematics within an ITS and also observed significant performance increases during school closures as compared to before. Importantly, however, these studies also used a between-student design comparing different cohorts of students using ITS either before or during school closures. In another sample, considering a sample of German students ($n \approx 16{,}000$; age range: 10 to 16), Spitzer et al. [49] evaluated data not only from the first period of school closures but also from the second period of school closures in Germany and observed performance increases in a cohort of students who worked on fraction, algebra, or percentages problems the first and second COVID-19 related school closures compared to another cohort of students who worked on these problems during the same months but in the previous three years before COVID-19 related school closures. However, these improvements were only observed for students who got problem sets assigned by their teachers and not if they selected problem sets on their own and if these problem sets were assigned in small chunks and not in larger bundles.[1]

While these studies reported between-student analyses, Spitzer and Musslick [50] followed a sample of German students ($n \approx 2500$; age range: 10 to 16) who studied mathematics with another ITS longitudinally. In particular, they evaluated the performance of the same students (within-student design) during school closures in Germany compared to the same time period in the year before with no school closure. Results indicated significant performance increases during school closures as compared to the year before. Moreover, Spitzer and Musslick [50] observed that lower-performing students showed particular performance increases within the ITS during school closures.

Together, these results provide evidence suggesting that learning within ITS may have been beneficial, especially for lower-performing students who had been observed to struggle most during COVID-19 related school closures in other panel studies. However, it is important to evaluate the performance trajectories of lower- and higher-performing students not only during school closures (as compared to before) but also after periods of school closures (as compared to before

[1] Note that several assignment possibilities exist within Bettermarks. Students may self-select problem sets (a problem set comprises nine single problems; see Methods), or students can get problem sets assigned by their teachers. Teachers can either assign single problem sets from books to their students (which we refer to as small chunks), or, alternatively, assign all problem sets of a book (which we refer to as larger bundles).

school closures) to better understand the long-term effects of school closures on students' performance outcomes in ITS. In other words, it is unclear whether the performance increases observed within ITS during school closures persisted over the following months after the first and period of school closures when schools opened again.

### *The present study*

The present study set off to investigate longitudinal performance trajectories of students who studied mathematics using an ITS, with a particular focus on differential effects for lower-performing and higher-performing students. We considered data (from $n \approx 2700$ students; who worked on $n \approx 5$ million mathematical problems) collected between January 2017 and May 2021 from the ITS *Bettermarks* (*Bettermarks* GmbH, Berlin, Germany). Importantly, we only considered students who had been using the ITS before COVID-19 related school closures and who continued using it throughout 2021 after school closures (i.e., we considered a *within-student* study design). As such, we were able to evaluate longitudinal performance changes within students across a time span of about 4.5 years.

To investigate differences in performance trajectories between lower- and higher-performing students, we separated students according to their pre-pandemic performance in two groups (lower- vs. higher-performing students) and evaluated the performance trajectories of these students (i) during the period of first school closures (March-May 2020) as compared to their performance in the previous three years during these months, (ii) between periods of school closures (June-December 2020) as compared to their performance in the previous three years during these months, (iii) during the second period of school closures (January-February 2021) compared to their performance in the previous four years during the same months, and in a follow-up analysis (March-May 2021) compared to their performance in the previous four years during the same months.

This allowed us to evaluate the effects of COVID-19 related school closures during the first and second period of school closures, but also on the time period between these two school closures (June-December 2020) and following school closures (March-May 2021) within the same students. For each time period comparison, we also evaluated differential effects for lower-and higher- performing students.

Based on previous studies considering data from ITS [50], we expected performance increases during the first and second period of school closures as compared to the previous three years for both low-and higher-performing students but that lower-performing students in particular would increase their performance. We had no specific hypotheses on the persistence of these positive effects after schools opened again after the first and second periods of school closures.

## Methods

### *The intelligent tutoring system*

#### *Overview*

The study considered data obtained from the ITS *Bettermarks* for learning mathematics which is used in schools across Germany since 2008 (e.g., [46,51]). *Bettermarks* is used in different types of schools, including and vocational track schools, in all regions of Germany, making the students who use the software representative of the population of students in Germany. It covers the mathematics curriculum for grades 4 to 12, spanning an age range from 9 to 18 years, and includes over 100 book topics that encompass a range of mathematical themes, such as basic and advanced arithmetic, decimals, fractions, algebra, geometry, probability, and statistics. Each book topic provides an overview of the mathematical concept and contains several problem sets, with an average of nine individual problems per set. Students work on these problem sets to learn mathematics.

#### *Usage and adaptive features*

The ITS can be used in the classroom or at home, with teachers assigning problem sets to students to complete in class or at home. Additionally, students can also self-select problem sets for independent practice and receive feedback on their accuracy. Teachers can assign single problem sets or whole books which reflect a topic (e.g., "Addition and subtraction of fractions") and include all problem sets of the book. As such, teachers can assign small chunks of mathematical problems or larger bundles of problems.

One specific feature of the ITS is that the system is capable of identifying knowledge gaps when students work on specific problems and provides students with additional learning material (other problem sets) suited to fill the knowledge gaps. Another feature of the ITS is that students may ask for hints when working on a problem and the ITS provides hints for students. Additionally, students are able to repeat problem sets, however, the parameterization of problem sets changes with each new attempt. Thus, students who try to memorize the results of previous attempts to game the system by inserting the memorized result on a new attempt get feedback that their memorized answer on the new attempt is incorrect. Finally, the ITS provides feedback to students and if the problem sets students worked on were assigned by their teachers, feedback on the accuracy of problem sets is also provided to teachers. Moreover, students receive feedback on the problems they work on. For some problems, the feedback is specific to the errors students made. For example, when students add the following two fractions "1/2 + 2/5", and they falsely add numerator and denominator separately and thus inserts "3/7" as the answer, the ITS will provide the following feedback to the student: "Don't add the numerators and the denominators. Find the lowest common denominator.".

The above-described adaptive features of the ITS *Bettermarks* are commonly implemented in other ITSs as well (e.g., [1,22,23,27,35]). For instance, the ITS *Mathia* (© 2023 Carnegie Learning, Inc.), also provides students with adaptive feedback when solutions are incorrect. Additionally, students may also request hints, and content specific feedback on well-known misconceptions for learning mathematics. However, *Bettermarks* differs from *Mathia* as it does not incorporate so called knowledge components implemented in some of their problems. Instead, knowledge gaps are triggered within *Bettermarks* (see above). Nevertheless, the implementation of both—knowledge components and knowledge gaps—serve to aid students' learning progress enabling them to master their learning goals.

Finally, the collected data includes students' accuracy rates on problem sets, problem IDs, and the date and time of working on them. The data is fully anonymized, and therefore demographic information is not available. Moreover, the analysis was a secondary data analysis from data which is continuously stored by *Bettermarks* with data available upon request from the first author. As this was a secondary retrospective data analysis no ethics approval was needed. However, all users agreed that their data will be stored and analyzed. However, it is not possible to trace back any of our data to individuals.

### *Inclusion criteria*

We considered the following inclusion criteria. First, we included data on problem sets that were worked on by students between January 1st, 2017 and May 31st, 2021. Second, we only considered students who worked through 5 or more problem sets during *each* of the following periods before, during, and after COVID-19 related school closures. These periods were: (i) first school closures: March – May,[2] 2020 compared to 2017–2019; (ii) between school closures: June-December 2020 compared to 2017–2019; (iii) second school closures: January – February 2021 compared to 2017–2020; (iv) following school closures: March – May 2021 compared to 2017–2019. As we only considered

[2] The specific time period was March 16th -May 31st that we used here. Only data within this period was considered for data analysis.

students who already used the ITS before COVID-19 related school closures in each of the periods and continued to use the ITS during as well as after COVID-19 related school closures, we were able to evaluate students' longitudinal performance trajectories. Third, we only considered problem sets students got assigned by their teachers and if they got single problem sets assigned as this usage type is mostly used within *Bettermarks*. With these inclusion criteria applied, the dataset included 2708 students who computed 636,060 problem sets which sum up to more than 5 million mathematical problems.

*Learning measures*

The learning of students can be measured by their average accuracy rates which we refer to as absolute accuracy rates.[3] However, absolute accuracy rates depend on the difficulty of problems sets students worked on which needs to be considered. We accounted for problem set difficulty by computing the average accuracy rate for each problem set for all students who worked with *Bettermarks* during a reference period from January 2017 until the end of February 2020 (for a similar approach consider [50]). Low average accuracy rates during this period indicated that problem sets were comparably difficult while high average accuracy rates indicated that problem sets were relatively easy. These two measures—absolute accuracy rate and average accuracy rate of problem sets—were then related to each other by subtracting the average accuracy of a problem sets by students' accuracy from the same problem providing the relative accuracy rate of problem sets.

As an example, consider a student who got an accuracy rate of 90 % on a given problem set and the overall average accuracy rate of this problem set was 80 % (i.e., the average accuracy rate across all students who worked on this problem set). 90 % - 80 % then leads to a relative accuracy rate of 10 %, reflecting that the student performed 10 percentage points better than the average of students on this particular problem set during the reference period. As such, positive relative accuracy rates reflect that a student performed better than average whereas negative relative accuracy rates indicate that a student performed worse than average on a given problem set. This allowed us to interpret relative accuracy rates similarly to absolute accuracy rates with high values in both measures indicating better performance.[4] Finally, we also report the average accuracy for the entire time period (January 1st, 2017 and May 31st, 2021) and illustrated the average accuracy in Fig. 1.

## Data analysis

The R software was used for data analysis (R [41]). Students with a relative accuracy rate below 0 during the period before school closures (January 2017 to February 2020) were assigned to the lower-performing group, while students with a relative accuracy rate of 0 and higher during the period before school closures (January 2017 to February 2020) were assigned to the higher-performing group.

We simplified the quantitative analysis by running separate linear regression models for each time period of interest as well as separate regression models for low- and higher-performing students. In particular, we quantified differences in relative accuracy rates for the following time periods: (i) first school closures (March – May 2020) compared to the same months in 2017–2019; (ii) between school closures (June – December 2020) compared to the same months in 2017–2019; (iii); second school closures (January – February 2021) compared to the same months in 2017–2020; (iv) following school closures (March – May 2021) compared to the same months in 2017–2019.

For each of these periods, we conducted a hierarchical linear regression analysis with students' *relative accuracy* as the dependent variable and a binary *time* variable (before school closures vs. during/after school closures) and a binary *group* variable (lower-performing vs higher-performing) as the independent variables. Importantly, we included both independent variables as main effects but also the interaction in the hierarchical linear regression model. Both binary variables (time and group) were contrast coded with $-1$ and $+1$ for a full factorial design matrix. We also fitted a random intercept for students to account for variability between students.

## Results

Absolute accuracy rate, relative accuracy rate, and average difficulty separated for the two learning groups and for each month for years 2020 and 2021, and the average for 2017 to 2019 (2017 to 2020 for January and February) as the period before school closures are depicted in Fig. 1. We report the estimated means from the regression models of relative accuracy rates for each time period below, while we also list the unstandardized ß estimates and *t*-values of the hierarchical linear regression model comparing students' relative accuracy rates in the time periods of interest in Table 1.

*First school closures: March 16th – may 31st, 2020*

Considering performance differences during first school closures in 2020 as compared to 2017–2019, our regression revealed a significant main effect for time (*beta* $=0.02$; $t = 25.59$; $p < .001$), suggesting an overall increase in relative accuracies during the first period of school closures as compared to before. The significant main effect for group (*beta* $=0.08$; $t = 46.33$; $p < .001$) suggested that higher-performing students performed significantly better than lower-performing students. Interestingly, the interaction between time and group was significant (*beta* $< -0.01$; $t = -4.17$; $p < .001$) indicating that the performance increase in lower-performing students was relatively stronger than the performance increase of higher-performing students.

*Between school closures: June 1st – December 31st, 2020*

We observed a similar pattern of results for the time between the two periods of school closures. Regression result revealed a main effect for time (*beta* $=0.01$; $t = 8.08$; $p < .001$), suggesting an overall increase in relative accuracies for the time between the two periods of school closures as compared to the same time window before the pandemic. As in the previous analysis, there also was a significant main effect of group (*beta* $=0.08$; $t = 49.85$; $p < .001$) suggesting that higher-performing students performed significantly better than lower-performing students. Finally, the interaction between time and group also was significant (*beta* $<-0.01$; $t = -2.29$; $p < .001$) indicating that the performance of lower-performing students increased relatively more strongly than the performance of higher-performing students.

*Second school closures: January 1st – February 28th, 2021*

Our third regression model pointed to results similar to the previous two regression models, with significant main effects of time (*beta* $=0.01$; $t = 15.36$; $p < .001$) and group (*beta* $=0.08$; $t = 43.68$; $p < .001$). These suggested performance improvements during the second period of school closures as compared to the same two months in the previous four years and that higher-performing students significantly outperformed lower-performing students. Again, we also observed the interaction between time and group to be significant (*beta* $<-0.01$; $t = -2.29$; $p < .001$) suggesting that the increase in performance was more pronounced for lower-performing as compared to higher-performing students.

[3] We considered the best result of students for each problem set in case students repeated problem sets.

[4] We illustrate students' relative accuracy, students' absolute accuracy, and the average accuracy of problem sets. However, we considered to reduce the complexity of the already comprehensive analysis by only quantifying differences between students' relative accuracy rates.

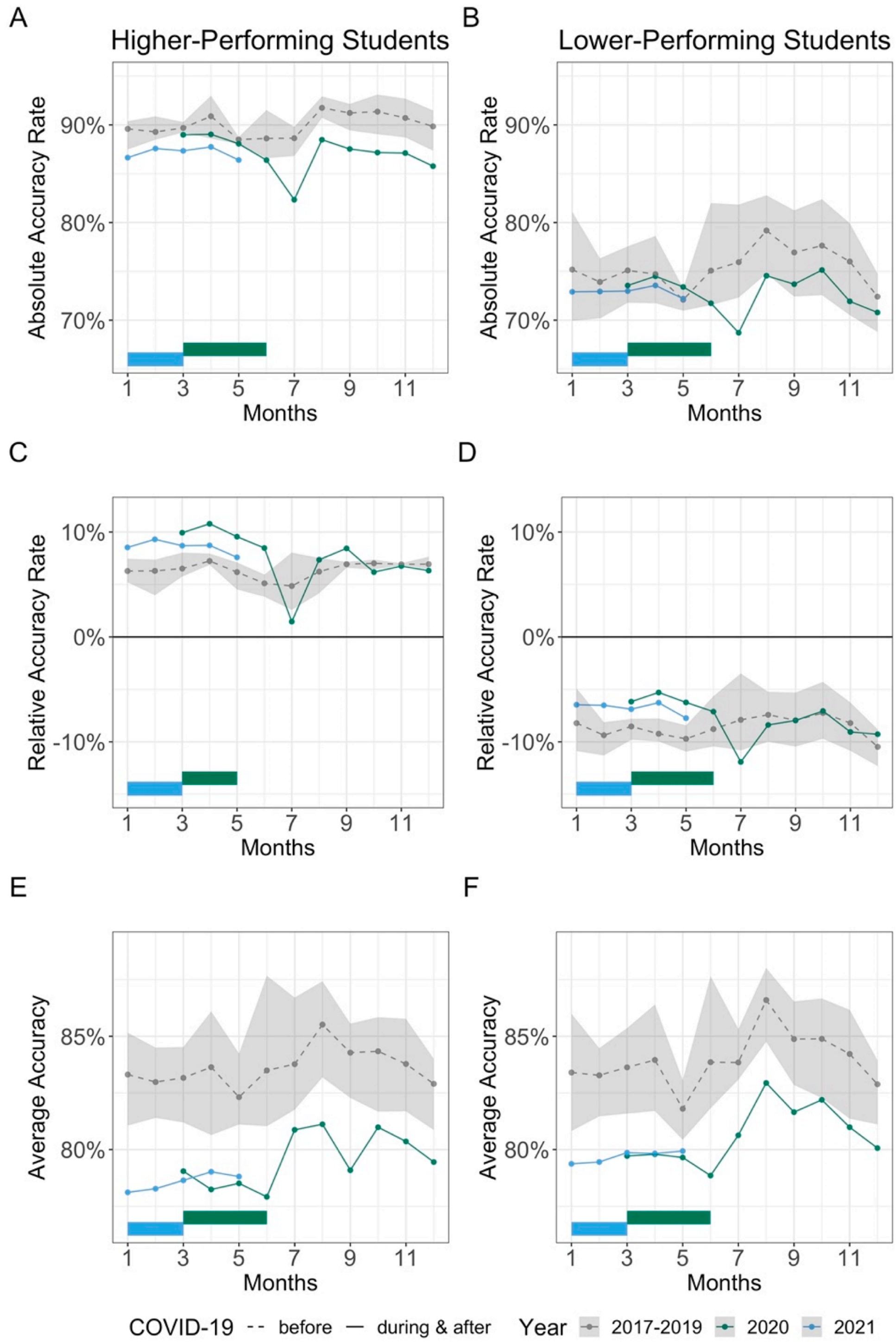


**Fig 1. Results for higher- (charts A, C, & E) and lower-performing students (charts B, D, & F).** The average results from the years before COVID-19 related school closures are depicted by grey points connected with grey dashed lines. Grey shaded areas indicate the minimum and maximum average result for these years. Periods of school closures are marked in the respective color bars: green bar = period of first school closure (March16th – May 31st, 2020); blue bar = period of second school closure (January 1st – February 28th, 2021). Results from the period of first school closures until the end of the year 2020 are reflected by green points and connecting lines. Results from the second period of school closures until the end of May 2021 are depicted by blue points and connecting lines. Each point results from aggregating the data for each participant for each month and year and then again for each month and each year. **A:** Absolute accuracy rates of higher-performing students. **B:** Absolute accuracy rates of lower-performing students. **C:** Relative accuracy rate of higher- performing students. **D:** Relative accuracy rate of lower- performing students. **E:** Average accuracy of problem sets of higher-performing students. **F:** Average accuracy of problem sets of lower- performing students. Note that we included all data from before first school closures in the Year 2017–2019 group and thus data from January and February 2020 is included in the grey dots and shades.

**Table 1**
Regression Results on Relative Accuracy Rate (estimates, t-values; and p-values indicated with asterix) of each Hierarchical Linear Regression Model.

| | 1. School Closures | | | Between School Closures | | | 2. School Closures | | | After School Closures | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Coeffcient* | *b* | *SE* | *t-Value* | *b* | *SE* | *t-Value* | *b* | *SE* | *t-Value* | *b* | *SE* | *t-Value* |
| (Intercept) | <−0.01 | <0.01 | −0.58 | −0.01 *** | <0.01 | −8.30 | −0.01 * | <0.01 | −2.20 | −0.01 *** | <0.01 | −4.74 |
| Time | 0.02 *** | <0.01 | 25.59 | <0.01 *** | <0.01 | 8.08 | 0.01 *** | <0.01 | 15.36 | 0.01 *** | <0.01 | 12.19 |
| Group | 0.08 *** | <0.01 | 46.33 | 0.08 *** | <0.01 | 49.85 | 0.08 *** | <0.01 | 43.68 | 0.08 *** | <0.01 | 44.11 |
| Time × Group | <−0.01 *** | <0.01 | −4.17 | <−0.01 * | <0.01 | −2.29 | <−0.01 ** | <0.01 | −3.21 | <−0.01 *** | <0.01 | −4.41 |
| $N_{Students}$ | 2708 | | | 2708 | | | 2708 | | | 2708 | | |
| $N_{Computed\ Problem\ Sets}$ | 164,986 | | | 291,725 | | | 133,639 | | | 125,193 | | |

* $p$<.05.
** $p$<.01.
*** $p$<.001.

*Following school closures: March 1st – May 31st, 2021*

Our fourth and final regression model again replicated the results of the previous regression models indicating a significant main effect of time (*beta* =*0*.01; $t = 12.19$; $p < .001$) and group (*beta* =*0*.08; $t = 44.11$; $p < .001$). These suggested performance improvements between March and May 2021 compared to the March to May period in years 2017 to 2019 as well as that higher-performing students performed significantly better than lower-performing students. We also observed a significant interaction of time and group (*beta* $< -0.01$; $t = -4.41$; $p < .001$) suggesting that the performance of lower-performing students increased relatively more than that of higher-performing students.

## Discussion

In this study, we investigated longitudinal performance trajectories of students who learned mathematics within an ITS between January 2017 and May 2021, using a within-student study design. We were particularly interested in how the two periods of COVID-19 related school closures (March-May 2020 and January-February 2021 in Germany) affected the performance of lower- and higher-performing students. In addition, we were interested in students' performance changes i) in the time between these two periods of school closures as well as ii) in performance changes after the periods of school closures as compared to the same times in the years before the pandemic.

We observed that the performance of both lower- and higher-performing students increased during the first period of school closures (see Fig. 1). Interestingly, however, the observed increase was more pronounced for lower-performing students. These results are in line with previous studies also considering data from ITS indicating performance increases during the first period of school closures [35, 48–50].

Nevertheless, our results also go beyond those of earlier studies as we followed the performance trajectories of the *same* students (both higher- and lower-performing students) after the first period of school closures to evaluate whether the effects observed for this period of school closures lasted. For the time between school closures (i.e., June-December 2020) our results suggested that students' performance remained on an increased level compared to previous years during these months. Again, the respective performance increases in comparison to the years before the pandemic were more pronounced for lower- as compared to higher-performing students. To the best of our knowledge, this study is the first to provide longitudinal evidence on students' performance trajectories suggesting that students' performance in mathematics not only increased during this period of school closures but also that this effect seemed to persist over months when schools opened again. Moreover, our results suggest that the effect of more pronounced performance increases for lower-performing students also remained stable during the time when schools opened again after the first period of school closures in 2020.

Then schools were closed for a second time in 2021. Again, our analysis suggested that the performance of both groups significantly increased during the second period of school closures (January and February 2021) as compared to the same period in previous years (2017–2020). However, similar to our results for the first period of school closures and the time between these two periods of school closures, the performance of lower-performing students increased more strongly compared to that of higher-performing students. And similar to the year 2020, when schools were opened again in 2021, this pattern of results remained the same with overall significant performance increases for both groups of students compared to the same time periods in previous years before school closures but more pronounced performance increases for lower-performing students compared to higher-performing students.

Together, these results basically replicated the pattern found for the succession of the first period of school closures and school re-openings in 2020 reflecting that the performance increase observed during the second period of school closures persisted into the time when schools were opened again in 2021 and was again more pronounced for lower-performing students. As such, our results suggest significant performance increases during, between, and after both periods of COVID-19 related school closures for lower- and higher-performing students—but also indicate that the performance of lower-performing students increased more strongly during, between, and after school closures. Our results thus provide evidence that the gap between lower- and higher-performing students decreased during, between, and after periods of school closures, which seems at odds with results reported based on data collected in the classroom and not within an ITS (e.g., [43]).

Understanding the effect of the COVID-19 pandemic on students' academic achievements in general is crucial for several reasons. First, following students' longitudinal performance trajectories is important as academic achievements are closely tied to future career, but also life prospects more generally [34,37,38]. Thus, assessing the pandemics' impact on students' learning process may help educators, policymakers, and parents to make informed decisions to minimize long-term consequences and provide the necessary support for students to catch up (e.g., via additional tutoring; see [14]). However, it is important to note that in contrast to other studies (e.g., [43]), we observed that students' performance improved—rather than decreased—during and after school closures. This, in turn, indicates that students who had access to remote learning infrastructures during times of crises were at least not negatively affected by the COVID-19 pandemic, as most studies from in-person assessments observed (for a meta-analysis see: [4]). Thus, our results suggest that students with access to ITSs may be better prepared for future situations that prohibit students to learn in schools and necessitate distance learning. This does not necessarily need to be situations as severe as another pandemic but may well include periods of absence from in-person teaching (e.g., when teachers are on sick leave), in which students would also benefit from digital learning opportunities. Thus, our results suggest policy makers and education administration to ensure that students are equipped with and prepared for the use of such digital learning software so that their education is not disrupted in the face of future challenges. In sum, it is essential to ensure that all students have the opportunity to reach their full academic potential, regardless of the challenges posed by external constraints. Our findings suggest that students who have the opportunity to (also) learn with digital learning software, such as ITSs, seem less affected and may even learn better within ITSs during situations of distance learning.

*Future works*

While this study aimed to investigate students' performance trajectories longitudinally, an interesting future avenue may be to identify specific mechanisms on why lower-performing students increased their performance more strongly when learning from a distance at home compared to higher-performing students. One potential explanation for the overall increased performance during school closures is that teachers may have incentivized students' use of and learning within the ITS differently during periods of school closures as compared to before and after. For instance, teachers may have told their students that their performance within the ITS counts towards their grades during school closures. However, this explanation does not account for the differential effect for lower- vs. higher-performing students during, between, and after school closures as well as the persisting effect of performance increases between and after school closures.

Another aspect which might explain our differential effects for lower- vs. higher-performing students is that the former experienced less math anxiety when studying mathematics at home, compared to being at school [2,3,6–8,33,52]. As math anxiety is known to impair cognitive processes involved in studying mathematics (e.g., working memory), it might be that these cognitive processes are less impaired when studying at home in absence of teachers—particularly for lower-performing students. Against this background, an interesting future avenue may be to have a closer look at whether math anxiety is indeed reduced in lower-performing students when learning mathematics from home.

Another interesting future research avenue may closer inspect whether differential effects can be observed for students' learning progress depending on whether they study in the classroom while their teachers are present who may provide students with additional help beyond the learning aid provided by the ITS, or at home, when teachers cannot help students' during their learning progress.

*Limitations*

It is important to take into account certain limitations of our study. First, our study only considered the effect of COVID-19 related school closures on students' performance within an ITS and these students worked with this ITS before, during and after school closures. We observed that students' performance improved during and after school closures as compared to before while other studies from in-person assessments reported performance decreases after school closures. Importantly, we did not compare our sample of students with other students who did not use an ITS before, during, and after COVID-19 related school closures. Contrasting the effects of students who study with an ITS against students who do not study with an ITS during the COVID-19 related school closures would have been ideal for understanding the specific effect of having access to ITSs during times of crisis. However, this data was not available which limits our study in this respect. Future work may consider assessing the general effect of ITSs on students' performance against another cohort of students without access to ITSs with a randomized control trial study.

Another limitation of our study is that we restricted our analysis to a relatively broad age range of students (9 to 18 years) as students within this range use *Bettermarks*. Thus, our hypotheses only considered average trends of students within this age range. An interesting future research avenue may address whether our observed results are similar for different age groups.

Finally, our result that lower-performing students improved more than higher-performing students may be due to a potential ceiling effect. That is, higher-performing students potentially already performed at their maximum level before school closures leaving little room for improvement. In contrast, lower-performing students had much more room to improve which could explain our findings.

## Conclusion

Altogether, our study provides evidence from a rich longitudinal dataset on longitudinal performance trajectories within the same students indicating that lower-performing students particularly benefited from engaging with an ITS for mathematics during, between, and after school closures. We hope that our study may serve as a foundation for relevant future research on how to aid learning mathematics in general and during times of crisis—especially in those who fall behind.

## Declaration of competing interest

There are no known conflicts of interest associated with this publication.

## References

[1] Almoubayyed H, Fancsali SE, Ritter S. Instruction-embedded assessment for reading ability in adaptive mathematics software. In: Proceedings of the ACM international conference proceeding series; 2023. p. 366–77. https://doi.org/10.1145/3576050.3576105.

[2] Ashcraft MH, Kirk EP. The relationships among working memory, math anxiety, and performance. J Exp Psychol Gen 2001;130(2):224–37. https://doi.org/10.1037/0096-3445.130.2.224.

[3] Ashcraft MH, Krause J. Working memory, math performance, and math anxiety. Psychon Bull Rev 2007;14(2):243–8.

[4] Betthäuser BA, Bach-Mortensen AM, Engzell P. A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. Nat Hum Behav 2023;2023:1–11. https://doi.org/10.1038/s41562-022-01506-4.

[5] Birkelund JF, Karlson KB. No evidence of a major learning slide 14 months into the COVID-19 pandemic in Denmark. Eur Soc 2022:1–21. https://doi.org/10.1080/14616696.2022.2129085.

[6] Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020;287:1–5. https://doi.org/10.1016/j.psychres.2020.112934.

[7] Carey E, Devine A, Hill F, Szucs D. Differentiating anxiety forms and their role in academic performance from primary to secondary school. PLoS One 2017;12(3):1–20. https://doi.org/10.1371/journal.pone.0174418.

[8] Caviola S, Toffalini E, Giofrè D, Ruiz JM, Szűcs D, Mammarella IC. Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: a meta-analysis on 906,311 participants. Educ Psychol Rev 2022;34(1). https://doi.org/10.1007/s10648-021-09618-5. Issue.

[9] Corbett A, Anderson JR. Knowledge tracing : modeling the acquisition of procedural knowledge. User Model User-Adapt Interact 1995;4:253–78.

[10] Donnelly R, Patrinos HA. Learning loss during COVID-19: an early systematic review. Prospects 2022;51(4):601–9. https://doi.org/10.1007/s11125-021-09582-6.

[11] Engzell P, Frey A, Verhagen MD. Learning loss due to school closures during the COVID-19 pandemic. Proc Natl Acad Sci U S A 2021;118(17). https://doi.org/10.1073/PNAS.2022376118.

[12] Förster N, Forthmann B, Back MD, Souvignier E. Effects of the COVID-19 pandemic on reading performance of second grade children in Germany. Read Writ 2023;36(2):289–315. https://doi.org/10.1007/s11145-022-10379-y.

[13] Fütterer T, Hoch E, Lachner A, Scheiter K, Stürmer K. High-quality digital distance teaching during COVID-19 school closures: does familiarity with technology matter? Comput Educ 2023;199. https://doi.org/10.1016/j.compedu.2023.104788. MarchArticle 104788.

[14] Gambi, L., De Witte, K., De Meyst, M., Derks, A., Groenez, S., Maldonado, J., Chénier, M.A., Palmaccio, S., & Decoster, A. (2021). The resiliency of school outcomes after the COVID-19 pandemic.Standardised test scores and inequality one yearafter long term school closures The resiliency of school outcomes after the COVID-19 pandemic. Standardised test scores and inequality one year. October.

[15] Gore J, Fray L, Miller A, Harris J, Taggart W. The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study. Australian Educational Researcher (Issue 0123456789) 2021. https://doi.org/10.1007/s13384-021-00436-w.

[16] Grewenig, E., Werner, K., & Zierow, L. (2020). COVID-19 and educational inequality : how school closures affect low- and high-achieving students. https://www.cesifo.org/en/publikationen/2020/working-paper/covid-19-and-educational-inequality-how-school-closures-affect-low.

[17] Haelermans C, Korthals R, Jacobs M, de Leeuw S, Vermeulen S, van Vugt L, Aarts B, Prokic-Breuer T, van der Velden R, van Wetten S, de Wolf I. Sharp increase in inequality in education in times of the COVID-19-pandemic. PLoS One 2022;17. https://doi.org/10.1371/journal.pone.0261114. VolIssue 2 February.

[18] Hammerstein S, König C, Dreisörner T, Frey A. Effects of COVID-19-related school closures on student achievement-a systematic review. Front Psychol 2021;12:1–8. https://doi.org/10.3389/fpsyg.2021.746289. September.

[19] Helm C, Huber SG. Predictors of central student learning outcomes in times of COVID-19: students', parents', and teachers' perspectives during school closure in 2020—a multiple informant relative weight analysis. Front Educ 2022;7:1–22. https://doi.org/10.3389/feduc.2022.743770 (Lausanne)February.

[20] Helm C, Huber S, Loisinger T. Meta-Review on findings about teaching and learning in distance education during the Corona pandemic—evidence from Germany, Austria and Switzerland. Z Erziehungswiss 2021;24(2):237–311. https://doi.org/10.1007/s11618-021-01000-z.

[21] Hevia FJ, Vergara-Lope S, Velásquez-Durán A, Calderón D. Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. Int J Educ Dev 2022;88. https://doi.org/10.1016/j.ijedudev.2021.102515.

[22] Hilz A, Guill K, Roloff J, Aldrup K, Köller O. The relationship between individual characteristics and practice behaviour within an adaptive arithmetic learning program. J Comput Assist Learn 2023;39(3):970–83. https://doi.org/10.1111/jcal.12780.

[23] Hilz A, Guill K, Roloff J, Sommerhoff D, Aldrup K. How to continue? New approaches to investigating the effects of adaptive math learning programs on students' performance, self-concept, and anxiety. J Intell 2023;11(6):108. https://doi.org/10.3390/jintelligence11060108.

[24] Huber SG, Helm C. COVID-19 and schooling: evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. Educ Assess Eval Account 2020;32:237–70.

[25] Jaekel AK, Scheiter K, Göllner R. Distance teaching during the COVID-19 crisis: social connectedness matters most for teaching quality and students' learning. AERA Open 2021;7(1):1–14. https://doi.org/10.1177/23328584211052050.

[26] Koedinger KR, Anderson JR, Hadley WH, Mark MA. Intelligent tutoring goes to school in the big city. Int J Artif Intell Educ 1997;1(1):30–43.

[27] Koedinger KR, Carvalho PF, Liu R, Mclaughlin EA. An astonishing regularity in student learning rate. Proc Natl Acad Sci 2023;120(13):e2221311120. https://doi.org/10.1073/pnas.

[28] Kuhfeld M, Soland J, Lewis K. Test score patterns across three COVID-19-impacted school years. Educ Res 2022;51(7):500–6. https://doi.org/10.3102/0013189×221109178.

[29] Kuhfeld M, Soland J, Lewis K, Ruzek E, Johnson A. The COVID-19 school year: learning and recovery across 2020-2021. AERA Open 2022;8(1):1–15. https://doi.org/10.1177/23328584221099306.

[30] Lerkkanen MK, Pakarinen E, Salminen J, Torppa M. Reading and math skills development among Finnish primary school children before and after COVID-19 school closure. Read Writ 2022;36(2):263–88. https://doi.org/10.1007/s11145-022-10358-3.

[31] Lichand G, Doria CA, Leal-Neto O, Fernandes JPC. The impacts of remote learning in secondary education during the pandemic in Brazil. Nat Hum Behav 2022;6(8):1079–86. https://doi.org/10.1038/s41562-022-01350-6.

[32] Maldonado JE, De Witte K, Maldonado J. The effect of school closures on standardised student test outcomes. KU Leuven – Faculty of Economics and Business; 2020. p. 20–48. September, https://lirias.kuleuven.be/3189074.

[33] Maloney EA, Beilock SL. Math anxiety: who has it, why it develops, and how to guard against it. Trends Cogn Sci Regul Ed 2012;16(8):404–6. https://doi.org/10.1016/j.tics.2012.06.008.

[34] Mbekeani PP, Papay J, Mantil A, Murnane RJ. Understanding high schools' effects on longer-term outcomes 2023;23:23–729. https://doi.org/10.26300/qwp6-hk05.

[35] Meeter M. Primary school mathematics during Covid-19: no evidence of learning gaps in adaptive practicing results. Trends Neurosci Educ 2021;25:100163. https://doi.org/10.1016/j.tine.2021.100163. October.

[36] Molnár G, Hermann Z. Short- and long-term effects of COVID-related kindergarten and school closures on first- to eighth-grade students' school readiness skills and mathematics, reading and science learning. Learn Instr 2023;83. https://doi.org/10.1016/j.learninstruc.2022.101706. May 2022.

[37] Murnane RJ, Willett JB, Levy F. The growing importance of cognitive skills in wage determination. Rev Econ Stat 1995;77(2):251–66.

[38] National Mathematics Advisory Panel. Foundations for success: the final report of the national mathematics advisory panel. U.S. Department of Education; 2008.

[39] Panagouli E, Stavridou A, Savvidi C, Kourti A, Psaltopoulou T, Sergentanis TN, Tsitsika A. School performance among children and adolescents during covid-19 pandemic: a systematic review. Children 2021;8(12):1–12. https://doi.org/10.3390/children8121134.

[40] Pelikan ER, Lüftenegger M, Holzer J, Korlat S, Spiel C, Schober B. Learning during COVID-19: the role of self-regulated learning, motivation, and procrastination for perceived competence. Z Erziehungswiss 2021;24(2):393–418. https://doi.org/10.1007/s11618-021-01002-x.

[41] R Core Team. (2013). R: a language and environment for statistical computing. http://www.r-project.org/.

[42] Ritter S, Anderson JR, Koedinger KR, Corbett A. Cognitive tutor : applied research in mathematics education. Psychon Bull Rev 2007;14(2):249–55.

[43] Schult J, Mahler N, Fauth B, Lindner MA. Long-term consequences of repeated school closures during the COVID-19 pandemic for reading and mathematics competencies. Front Educ 2022;13:1–13. https://doi.org/10.3389/feduc.2022.867316 (Lausanne)April.

[44] Schult J, Nicole M, Fauth B, Lindner MA. Did students learn less during the COVID-19 pandemic? Reading and Mathematics Competencies Before and After the First Pandemic Wave; 2021. https://psyarxiv.com/pqtgf/.

[45] Schuurman TM, Henrichs LF, Schuurman NK, Polderdijk S, Hornstra L. Learning loss in vulnerable student populations after the first COVID-19 school closure in the Netherlands. Scandinavian Journal of Educational Research; 2021. https://doi.org/10.1080/00313831.2021.2006307.

[46] Spitzer MWH. 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 2022;37(1):39–53. https://doi.org/10.1007/s10212-021-00546-0.

[47] Spitzer MWH, Gutsfeld R, Wirzberger M, Moeller K. Evaluating students ' engagement with an online learning environment during and after COVID-19 related school closures : a survival analysis approach. Trends Neurosci Educ 2021;25:100168.

[48] Spitzer MWH, Moeller K. Performance increases in mathematics during COVID-19 pandemic distance learning in Austria: evidence from an intelligent tutoring system for mathematics. Trends Neurosci Educ 2023;31:100203.

[49] Spitzer MWH, Moeller K, Musslick S. Assignment strategies modulate students' academic performance in an online learning environment during the first and second COVID-19 related school closures. PLoS ONE; 2023.

[50] Spitzer MWH, Musslick S. 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. PLoS One 2021;16(8):1–16. https://doi.org/10.1371/journal.pone.0255629.

[51] Stapel M, Zheng Z, Pinkwart N. An ensemble method to predict student performance in an online math learning environment. In: Proceedings of the 9th international conference on educational data mining; 2016. p. 231–8.

[52] Stella M, de Nigris S, Aloric A, Siew CSQ. Forma mentis networks quantify crucial differences in STEM perception between students and experts. PLoS One 2019;14(10):1–21. https://doi.org/10.1371/journal.pone.0222870.

[53] Tomasik MJ, Helbling LA, Moser U. Educational gains of in-person vs. distance learning in primary and secondary schools: a natural experiment during the COVID-19 pandemic school closures in Switzerland. Int J Psychol 2020;56(4):566–76. https://doi.org/10.1002/ijop.12728.

[54] UNESCO. Education: from disruption to recovery. School Clusures Caused by Coronavirus (COVID-19); 2020.

[55] van der Velde M, Sense F, Spijkers R, Meeter M, van Rijn H. Lockdown learning: changes in online foreign-language study activity and performance of dutch secondary school students during the COVID-19 pandemic. Front Educ 2021;6:1–8. https://doi.org/10.3389/feduc.2021.712987 (Lausanne)August.

[56] Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2020). Coronavirus crisis cut the time children spend on schooling in half. https://www.ifo.de/en/node/57298.

[57] Zierer K. Effects of pandemic-related school closures on pupils' performance and learning in selected countries: a rapid review. Educ Sci 2021;11(6). https://doi.org/10.3390/educsci11060252 (Basel).