frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 08 September 2023
DOI 10.3389/feduc.2023.1213857

Check for updates

OPEN ACCESS

EDITED BY
Alberto Crescentini,
University of Applied Sciences and Arts of Southern Switzerland, Switzerland

REVIEWED BY
Angel Mukuka,
Walter Sisulu University, South Africa
Ann Dowker,
University of Oxford, United Kingdom

*CORRESPONDENCE
Hedwig Gasteiger
✉ hedwig.gasteiger@uni-osnabrueck.de

RECEIVED 28 April 2023
ACCEPTED 28 August 2023
PUBLISHED 08 September 2023

CITATION
Gasteiger H, Sachse KA, Schumann K, Gerve M, Schulz A and Engelbert-Kocher M (2023) COVID-19-related school closures and mathematical performance—findings from a study with grade 3 students in Germany. *Front. Educ.* 8:1213857.
doi: 10.3389/feduc.2023.1213857

COVID-19-related school closures and mathematical performance—findings from a study with grade 3 students in Germany

Hedwig Gasteiger[1]*, Karoline A. Sachse[2,3], Kristoph Schumann[2], Mona Gerve[1], Axel Schulz[4] and Maria Engelbert-Kocher[2]

[1]Mathematics Education, Institute of Mathematics, Mathematics/Computer Science/Physics, Osnabrück University, Osnabrück, Germany, [2]Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin, Berlin, Germany, [3]Department of Psychology, University of Potsdam, Potsdam, Germany, [4]Institute for Didactics of Mathematics, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany

Introduction: Due to the COVID-19 pandemic, measures were taken that had a considerable impact on the situation in schools. In Germany, these measures lasted more than a year and ranged from school closures and distance learning to alternating teaching phases with small groups. In the present study, we examined whether third-grade students' mathematics performance changed in different content domains before and after the COVID-19-related changes in school.

Methods: In a repeated cross-sectional design, we compared two cohorts of third graders (2019: $N = 1{,}905$; 2021: $N = 3{,}203$) based on standardized mathematics tests, constructed according to the German National Educational Standards, which allowed for a differential competence diagnostic for five content domains. Generalized linear mixed models were used to model item and person properties and assess their effects on performance.

Results: There was a significant drop in performance overall. While the drop in the content domain Numbers & Operations was smaller than the overall drop in performance, the content domains Space & Shape and Data, Frequency, Probability were more affected.

Discussion: The findings of this study may be explained by the results of numerous surveys of students, teachers, and parents, which indicated that a lot of time was spent on exercises and reproduction tasks during distance learning.

KEYWORDS

COVID-19, comparative study, mathematics, primary school, distance learning, content domains

1. Introduction

Over the course of the COVID-19 pandemic starting at the beginning of 2020, various measures were taken to keep the spread of the virus low and minimize the associated consequences. One of the measures taken in almost all countries in the world was the closure of schools in Spring 2020 (for an overview: UNESCO Institute of Statistics, 2017). In Europe, the situation varied from country to country. While schools in Sweden, for example, remained open, 23 European countries closed their schools for between 30 and 109 days (including Germany, which closed schools for 52 days; OECD/European Union, 2020). In Germany, the period of school closure was followed by phases in which classroom instruction alternated with phases of alternating instruction (half of the class in the classroom, the other half in distance learning) for several months and regionally until summer 2021 depending on the regional incidence value (Schneider et al., 2022).

Some studies and surveys show that teaching and students' learning changed dramatically during this time: the school closure and phases of alternating instruction challenged teachers to organize distance learning and students and parents to learn at home. An online survey (~1,700 teachers in Germany, Austria, and Switzerland—all types of school) showed that 29% of the teachers contacted their students daily, while 27% did this only once a week or less (Fobizz, 2020), though regular contact between teachers and students (secondary school) or teachers and parents (primary school) was the most important predictor of teacher-reported achievement of learning goals (Schneider et al., 2021). Asking teachers, parents or students[1] about digital lessons, the results are similar: only a third of teachers or less taught students digitally (Helm et al., 2021). About a third of parents of 4–10-year-olds surveyed in a New Zealand study had a negative view of home-learning in mathematics, criticizing, for example, the lack of support from school (Darragh and Franke, 2021). Students' learning time per week decreased significantly during the time they had to stay at home. While some studies report that about 40% of the students stated spending 2 h or less on learning per day (Anger et al., 2020; Wacker et al., 2020), other studies report an average reported learning time of between 3.6 h (Wößmann et al., 2020) up to 5 h (Huber and Helm, 2020; Schober et al., 2020). In addition, students had very different conditions for learning at home. For example, the extent to which children had access to electronic devices was an important factor influencing learning at home and children's performance (Orbach et al., 2023).

It stands to reason that this situation also affected primary students' learning development and learning success in Germany. Though there is one study showing no decline in performance (Depping et al., 2021), larger studies based on standardized tests identified a drop in primary students' performance. Ludewig et al. (2022) found a substantial decline in reading performance comparing data from 2016 and 2021. Likewise, Stanat et al. (2022) found a substantial decline between 2016 and 2021 in all competencies under study, including reading and mathematics, based on German nationally representative samples of more than 25,000 fourth-graders per cohort. In the latter two studies, performance was examined at 5-year intervals. Although the authors emphasized that it is not possible to determine unequivocally the extent to which the observed changes in performance are due to pandemic-related restrictions, they considered it likely that these restrictions played a significant role.

In addition to the standardized comparison studies, there are some survey results concerning students' performance. Nearly two-thirds of parents surveyed in different European countries or in Germany exclusively expressed concern that their children learned less during the period of distance learning (Wößmann et al., 2020; Institut für Demoskopie Allensbach, 2021; Kantar Market Research Institute, 2021). More detailed analyses suggest differences with regard to gender. While parents of students in secondary school indicated for 44% of boys that they learned less during this time than before, this was true for only 30% of girls (Attig et al., 2020; Nusser et al., 2021). The effects of gender on achievement—especially for primary school—have been less studied so far.

Our study will contribute to existing research by providing a large database on children's mathematical competence in various mathematical content domains in relation to the German National Education Standards (time frame before and after the serious changes in the school situation). We address the question of how students' mathematical achievement in grade 3 changed between 2019 and 2021 across different mathematical content and skills and examine possible gender effects. In a repeated cross-sectional design, we compare two cohorts of third graders using standardized mathematics tests that are constructed according to the German Educational Standards and that allow differential competence analyses for five mathematical content domains and different levels of requirement.

2. Theoretical background and review of related literature

In order to make specific hypotheses about possible changes in mathematical performance due to school closures, we first briefly describe the goals of mathematics education in primary school and how they can be achieved. Findings on teaching and learning during school closures with a particular focus on mathematics are then reported.

2.1. Goals of mathematics education in primary school

National standards describe the mathematical competence that students should achieve at a given time in their school careers (KMK, 2004; Common Core State Standards Initiative, 2010).[2] Content and process-related standards specify the mathematical content students should know, and how they can learn and apply it. The German Educational Standards for primary school list expectations along the five guiding content domains of Numbers & Operations; Space & Shape; Measurement; Pattern & Structures; Data, Frequency, Probability and include Problem Solving, Mathematical Communicating, Modeling, Reasoning, and Representing as process-related proficiencies (KMK, 2004). Thus, mathematical competence encompasses not only knowledge, but also skills that are demonstrated in actively applying the content. In addition to content domains and process-related proficiencies, the German standards describe requirement levels (Anforderungsbereiche) that should be covered in teaching and in the selection of tasks. The spectrum of requirements ranges from rather simple reproduction of results to the more complex tasks of making connections to generalizing and reflecting.

1 The numerous surveys cited here and below rarely report results separated by type of school. If statements can be made in this regard, this is explicitly mentioned.

2 There is a renewed version of the KMK German National Education Standards from 2022. As this study is based on the KMK German National Education Standards from 2004, this version is cited here.

To reach the described goals, a high quality of mathematics instruction is necessary. We know that instructional support, among other factors, is related to student achievement in general (OECD, 2020). In particular, cognitive activation, clarity of instruction, and a supportive climate play important roles in this context (Blömeke et al., 2016). Schoenfeld (2018) differentiates these three aspects with his five dimensions for robust understanding in mathematics: the *Content* has to be presented in a way that allows for focused and coherent discussions and provides opportunities to make connections. Mathematics teaching should guarantee *Cognitive Demand*, which means to challenge students with tasks ranging from moderate to demanding (as described in the level of requirements) and provide situations of "productive struggle." Students should have *Equitable Access to Content*: all students need to engage with the content and be involved in meaningful ways. Mathematics education needs to provide opportunities to discuss disciplinary ideas and to build on each other's ideas (*Agency, Ownership,* and *Identity*), which will train especially the process-related proficiencies reasoning and communicating. The mathematics teacher should pay attention to *Formative Assessment* and give students the opportunity to deepen their understanding with appropriate guidance for further work, tailored to their level of learning.

To assess whether the mathematical proficiencies required by the German Educational Standards are achieved by the students, proficiency level models based on the standards were developed by the Institute for Educational Quality Improvement (IQB) together with subject experts. In comprehensive standard-setting processes experts divided the continuous scale in every content domain into multiple sections referred to as proficiency levels. For each level, there is a description of the cognitive requirements that students can meet once they reach that proficiency level. In this way, it is possible to assign qualitative descriptions of the proficiencies acquired by students and to determine what percentage of students are most likely to be able to meet specific requirements. To examine the effects of school closures on mathematics performance, the broad spectrum of mathematical competence with all the proficiencies that students should acquire needs to be considered.

2.2. Quality of distance learning

Whether a high quality of mathematics instruction as described above was realized in distance learning settings is presented below. Since the numerous surveys on distance learning often do not differentiate between school forms and surveys of primary school children are rather scarce, results of all school forms are reported below. If the school forms can be differentiated, this is indicated.

The instruction given to students (aged 10–16 years) during distance learning in mathematics consisted mainly of solving mathematical tasks (Heller and Zügel, 2020). 80% of parents reported that their children had to work on worksheets provided by their teachers, whereas only 29% reported that classroom teaching via video took place several times a week or daily (Wößmann et al., 2020). About 40% of the parents and also of the teachers stated that there was no possibility for a communicative exchange with classmates (Huber et al., 2020; Wößmann et al., 2020). These results were confirmed by a qualitative analysis of teaching descriptions of 700 teachers from four European countries (including Germany). The descriptions showed that more than two-thirds of the teachers tended to use a transmissive teaching style in distance learning. Some teachers stated that they changed their teaching style in distance learning compared to classroom teaching, because the particular conditions did not allow for collaborative teaching, or because transmissive learning was easier to organize (Aldon et al., 2021). A survey of 82 Australian elementary teachers found that the teachers' attitudes toward productive struggle or cognitive demand differed depending on whether the instruction was distance or face-to-face (Russo et al., 2021). The main reasons given by the teachers were the absence of a teacher-facilitated, synchronous learning environment, as well as parents' negative attitudes toward struggle when learning mathematics. These are clear indications that less cognitive activation took place in distance learning and that students had fewer opportunities to engage in discussions about disciplinary ideas and their ways of thinking (for an overview, see also Helm et al., 2021).

There is also evidence that instruction may have changed in terms of content and performance expectations. Secondary school teachers in structurally disadvantaged schools in particular argued for lowering the standard level during the period of distance learning (Bremm, 2021), and there are indications that the procedural aspect of mathematical concepts (including calculating, applying algorithms, or exercising skills and techniques) was more present in distance learning than conceptual understanding (Aldon et al., 2021).

In terms of a supportive climate or formative assessment, there is some indication that teachers gave comparatively less positive reinforcement to their students at home (Mantasiah et al., 2021). While parents reported that almost all students received worksheets to work on weekly, just under two-thirds of the students received feedback at least once a week (Wößmann et al., 2020). In particular, students of parents without an academic background and lower-performing students very rarely had individual contact with their teachers (Wößmann et al., 2020). There is some evidence that teachers supported students who struggled with given tasks primarily by generally providing *more* materials and additional explanations rather than offering individualized learning tasks (Aldon et al., 2021). On the other hand, only a few students and parents indicated that teachers were not available to answer students' questions (Helm et al., 2021). Regarding parental support for home-learning in mathematics, some parents stated that they found it difficult to provide the best possible support for their children (Darragh and Franke, 2021).

To summarize, there are indications that distance learning was less cognitively activating and more focused on procedural aspects such as calculating or exercising skills than on conceptual understanding. Moreover, it seems that students received less individual support and teachers considered lowering the standards.

2.3. Gender differences in coping with distance learning

As reported in the introduction section, home schooling significantly reduced the children's weekly learning time. Students report between 3.6 h (Wößmann et al., 2020) and 5 h of learning time per day (Schober et al., 2020; Helm et al., 2021). The reported learning time appears comparatively low if one assumes that a school day has at least 4.5 h of pure lesson time (six times 45 min) plus time for homework. A survey in seven European countries (including

Germany) with 6,285 parents and their children ($n = 5,767$, 10–18 years) found that girls spent 18 h of per week studying in addition to any online sessions, and boys spent 17 h (Kantar Market Research Institute, 2021). The fact that boys spent less time for learning than girls was also confirmed in other surveys (Wößmann et al., 2020; Lampert and Thiel, 2021; Nusser et al., 2021).

It seems that girls adapted to distance learning better than boys (Forsa, 2020). Boys more often did not know what to do for school compared to girls (Sturzbecher et al., 2021), and parents of boys were more likely than parents of girls to report that their child was more difficult to motivate to learn at home (Lockl et al., 2021). They saw a main challenge in their boys' ability to concentrate in online learning sessions, while girls' parents saw the lack of social connection as their child's main challenge during school closure (Kantar Market Research Institute, 2021).

2.4. Learning progress and performance development in mathematics

All the findings on distance learning ("Quality of distance learning") show that the conditions for teaching to achieve the desired objectives ("Goals of mathematics education in primary school") are poorly fulfilled. There are already some results on whether this also reflected in lower performance in mathematics. We report performance comparisons (time frame before and after the school closures) first from Germany, followed by international comparisons.[3]

Schult et al. (2022) used a mandatory, teacher-administrated test in the German federal state of Baden-Wuerttemberg in grade 5 to compare students' performance in the three pre-pandemic-years 2017–2019 with performance in September 2020 (Schult et al., 2022). After 2 months of school closure, students' performance was slightly below the average results of the 3 years prior to the pandemic. The difference was more pronounced in the subtest Operation ($d = -0.09$) than in the subtest Number ($d = -0.03$) and more in low-achieving than in high-achieving students.

The study by Depping et al. (2021) showed hardly any differences in performance before and after the school closures in the German federal state of Hamburg. Depping et al. compared the results of teacher-administered tests at the end of grade 3 before the pandemic (2019) in the content domains Space & Shape as well as Data, Frequency, Probability with teacher-administered tests at the beginning of grade 4 (2020). Additionally, they compared mathematical performance in all content domains at the beginning of grade 5 (2019 vs. 2020). Although both group comparisons showed slightly lower performance in 2020 than in 2019 (the loss in performance was most pronounced for the disadvantaged school group in grade 5), these differences proved to be statistically nonsignificant ($d < 0.2$).

Some international comparison studies show a drop in performance. Tomasik et al. (2021) used data generated from a computer-based formative assessment tool in Switzerland and compared students' learning gain in Mathematics and German 8 weeks before the school closure with the learning gain during 8 weeks of school closure separately for primary and secondary school students. While no differences were evident at the secondary level, learning gains at the primary level were twice as high in the 8 weeks prior to school closures than during the period of school closure. In addition, heterogeneity in achievement increased significantly (Tomasik et al., 2021).

In the Netherlands, schools were closed for 8 weeks. Engzell et al. (2021) report a loss of performance in mathematics, spelling and reading of about one fifth of a school year (3 percentiles or 0.08 standard deviations) during school closures. The authors analyzed a sample of students from 15% of all Dutch schools (aged 8–11). They compared scores from standardized tests immediately before and after the school closures with scores from the same periods in the 3 years prior to the pandemic. Learning loss was particularly pronounced in mathematics and among students from disadvantaged families, with no overall gender differences found (Engzell et al., 2021).

Similar results were reported by studies with standardized tests in California (2.5-month learning lag in mathematics in grade 4–8; Pier et al., 2021), in Belgium (school mean scores decreased by 0.19 standard deviation; Maldonado and de Witte, 2022) and in South Africa (schools were closed for 155 days, loss from 59 to 50%—interpreted as a loss of almost a full school year; van der Berg et al., 2022). The South African study tested five content domains in grade 3, but was not balanced across content domains: more than 60% of the items tested the Number, Operation, and Relationships domain. The other for domains were tested with four to six items (7–10% of all test items). The greatest loss of learning was reported in Data Handling (13%), followed by Numbers, Operations and Relationships (9%), Measurement (8%), Patterns, Functions and Algebra (7%), and Geometry (6%; van der Berg et al., 2022).

However, there are further studies that show no significant drop in performance. A study of more than 2,000 students (average age 9.7) in Australia did not find an impact of school closures on performance in general ($d = 0.06$) but significant negative effects on performance among students from disadvantaged schools ($d = -0.16$; Gore et al., 2021). A drop in performance was also not evident in the study by Meeter (2021), who analyzed successively collected performance data from a digital learning environment in the Netherlands. During the school closure, younger children in particular even showed higher performance than children in previous years. It seems that they were able to use the lockdown period to practice key skills. However, this performance advantage disappeared when schools reopened (Meeter, 2021). Since the situation after the initial school closures differed in many countries, the findings of other countries cannot be easily transferred to the German context. Some of the German studies refer to a long period of time (see Introduction, Stanat et al., 2022), to only one federal state (Depping et al., 2021), not to the primary level (Schult et al., 2022) or only to some content domains (Depping et al., 2021). For this reason, specific results from the primary level that do not only focus on mathematical competence in general, but also on the different content domains in an adequate time frame around the school closures, is thus far missing.

3. Summary and research questions

From the numerous surveys about the school situation during the pandemic, it is evident that there were massive changes in learning at

3 We primarily report results from studies at the primary level – otherwise, the school level is indicated.

school and at home. Students' weekly learning time shortened—for boys even more than for girls—school-based learning focused primarily on completing tasks that students were expected to work on at home, and often there was no regular exchange comparable to face-to-face classroom discussion. There is strong evidence that students were less cognitively activated in their mathematics learning in part due to fewer communicative exchanges and the fact that more time was allocated for the procedural aspect of mathematics than to conceptual understanding. Because of the reduced face-to-face contact between teachers and students, there was also apparently less feedback provided and less individualized adaptive learning support. Thus, it can be assumed that the changed school situation also had an impact on students' performance.

To analyze this assumption, the present study addresses the following research questions.

RQ1: Did third-grade students demonstrate lower math performance in 2021 than third-grade students in 2019?

Referring to the results of national and international studies, we hypothesize lower math performance in 2021.

RQ2: Are there differences between 2019 and 2021 in student performance with respect to the content domains?

Given that there was a focus on procedural learning (arithmetic, algorithms, and skills) during the distance-learning phases (Aldon et al., 2021), we hypothesize that performance in the content domain of Numbers & Operations was less affected than that in other domains.

RQ3: Are there differences between 2019 and 2021 in student performance with respect to gender?

This is an open question. Some surveys provide evidence for differential gender experiences mostly for secondary school students during the pandemic (e.g., Kantar Market Research Institute, 2021; Lockl et al., 2021; Sturzbecher et al., 2021). There is only little information about different gender experiences in primary school, and only a few studies analyzed performance data according to gender differences (Engzell et al., 2021).

RQ4: Are there differences between 2019 and 2021 in student performance with respect to requirement level?

Results from previous studies suggest that instruction and mathematical learning were less cognitively demanding and placed more emphasis on skill mastery (Aldon et al., 2021), so it can be hypothesized that performance on simpler tasks focusing on skill mastery was less affected by the altered school situation than performance on more complex problem-solving tasks that require, for example, making connections or generalizations.

4. Materials and methods

4.1. Sample

This study is based on data from two Germany-wide pilot studies in elementary schools (grade 3) with an overlapping set of mathematics items. The data were collected in May 2019 and May 2021 as preparation of the nationwide written comparison tests VERA (*Vergleichsarbeiten*), which students take in the third grade in Germany as part of the national educational monitoring strategy. The samples were comparable in that they were drawn as stratified cluster-random samples that were intended to cover the main characteristics of German third graders with respect to the outcome. A total of 1,905 students from 100 classes in eight German federal states completed the mathematics assessment in the pilot study in 2019 and 3,203 students from 172 classes in mostly different eight German federal states in 2021. There was always one class per school sampled. The students' mean age was 9.48 years ($SD = 0.65$) in 2019 and 9.52 years ($SD = 0.62$) in 2021. In 2019 and 2021, 49.0 and 50.3% of the children were female, respectively.

4.2. Measures

In both years, the test consisted of two parts, with working times of 40 min (part 1) and 20 min (part 2) and a break of 10 min in between.

For our analysis, we used data from 537 test items. Among these, 85 were identical, 145 items were uniquely administered in 2019 and 307 items were uniquely administered in 2021.

The test items were developed by teachers in close collaboration with experts in mathematics education and psychometricians and were based on the proficiency level models that were developed according to the German Educational Standards as briefly described above. They cover all five content domains defined in the German Educational Standards. The items have multiple-choice format (e.g., "Put a cross next to the drawing showing this building from above."—Content domain: Space and Shape) and constructed-response formats (e.g., "Anna, Bert and Cesar are sitting on a bench. Write down all the ways the children can sit."—Content domain: Data, Frequency, and Probability) and were grouped into disjoint blocks that were combined to booklets according to a multi-matrix design (e.g., Frey et al., 2009). Because our performance measure is based on proficiency level models, we will use the terms performance and proficiency in the following interchangeably.

In order to examine measurement invariance assumptions as prerequisites for linking, we calculated differential item functioning (DIF) statistics according to the Educational Testing Service (ETS) classification (e.g., Penfield and Algina, 2006). We found 11 of the 85 items to exhibit DIF. We inspected these 11 items and the students' answers to these items in order to examine whether there could be any construct-irrelevant source of variance that could have caused the differential functioning. During this procedure, we found two items that had to be excluded for the year 2019 because their coding guidelines were misunderstood. The remaining nine DIF items were omitted from the link by relaxing measurement invariance for these items opting for partial invariance but were retained in the analysis with unique difficulty parameters for each year. Consequently, there were 74 common items remaining (13 *Numbers & Operations*, 19 *Space & Shape,* 12 *Pattern & Structures,* 21 *Measurement*, and 9 *Data, Frequency, Probability*).

4.3. Analysis

Generalized Linear Mixed Models (GLMMs) allow us to model item and person properties and assess their effects on the individuals' responses. In order to investigate the effects of the factors addressed in our research questions on students' correct responses (their score) to the 546 mathematics items in 2019 and 2021, we specified a linear logistic test model with additional error term in the GLMMs framework with a multi-level structure (students were nested in classes). Further, GLMMs allow for the specification of fixed and random parts. Our fixed-effects part consisted of the following predictors: *content domain*, *proficiency level* and *requirement level* as item features on the item side and *gender* and *year of measurement* (two distinct samples in 2019 and 2021) on the person side. Interactions with *year of measurement* were allowed for all factors in order to investigate differential changes for different types of items as well as for *boys* and *girls*. The models' random part consisted of random item and random person effects that allow for individual deviations for each item and each person from the predicted value of the models' fixed part.

The factor content domain consists of five levels (*Numbers & Operations, Space & Shape, Pattern & Structures, Measurement, and Data, Frequency, Probability*) that were effect coded. With effect coding, the intercept is equal to the grand mean of all five content domains. Thus, we were assuming that all content domains contribute with equal weight to the proficiency level models. The same was done for the factor gender. The factor requirement level consists of three levels (I–III) as described in the German Educational Standards. As there were only a few items in the third category, we decided to collapse categories II and III into one category (II-III). The binary factor requirement level was weighted effect coded. In weighted effect coding, all levels of the factor are weighted by their sample proportions. This was done because we assumed that the three requirement levels are not equally distributed in the potential item pool. The factor year of measurement was dummy coded with 2019 as reference group. Because the proficiency levels are based on a continuous scale, this covariate was added as continuous and centered at the medium stage (3 out of 5). Thus, the model intercept can be interpreted as the average across all content domains and all boys and girls at a moderate proficiency level and over not necessarily equally frequent levels of requirement level in the year 2019. The models were estimated using the lme4 R package (Bates et al., 2022). In lme4, the model specifies as score ~ year*content_domain + year*sex + year*requirement_level + year*proficiency_level + (1|class/student) + (1|item).

As mentioned above, with the exception of score and proficiency level, all variables were factors with 5,108 levels for student, 272 levels for class, 546 levels for item, five levels for content domain, two levels for gender, and two levels for requirement level. In addition to the main effects and interaction effects, we investigated also simple effects in order to explore the exact nature of our interactions using the *emmeans* R package (Lenth et al., 2022). In particular, we compared the mean of each content domain in the year 2021 to the corresponding mean in 2019. The same was done for *gender* and *requirement level*.

5. Results

5.1. Mathematical proficiencies in 2019 compared to 2021 in general (RQ1) and by domain (RQ2)

The results for answering RQ1 showed a negative main effect of the *year*, with a lower probability to respond correctly to an item in 2021 compared to 2019, therefore indicating an overall increase in item difficulty or a drop in performance ($p < 0.001$, see Table 1). This result can be interpreted as averaged over *content domains*, and for items at a moderate *proficiency level* (3 out of 5) for a population with similar characteristics regarding *gender*, controlled for differences in class means due to other variables, and for items with similar *requirement levels* as in our sample.

This general finding differed across the different *content domains* (RQ2): the interaction effects in Table 1 indicate an average decline for *Measurement* (*M*) and *Pattern & Structures* (*P&S*). Contrastingly, the drop compared to the average decline was significantly larger for *Data, Frequency*, *Probability* (*DFP*; $p = 0.006$) and *Space & Shape* (*S&S*; $p = 0.003$) on the one hand and smaller for *Numbers & Operations* (*N&O*; $p = 0.017$) on the other hand.

Figure 1 depicts these domain-specific effects in terms of model-derived estimates for the probability of a correct answer in the different *domains* and *years*, as well as the significance of the simple effect of the year for each *domain*. The probabilities to answer an item correctly were mostly homogenous for all domains in 2021, ranging from 38 to 40%. In contrast, in 2019 we could identify some heterogeneity; that is, a group of *domains* had correct answer probabilities of around 42% (*N&O, M*, *P&S*), and a group had probabilities around 48% (*S&S* and *DFP*), indicating a lower difficulty for the latter in 2019 and a larger decline between 2019 and 2021. Nevertheless, all simple effects except for *Numbers & Operations* and *Pattern & Structures* were significant, indicating a relevant change in the probability to answer an item correctly. Interestingly, seven (2 *DFP*, 5 *S&S*) out of the nine aforementioned items with severe DIF according to the ETS classification that were excluded from the link were omitted more often and answered correctly less often in 2021 than in 2019. In contrast, the last two DIF items (both *N&O*) showed no such differences in omission, but even had a higher rate of correct answers in 2021 than in 2019. That is, the qualitative analysis of the DIF items is in line with the quantitative effects reported above.

Regarding the first two research questions, we can confirm that students' proficiencies in mathematics decreased in general between 2019 and 2021 (RQ1), with differences between *domains* (RQ2). *Space & Shape* and *Data, Frequency, Probability* decreased more than the other domains, whereas *Numbers & Operations* decreased less. *Pattern & Structures* and *Measurement* did not decrease differentially from the average of all domains.

5.2. Gender effects (RQ3)

The main effect of *gender* in Table 1 indicates that *boys* were more likely to answer an item correctly than *girls* in 2019 ($p < 0.001$). For both genders the estimated probability of a correct response declined in 2021. This decline appeared to be slightly larger for girls. However, this interaction effect is not significantly different from zero ($p = 0.057$).

As can be seen in Figure 2, the probability of a correct answer for girls declines from 42 to 34% between 2019 and 2021, and for boys from 47 to 42%, also showing a slightly larger decline for girls. Both drops can be regarded as small effects in terms of Cohen's *d* (−0.161 and −0.092, respectively), when converting the corresponding odds ratios to effect size *d* according to Sánchez-Meca et al. (2003).

Thus, to answer the third research question, we can attest that the performances of *boys* and *girls* do not decline significantly differently, but clearly do decline and more for girls than for boys.

TABLE 1 Results of the GLMM: main effects and interactions with year of measurement.

Parameter	Estimate	*SE*	*p*	
Fixed effects				
Intercept	−0.25	0.07	<0.001	***
Content domain Data, Frequency, Probability (DFP)	0.20	0.06	0.002	**
Content domain Numbers & Operations (N&O)	−0.13	0.05	0.006	**
Content domain Measurement (M)	−0.05	0.05	0.273	
Content domain Pattern & Structures (P&S)	−0.10	0.05	0.058	
Content domain Space & Shape (S&S)	0.09	0.05	0.052	
Gender	0.20	0.05	<0.001	***
Requirement level	−0.18	0.06	0.004	**
Proficiency level	−0.87	0.02	<0.001	***
Year of measurement	−0.27	0.08	<0.001	***
Year of measurement × DFP	−0.14	0.05	0.006	**
Year of measurement × N&O	0.11	0.05	0.017	*
Year of measurement × M	0.06	0.04	0.151	
Year of measurement × P&S	0.09	0.05	0.054	
Year of measurement × S&S	−0.12	0.04	0.003	**
Year of measurement × Gender	0.12	0.06	0.057	
Year of measurement × Requirement level	0.01	0.06	0.886	
Year of measurement × Proficiency level	−0.01	0.02	0.665	
Random effects				
σ^2	3.29			
$\tau_{00id:class}$	1.00			
τ_{00item}	0.18			
$\tau_{00class}$	0.30			
ICC	0.31			
N_{id}	5,066			
N_{class}	272			
N_{item}	546			
Observations	219,257			
Marginal R^2/Conditional R^2	0.25/0.48			

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

5.3. Task-related effects (RQ4)

To assess the influence of the complexity of the task we compared test items with a *requirement level* of I to those with a level of II or III. In 2019, the probability of a correct answer was higher for *requirement level* I than for level II–III ($p = 0.004$). This was also the case in 2021 (i.e., there is no significant interaction effect; see Table 1). Figure 3 shows that in both years items at *requirement level* I have a higher probability of being answered correctly (in 2019, about 47%; in 2021, about 41%) than items from level II-III (in 2019, about 43%; in 2021, about 37%). The average probability for items at both levels declined in 2021 by 6% compared to 2019. Both declines are significant.

With regard to the fourth research question, we were therefore able to conclude that the decrease was not more severe for more complex tasks.

6. Discussion

One main objective of the present study was to compare students' mathematics performance in 2019 and 2021 with a repeated cross-sectional design to explore the extent to which the school closures due to the pandemic may have had an impact on students' learning.

Our results show a significant decrease in students' mathematical performance (RQ1). These results are in line with the results of some other studies (Engzell et al., 2021; Pier et al., 2021; Tomasik et al., 2021). Possible reasons for the decrease of mathematical performance in general are that teaching and learning during the pandemic was less cognitively demanding but more transmissive ("Quality of distance learning") and that procedural aspects of mathematical learning were more present than conceptual understanding in distance learning (Aldon et al., 2021). Another possible reason might be the decrease of effective learning time reported by students and parents.

Analyzing mathematical performance separately for the five content domains (RQ2) reveals a significant decrease in performance in all of these domains except *Numbers & Operations* and *Pattern & Structures*. The decrease is particularly pronounced in the domains *Space & Shape* and *Data*, *Frequency*, *Probability*. van der Berg et al. (2022) also reported the largest drop in performance for the domain of Data Handling, but the smallest for geometry, and our results are in complete contrast to those of Depping et al. (2021) who found no significant difference in students' performance in the content domains *Space & Shape* and *Data, Frequency, Probability* between 2019 and 2020. However, Depping et al. (2021) compared students' performance in spring 2019 in grade 3 with students' performance in August/September 2020 in grade 4 after the pandemic. Both tests had the same items, but in 2020, the students were already in grade 4. The results could be biased by the different points of measurement. In addition, students were on summer vacation immediately before they were tested in 2020, and many schools organized and used catch-up programs during vacation so that learning deficits could be compensated (Depping et al., 2021).

Several factors may have caused performance in the domains *Space & Shape* and *Data, Frequency, Probability* to decline more clearly than the average of all domains. The fact that performance decreased particularly in the domain *Space & Shape* could be due to fewer hands-on activities having taken place during home schooling. Geometric activities (such as cutting, folding, building with cubes, moving in space, etc.) can lead to a deeper understanding and thus better performance in this domain, but these activities were probably rarely carried out in distance-learning settings. This is similarly the case for the domain of *Data*, *Frequency*, *Probability*, where hands-on experiments play an important role as well (Winter, 1976). A lack of classroom discussion could also be a reason for the drop in

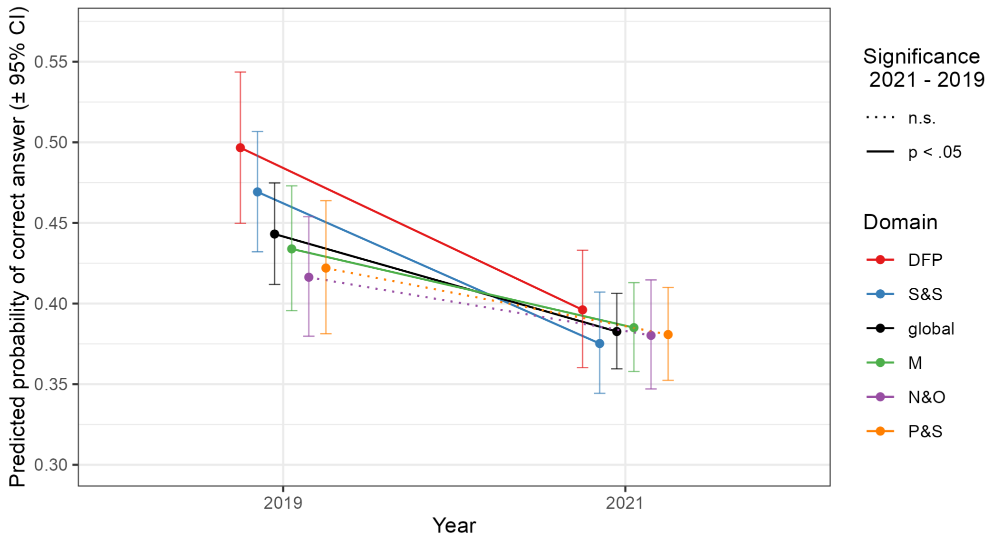

FIGURE 1
Interaction of domain and year of measurement via simple effects including a test of significance of the difference between 2019 and 2021. The simple effect of the average of all domains ("global") is shown as a reference (black line).

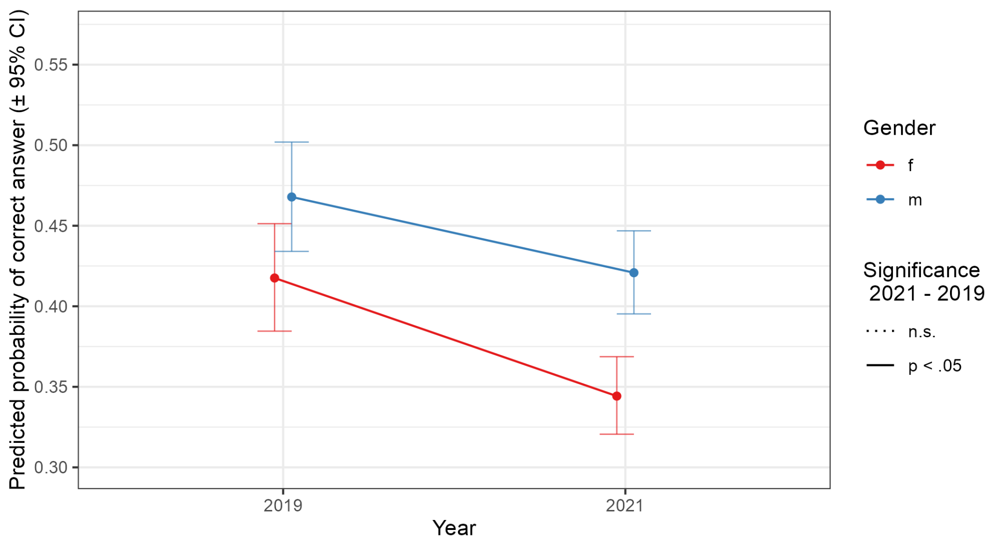

FIGURE 2
Interaction of gender and year of measurement via simple effects including a test of significance of the difference between 2019 and 2021.

performance in the *Data*, *Frequency*, *Probability* domain. Problem solving in this domain involves reasoning about probabilities or graphs. Coskun and Kara (2022) showed that school closures had a negative impact on mathematical reasoning skills. With lower reasoning skills, students may have difficulty discussing whether one random event is more likely than another and why, and therefore difficulty in solving problems in this domain (Coskun and Kara, 2022). Moreover, as discussed above, students performed better in 2019 in the domains *Space & Shape* and *Data*, *Frequency*, *Probability* than in the other domains. These two domains were also tested in the VERA-test conducted nationwide in 2019, and teachers may therefore have prepared their students in these domains. Unfortunately, it is not possible to disentangle these effects in our study.

Numbers & Operations is the only content domain where the decline in performance between 2019 and 2021 is less than the average decline across all domains. We expected that performance in the content domain *Numbers & Operations* would be less affected by the school closures, as it could be assumed that teachers concentrated on skills—especially on calculating ("Quality of distance learning"). Meeter (2021) study provides evidence that time in distance learning was used primarily to practice skills, and the results of Aldon et al. (2021) corroborate this: calculating or exercising algorithms, which are both part of the content domain *Numbers & Operations,* were clearly preferentially done in distance learning. Another possible explanation would be that in the content domain *Numbers & Operations* in particular, parents were better able to provide support

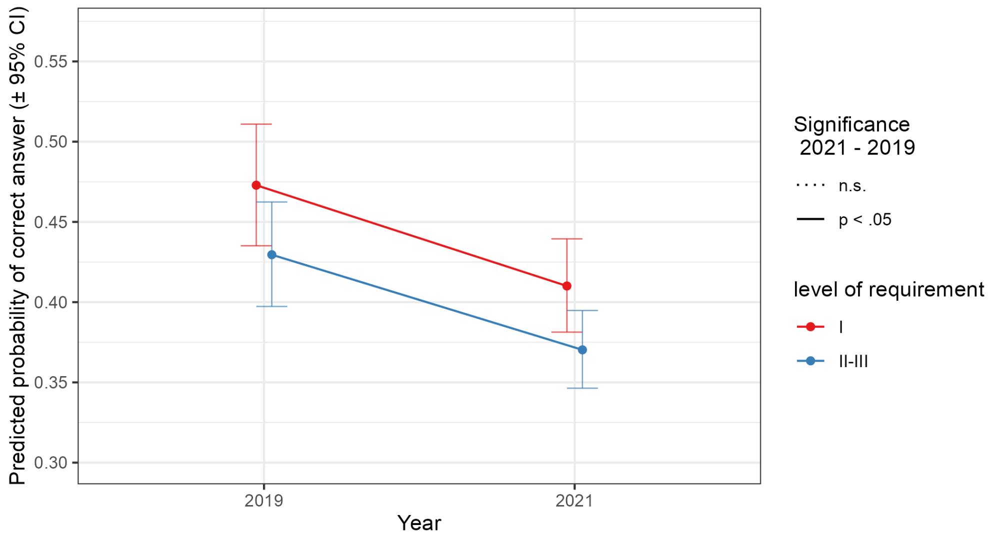

FIGURE 3
Interaction of requirement level and year of measurement via simple effects including a test of significance of the difference between 2019 and 2021.

in distance learning than, for example, in the content domains *Space & Shape* or *Data, Frequency, Probability*.

With regard to *gender*, there were no significant differences between girls' and boys' decline in performance between 2019 and 2021 (RQ3). This is in line with the results of Engzell et al. (2021), where also no gender differences were found. The study by Engzell et al. (2021) was carried out in the Netherlands, where schools were closed for only 8 weeks. In Germany, on the other hand, there was a special situation at schools that lasted for more than 1 year, characterized by alternating instruction and repeated temporary school closures. According to the results of the survey studies (e.g., Kantar Market Research Institute, 2021; Lockl et al., 2021; Sturzbecher et al., 2021), older boys and girls had different difficulties during this period. The similarity of performance results between boys and girls suggests that this was not the case for primary school children or, if it was, did not affect boys' and girls' performance in mathematics differently.

In addition, it was of interest whether the unusual situation in schools had a different effect on items with higher or lower *requirement levels* (RQ4). In contrast to our assumption, the observed decrease was similar for less and more complex tasks. It may be that these results reflect the fact that children had different conditions in learning at home, which is also influenced by socioeconomic status (Schneider et al., 2022; Orbach et al., 2023). For example, it could be that children who received a low degree of parental support in learning at home were not even able to solve reproduction tasks at a lower requirement level. This may have led to the significant decrease in performance even on less complex tasks (Level I). Targeted training in these skills did not appear to prevent performance from deteriorating.

7. Limitations and perspectives for future directions

Our study shows how primary student's mathematical proficiencies changed between 2019 and 2021. Stanat et al. (2022) already reported a substantial decline in mathematics achievement in 2021 compared with 2016, arguing that pandemic-related constraints were at least partly responsible for the observed changes, but the time period from 2016 to 2021 is comparatively long. Our study compared mathematics achievement in 2019—the last school year not affected by the pandemic—with achievement in 2021, when pandemic-related constraints were just beginning to phase out. The results of our study can therefore be taken as an indication that school closures and distance learning during the pandemic did indeed have an impact on students' mathematics achievement.

A limitation of our study is that we have no data on students' general cognitive ability, their socioeconomic status (SES) or overall mathematics achievement levels. This information would have been very interesting as many studies have shown a greater drop in performance for low-achieving students or students with low SES (Engzell et al., 2021; Schult et al., 2022). Another limitation is that our study has a cross-sectional design, with no information regarding individual developmental processes, and that the two assessments were carried out to a great extent in different federal states. It must also be mentioned that the numerous surveys referred to in the theoretical background section do not always allow for clear statements to be made about the primary level. In interpreting the results, however, the findings from the surveys were used in part without being able to say with certainty to what extent the survey results also apply to the primary level.

Nevertheless, the significant drop in performance that varied by *content domain* does suggest that the school closures and all pandemic-related constraints did have effects. Since mathematics is a hierarchical subject, this drop in performance is likely to have serious implications for further research and practice. It would be important to keep track of students' learning development and to investigate whether and how the gaps that have opened up can be closed again in the longer term. In upcoming comparative studies, it would be useful to conduct additional surveys on students' coping during school closures and distance learning to generate further insights. Furthermore, these results can be used to inform policy decisions in similar situations, regardless of the country.

To counter the reported effects in practice, it is important for teachers to take into account achievement differences among students in planning and teaching lessons. It is important for teachers to use informal diagnostic measures as soon as possible, and to use these consistently over the following several years to gain insight into their students' learning after this exceptional situation. Furthermore, in addition to appropriate interventions in the classroom, effective support programs should be set up as soon as possible for children who are particularly far behind. Since the content in mathematics builds strongly on each other, closing existing gaps should be given absolute priority.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the datasets presented in this article are subject to restrictions governing their use and disclosure, which limit their availability. Access to the data may be granted through a contractual agreement with the IQB. Requests for access should be directed to the designated point of contact at the IQB, who will assess the request and determine whether access can be granted, subject to the relevant legal and ethical considerations. Requests to access these datasets should be directed to https://www.iqb.hu-berlin.de.

Ethics statement

Ethical approval was not required for the studies involving humans because no personal data were collected. It is not possible to draw conclusions about the individual in this study. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

HG, MG, and AS were involved in the design of the items. ME-K organized the pilot study and the development process of the items. KSa and KSc performed the data analysis. HG, KSa, and KSc were responsible for the first draft of the manuscript. HG, KSa, KSc, MG, AS, and ME-K contributed to the interpretation of the results and drafted parts of the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aldon, G., Cusi, A., Schacht, F., and Swidan, O. (2021). Teaching mathematics in a context of lockdown: a study focused on teachers' praxeologies. *Educ. Sci.* 11, 1–21. doi: 10.3390/educsci11020038

Anger, S., Bernhard, S., Dietrich, H., Lerche, A., Patzina, A., Sandner, M., et al. (2020). Schulschließungen wegen Corona: Regelmäßiger Kontakt zur Schule kann die schulischen Aktivitäten der Jugendlichen erhöhen [school closures due to Corona: Regular contact with school can increase young people's school activities]. IAB-Forum. Available at: www.iab-forum.de/schulschliessungen-wegen-corona-regelmassiger-kontakt-zur-schule-kann-die-schulischen-aktivitaten-der-jugendlichen-erhohen/

Attig, M., Wolter, I., and Nusser, L. (2020). Zufriedenheit in unruhigen Zeiten: Welche Rolle die Kommunikation zwischen Eltern und Schulen während der Schulschließungen gespielt hat—Einschätzungen von Eltern zur Unterstützung durch die Schule und zum Lernerfolg ihrer kinder während des lockdowns [satisfaction in turbulent times: what role communication between parents and schools played during school closures—Parents' assessments of school support and their children's learning success during the lockdown]. LifBi. Available at: www.lifbi.de/Portals/13/Corona/NEPS_Corona-und-Bildung_Bericht_4-Schulkommunikation.pdf?ver=cG1QEP_NrmvYukcreWg5zw%3D%3D

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2022). lme4: Linear mixed-effects models using "Eigen" and S4

Blömeke, S., Olsen, R. V., and Suhl, U. (2016). "Relation of student achievement to the quality of their teachers and instructional quality" in *Teacher Quality, Instructional Quality and Student Outcomes. IEA Research for Education*. eds. T. Nilsen and J.-E. Gustafsson, vol. *2* (Germany: Springer Nature), 21–50.

Bremm, N. (2021). Bildungsbenachteiligung in der Corona-Pandemie. Erste Ergebnisse einer multiperspektivischen Fragebogenstudie [educational disadvantage in the Corona pandemic. First results of a multi-perspective survey]. *PraxisForschungLehrer*innenBildung* 3, 54–70. doi: 10.11576/pflb-3937

Common Core State Standards Initiative (2010). Common Core state standards for mathematics. Available at: https://www.nctm.org/uploadedFiles/Standards_and_Positions/Common_Core_State_Standards/Math_Standards.pdf

Coskun, K., and Kara, C. (2022). The impact of the COVID-19 pandemic on primary school students' mathematical reasoning skills: a mediation analysis. *Lond. Rev. Educ.* 20, 1–16. doi: 10.14324/LRE.20.1.19

Darragh, L., and Franke, N. (2021). Lessons from lockdown: parent perspectives on home-learning mathematics during COVID-19 lockdown. *Int. J. Sci. Math. Educ.* 20, 1521–1542. doi: 10.1007/s10763-021-10222-w

Depping, D., Lücken, M., Musekamp, F., and Thonke, F. (2021). "Kompetenzstände hamburger Schüler*innen vor und während der Corona-Pandemie [Competence Levels of Hamburg Students Before and During the Corona Pandemic]" in *Schule Während der Corona-Pandemie. Neue Ergebnisse und Überblick über ein Dynamisches Forschungsfeld (Die Deutsche Schule, Suppl. 17)*. eds. D. Fickermann and B. Edelstein (Germany: Waxmann Verlag), 51–79.

Engzell, P., Frey, A., and Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–7. doi: 10.1073/pnas.2022376118

Fobizz. (2020). So sehen Lehrkräfte die Corona-Krise in der Bildung [how teachers percieve the Corona crisis in education]. Available at: https://fobizz.com/umfrage-ergebnisse-so-sehen-lehrkraefte-die-corona-krise-in-der-bildung/ (Accessed June 2, 2022).

Forsa (2020). Homeschooling in Corona-Zeiten. Erfahrungen von Eltern und Schülern [homeschooling in times of Corona. Experiences of parents and students]. DAK. Available at: https://www.dak.de/dak/download/forsa-studie-2266768.pdf

Frey, A., Hartig, J., and Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: theory and practice. *Educ. Meas. Issues Pract.* 28, 39–53. doi: 10.1111/j.1745-3992.2009.00154.x

Gore, J., Fray, L., Miller, A., Harris, J., and Taggart, W. (2021). The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study. *Aust. Educ. Res.* 48, 605–637. doi: 10.1007/s13384-021-00436-w

Heller, S., and Zügel, O. (2020). Schule zu Hause"in Deutschland. Bestandsaufnahme im Corona-lockdown aus Perspektive der Schüler/−innen und Eltern ["School at Home" in Germany. Stocktaking in the Corona lockdown from the perspective of students and parents]. Deutsche Telekom Stiftung. Accelerom. Available at: https://www.telekom-stiftung.de/sites/default/files/files/media/publications/Ergebnisbericht-Homeschooling.pdf

Helm, C., Huber, S., and Loisinger, T. (2021). Was wissen wir über schulische Lehr-Lern-Prozesse im Distanzunterricht während der Corona-Pandemie? – Evidenz aus Deutschland, Österreich und der Schweiz [Meta-Review on findings about teaching and learning in distance education during the Corona pandemic – Evidence from Germany, Austria and Switzerland]. *Z. Erzieh.* 24, 237–311. doi: 10.1007/s11618-021-01000-z

Huber, S. G., Günther, P. S., Schneider, N., Helm, C., Schwander, M., Schneider, J. A., et al. (2020). *COVID-19 und aktuelle Herausforderungen in Schule und Bildung. Erste Befunde des Schul-Barometers in Deutschland, Österreich und der Schweiz [COVID-19 and Current Challenges in Schools and Education. First Results of the School Barometer in Germany, Austria and Switzerland]*. Germany: Waxmann Verlag

Huber, S. G., and Helm, C. (2020). 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.* 32, 237–270. doi: 10.1007/s11092-020-09322-y

Institut für Demoskopie Allensbach (2021). Lernen in Zeiten von Corona. Ergebnisse einer Befragung von Schülern und Eltern von Kindern der Klassenstufen 5 bis 10 im Frühjahr 2021 [learning in times of Corona. Results of a survey of students and parents of children in grades 5 to 10 in spring 2021]. Available at: https://www.telekom-stiftung.de/sites/default/files/files/media/publications/Lernen-in-Zeiten-von-Corona-Bericht.pdf

Kantar Market Research Institute (2021). GoStudent education report–quantitative research findings Europe. Available at: https://www.gostudent.org/static/images/educational_reports/GoStudent%20Educational%20Report%202021.pdf

KMK (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich [Educational Standards in Mathematics for Primary Education]. Beschluss vom 15.10.2004*. Germany: Luchterhand

Lampert, C., and Thiel, K (2021). Mediennutzung und Schule zur Zeit des ersten lockdowns während der Covid-19-Pandemie 2020. Ergebnisse einer online-Befragung von 10-bis 18-Jährigen in Deutschland [media use and school at the time of the first lockdown during the Covid-19 pandemic in 2020. Results of an online survey of 10-to 18-year-olds in Germany]. Verlag Hans-Bredow-Institut (No. 53). Available at: https://www.hans-bredow-institut.de/uploads/media/Publikationen/cms/media/k7d80jh_AP53_KiDiCoTi.pdf

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., et al. (2022). Emmeans: Estimated marginal means, aka least-squares means (1.7.4–1) [computer software]. Available at: CRAN.R-project.org/package=emmeans

Lockl, K., Attig, M., Nusser, L., and Wolter, I. (2021). Lernen im Lockdown: Welche Voraussetzungen helfen Schülerinnen und Schülern? [Learning in lockdown: what condition help students?]. LifBi. Available at: https://www.lifbi.de/Portals/13/Corona/NEPS_Corona-und-Bildung_Bericht_5-Motivation.pdf

Ludewig, U., Kleinkorres, R., Schaufelberger, R., Schlitter, T., Lorenz, R., König, C., et al. (2022). COVID-19 pandemic and student reading achievement: findings from a school panel study. *Front. Psychol.* 13:876485. doi: 10.3389/fpsyg.2022.876485

Maldonado, J. E., and de Witte, K. (2022). The effect of school closures on standardised student test outcomes. *Br. Educ. Res. J.* 48, 49–94. doi: 10.1002/berj.3754

Mantasiah, R., Yusri Sinring, A., and Aryani, F. (2021). Assessing verbal positive reinforcement of teachers during school from home in the COVID-19 pandemic era. *Int. J. Instr.* 14, 1037–1050. doi: 10.29333/iji.2021.14259a

Meeter, M. (2021). Primary school mathematics during Covid-19: no evidence of learning gaps in adaptive practicing results. *Trends Neurosci. Educ.* 25:100163. doi: 10.1016/j.tine.2021.100163

Nusser, L., Wolter, I., Attig, M., and Fackler, S. (2021). "Die Schulschließung aus Sicht der Eltern [The school closure from the parents' perspective]" in *Schule Während der Corona-Pandemie. Neue Ergebnisse und Überblick Über Ein Dynamisches Forschungsfeld (Die Deutsche Schule, Suppl. 17)*. eds. D. Fickermann and B. Edelstein (Germany: Waxmann Verlag), 33–50.

OECD (2020). *Global Teaching Insights: A Video Study of Teaching*. Berlin: OECD Publishing

OECD/European Union. (2020). *Health at a Glance: Europe 2020: State of Health in the EU Cycle*. Berlin: OECD Publishing

Orbach, L., Fritz, A., Haase, V. G., Dowker, A., and Räsänen, P. (2023). Conditions of distance learning and teaching and their relation to elementary school children's basic number skills after the suspension of face-to-face teaching during the COVID-19 pandemic. *Front. Educ.* 8:1083074. doi: 10.3389/feduc.2023.1083074

Penfield, R. D., and Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *J. Educ. Meas.* 43, 295–312. doi: 10.1111/j.1745-3984.2006.00018.x

Pier, L., Christian, M., Tymeson, H., and Meyer, R. H. (2021). COVID-19 impacts on student learning: evidence from interim assessments in California. Policy Analysis for California Education. Available at: https://edpolicyinca.org/publications/covid-19-impacts-student-learning

Russo, J., Bobis, J., Downton, A., Livy, S., and Sullivan, P. (2021). Primary teacher attitudes towards productive struggle in mathematics in remote learning versus classroom-based settings. *Educ. Sci.* 11:35. doi: 10.3390/educsci11020035

Sánchez-Meca, J., Marín-Martínez, F., and Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychol. Methods* 8, 448–467. doi: 10.1037/1082-989X.8.4.448

Schneider, R., Enke, F., Sachse, K. A., and Schipolowski, S. (2022). "Lernbedingungen im Fern-und Wechselunterricht während der coronavirus-Pandemie [Learning conditions in distance and alternate teaching during the coronavirus pandemic.]" in *IQB-Bildungstrend 2021*. eds. P. Stanat, S. Schipolowski, R. Schneider, K. A. Sachse, S. Weirich and S. Henschel (Germany: Waxmann Verlag), 115–126.

Schneider, R., Sachse, K. A., Schipolowski, S., and Enke, F. (2021). Teaching in times of COVID-19: the evaluation of distance teaching in elementary and secondary schools in Germany. *Front. Educ.* 6:702406. doi: 10.3389/feduc.2021.702406

Schober, B., Lüftenegger, M., Spiel, C., Holzer, J., Ikanovic, S. K., Pelikan, E., et al. (2020). Lernen unter COVID-19-Bedingungen. Erste Ergebnisse–Schüler*innen [Learning under COVID-19 conditions. First results—students]. Available at: https://lernencovid19.univie.ac.at/fileadmin/user_upload/p_lernencovid19/Zwischenergebnisse_Schueler_innen.pdf

Schoenfeld, A. (2018). Video analyses for research and professional development: the teaching for robust understanding (TRU) framework. ZDM–mathematics. *Education* 50, 491–506. doi: 10.1007/s11858-017-0908-y

Schult, J., Mahler, N., Fauth, B., and Lindner, M. A. (2022). Did students learn less during the COVID-19 pandemic? Reading and mathematics competencies before and after the first pandemic wave. *Sch. Eff. Sch. Improv.* 33, 544–563. doi: 10.1080/09243453.2022.2061014

Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., and Henschel, S. (2022). Zusammenfassung und Einordnung der Befunde [Summary and contextualization of the findings]. IQB-Bildungstrend 2021. Germany: Waxmann Verlag, 115–126.

Sturzbecher, D., Dusin, R., Kunze, T., Bredow, B., and Pöge, A. (2021). Jugend in Brandenburg 2020. Auswirkungen der Corona-Pandemie [youth in Brandenburg 2020. Impact of the Corona pandemic]. MBJS Brandenburg. Available at: https://mbjs.brandenburg.de/media_fast/6288/ifk-jugendstudie_2020_-_auswirkungen_der_corona-pandemie.pdf

Tomasik, M. J., Helbling, L. A., and Moser, U. (2021). 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.* 56, 566–576. doi: 10.1002/ijop.12728

UNESCO Institute of Statistics (2017). Dashboards on the global monitoring of school closures caused by the COVID-19 pandemic. Available at: https://covid19.uis.unesco.org/global-monitoring-school-closures-covid19/ (Accessed August 10, 2023).

van der Berg, S., Hoadley, U., Galant, J., van Wyk, C., and Böhmer, B. (2022). "Learning losses from COVID-19 in the Western cape; evidence from systemic tests", in *Research on Socio Economic Policy (Resep). chapter* 4. Stellenbosch University, 62–87.

Wacker, A., Unger, V., and Rey, T. (2020). ""Sind doch Corona-Ferien, oder nicht?". Befunde einer Schüler*innenbefragung zum "Fernunterricht. ["It's a Corona holiday, isn't it?" results of a student survey on distance learning.]"" in *Langsam Vermisse Ich Die Schule… Schule Während Und Nach der Corona-Pandemie, (Die Deutsche Schule, Suppl. 16)*. eds. D. Fickermann and B. Edelstein (Germany: Waxmann Verlag), 79–94.

Winter, H. (1976). Erfahrungen zur Stochastik in der Grundschule (Klasse 1-6) [experiences with stochastics in primary school (grades 1 to 6)]. *Didak. Mathemat.* 1, 22–37.

Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., and Zierow, L. (2020). *Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen? [Education in the Corona crisis: How have students spent the period of school closures, and what educational measures do Germans favor?], ifo institute–Leibniz Institute for Economic Research at the university of Munich Ifo Schnelldienst* 73, 25–39.