International Journal of Psychology

International Journal of Psychology, 2021
Vol. 56, No. 4, 566–576, DOI: 10.1002/ijop.12728

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

Martin J. Tomasik[1,2], Laura A. Helbling[2], and Urs Moser[2]

[1]Institute for Educational Evaluation, University of Zurich, Zurich, Switzerland
[2]Department for Psychology and Psychotherapy, University of Witten-Herdecke, Witten, Germany

Using data from a computer-based formative feedback system, we compare learning gains in the 8 weeks of school closures related to the COVID-19 pandemic in Switzerland with learning gains in the 8 weeks before these school closures. The school performance in mathematics and language of $N = 28{,}685$ pupils is modelled in second-order piecewise latent growth models with strict measurement invariance for the two periods under investigation. While secondary school pupils remain largely unaffected by the school closures in terms of learning gains, for primary school pupils learning slows down and at the same time interindividual variance in learning gains increases. Distance learning arrangements seem an effective means to substitute for in-person learning, at least in an emergency situation, but not all pupils benefit to the same degree.

Keywords: COVID-19; Distance learning; Learning progress; School achievement; School closures.

In an attempt to contain the spread of the COVID-19 virus, most governments around the world have temporarily closed schools and other educational institutions, affecting more than 1.2 billion pupils and students or almost three-quarters of the learner population (United Nations Sustainable Development Group, 2020). Learners are probably the single largest group to experience the pandemic's indirect effects. Some researchers and organisations (e.g., Burgess & Sievertsen, 2020; Education Endowment Foundation [EEF], 2020; Kuhfeld et al., 2020) have projected that school closures during the pandemic could have detrimental effects on learning gains and social disparities in learning. To the best of our knowledge, there is no empirical evidence yet on the school closure effect's actual direction and size. There is mainly more or less substantiated speculation, as educational researchers, like many other stakeholders, were unprepared for and overwhelmed by the situation. Few expected the pace and scope of the pandemic's development in spring 2020.

In this article, we analyse a coincidentally ongoing data collection to provide timely empirical evidence on the impact of distance learning in schools. If there was an effect of the school closures on learning and if this effect was sufficiently large to be reliably measured during the relatively short period of time, this would not only be relevant on its own or helpful to identifying pupils at-risk that would require special attention in the case of another school closure potentially to follow. We also know that educational achievement can have cascading effects into other developmental domains, such as employment or health and affect other developmental outcomes such us income or civic engagement—even years later. Of course, we cannot provide any evidence on these long-term effects with the data at hand. However, we wanted to mention them to indicate the potentially broad impact of the findings for the individual and the society as a whole. Because education is correlated with virtually every psychological trait and because it moderates many psychological processes, this article is meant as a more

Correspondence should be addressed to Dr Martin J. Tomasik, Institut für Bildungsevaluation, Wilfriedstrasse 15, 8032 Zurich, Switzerland. (E-mail: martin.tomasik@ibe.uzh.ch)

MT developed the concept of the study, drafted the manuscript and was responsible for the statistical hypothesis testing; LH prepared the data and supported MT in the data analyses; UM developed the computer-based formative feedback system, provided the data and revised the manuscript.

general contribution to the broader discussion on the psychological effects of the pandemic.

PANDEMIC POLICY ON EDUCATION IN SWITZERLAND

In Switzerland, all educational institutions closed on 16 March 2020 and reopened again on 11 May 2020. During these 8 weeks, schools virtually overnight switched to distance learning. It is safe to assume that most pupils, parents and teachers were unprepared for the situation. The school closures were announced at very short notice on the Friday afternoon before taking effect the following Monday. School authorities at first did not provide any guidance for parents and teachers on how to deal with the situation's challenges, resulting in high uncertainty on all sides. In the first days, for instance, it was not even clear whether pupils were allowed to pick up their workbooks that they had left in school. By the end of March, however, school authorities had put together and distributed information on best practice in distance instruction, and many EdTech companies had made their distance learning applications available to schools. Moreover, decisions had been made on how to handle attendance, grading and progression in the second half of the school year. While the school closures were mandated by federal directive, the definition of specific regulations and their implementation in school practice was decided and supervised by the 26 individual cantons. This makes it difficult to summarise them in brief. In most cantons, grading was suspended sooner or later and special regulations came into effect regarding attendance and progression. Pupils often received no grades until the end of the school year, even after schools reopened. Many pupils therefore received only a pro forma school report at the end of the 2019/2020 school year.

POTENTIAL IMPACT OF SCHOOL CLOSURES ON EDUCATIONAL GAINS

Although the situation leading to the school closures in the second term of the 2019/2020 school year and its scope are unprecedented in recent history, studies on school non-attendance for reasons other than a pandemic may be informative for estimating the effects of school closures and the rapid transition to distance learning due to COVID-19. In reviewing these studies, we orientate ourselves by the few highly topical works that have been published in a remarkably short time: the projection study by Kuhfeld et al. (2020), the short article by Burgess and Sievertsen (2020) and the meta-analysis by the EEF (2020). In all three publications, based on previous research, the authors expect that the school closures will have an impact on learning gains, although they disagree on its order of magnitude. Notably, none of these papers analyse data collected during the actual pandemic. Evidence is cited from studies on seasonal learning and school closures during natural disasters, comparative studies on instructional time and school absenteeism studies.

Seasonal learning studies

There is a long tradition of studies scrutinising the institutional effects on learning from the learning loss during regular school vacations. Early works on summer learning loss, summarised and meta-analysed, for instance, by Cooper et al. (1996) suggest not only that achievement declines over the summer break to an extent equivalent to one month of school learning but also that there are social disparities in this effect that contribute to a growing heterogeneity in achievement. More recent studies, however, question not only the summer learning loss but also the social disparity effect (e.g., von Hippel & Hamrock, 2019) and argue with the methodological issues related to scaling, the use of different test forms, the choice between a regressor variable model or a change score model and the way of modelling measurement error, asserting that all these issues can jeopardise the valid interpretation of findings from studies on learning loss. In any case, a large variance of effect sizes is reported, with estimates ranging from no loss at all up to $d = .010$ per vacation day. Extrapolating these figures to the 8 weeks or 40 days of school closures during the COVID-19 pandemic would result in learning losses of $.000 \leq d \leq .400$.

Comparative studies on instructional time

As Burgess and Sievertsen (2020) argue, studies investigating differences in instructional time on educational outcomes might be informative for estimating the impact of school closures related to COVID-19. The authors cite as exemplars two such studies that try to establish causality by investigating dosage-response patterns. In the study by Carlsson et al. (2015), Swedish males took a battery of cognitive tests in preparation for military service that randomly varied in date and hence in the time for preparation in school. In this study, just 10 days of extra schooling raised the scores of the recruits on the crystallised intelligence test by $d = .010$. Lavy (2015) took another approach by explaining international achievement gaps found in the Programme for International Student Assessments (PISA) by differences in schools' instruction time in the various countries. He found that one more hour per week in the main subjects increased test scores by about $d = .060$. Of course, this correlative study can only statistically control for all the other differences that exist between the educational systems in the various countries. Taking these two effect sizes and extrapolating them to the

40 days of school closures in Switzerland would result in a learning loss of only $d = .040$.

Studies on school absenteeism

Another strand of research relevant in the present context deals with school absenteeism and compares the learning progress of those who attend school with the learning progress of those who miss some lessons, hence, also applying a dosage-response paradigm. There are numerous reasons why pupils do not attend school, including their own illness or lack of access to reliable transportation. Minority status and low family income are also important correlates for school absenteeism. Research regularly reports a linear association between the number of days missed in school and end-of-year test scores, although the range of effect sizes reported tends to be large. While Aucejo and Romano (2016), Gershenson et al. (2017) and Liu et al. (2020) consistently report effect sizes of $.006 \leq d \leq .008$ for each school day missed, Goodman (2014) finds that one single day of absence reduces the pupils' mathematics scores by as much as $d = .050$. Extrapolating this to the 40 days of school closures during the COVID-19 pandemic would amount to a learning loss of $.240 \leq d \leq 2.000$.

School closures during natural disasters

Research on school closures due to natural disasters is less frequent, as these events rarely occur. Kuhfeld et al. (2020) identify three recent studies that observe the impact of severe weather events on learning. Hansen (2011) found that each day of school closures due to snow in Colorado reduced achievement by $.013 \leq d \leq .039$. Goodman (2014), analysing the effects of school closures due to snow in Massachusetts, found a comparable effect only for poor schools. The third study by Sacerdote (2012), which is not directly comparable in terms of the effect size, investigated learning loss in the aftermath of hurricane Katrina and found an overall effect of $d = .100$ due to displacement. Extrapolating the effect sizes of the snow studies to the 40-day school closures during the COVID-19 pandemic would yield effect sizes of $.510 \leq d \leq 1.560$, which is much larger than the overall effect reported for hurricane Katrina.

Generalisability of existing evidence to school closures during the COVID-19 pandemic

The generalisability of all four types of studies to the situation during the COVID-19 pandemic is somewhat limited for several reasons. Evidence from seasonal learning studies has the advantage of relying on numerous studies and sometimes very large samples. However, school vacations are predictable events that do not resemble sudden school closures. Parents and their children can prepare and often spend at least some vacation time together, especially when the children are still young. Comparative studies on instructional time are useful to estimate the effect of schooling but hardly control for cultural, curricular and other differences between single countries, so a number of alternative explanations are possible. Evidence from studies on school absenteeism probably offers the most direct way of studying the institutional influence on learning, and at the same time, large control groups of pupils attending school are available. However, a high degree of self-selection and low equivalence between attending and absent pupils on virtually all behavioural, cognitive and socio-demographic variables threaten the valid interpretation of the results as effects of schooling. Evidence from studies on school closures during natural disasters is probably best suited to generalise to the COVID-19 pandemic situation because such events usually occur suddenly, allow for little if any preparation (in contrast to seasonal learning studies), are locally confined, which limits the influence of cultural factors (in contrast to studies on instructional times), affect every single pupil and hence reduce self-selection bias (in contrast to studies on absenteeism). However, fortunately, such disasters are rare as is empirical evidence from large samples. Furthermore, the educational situation is different, as no systematic distance learning has so far been implemented as a means to compensate for in-person learning. All these reasons, together with the methodological uncertainties and the broad range of extrapolated effect size estimates ($.000 \leq d \leq 2.000$), make it extraordinarily difficult to provide reliable point estimates for the actual effect of the COVID-19 school closures.

Another substantial factor that significantly limits the comparability of existing studies—perhaps with the notable exception of some of those related to natural disasters—is that the COVID-19 pandemic brought about a high degree of psychological uncertainty for parents and their children, confronted many families with social, emotional and economic strain and compromised mental health in the general population (e.g., Serafini et al., 2020). Many working parents had to educate and care for their children besides meeting their jobs' old and new demands. Existing social ties, for instance, those to grandparents and other supportive networks, were disrupted. The risks of income or job loss during one of the major economic downturns in recent history together with tangible health threats were ubiquitous in many families. It is well known that such strains immediately translate from the distal to the proximal developmental contexts of individuals (Tomasik & Silbereisen, 2016) and can challenge the functioning of families, particularly those who are most vulnerable (Elder & Caspi, 1988; Silbereisen & Tomasik, 2011). All these considerations suggest that the effect of the school closures due to the COVID-19 pandemic could be larger than in previous

studies and produce more heterogeneity in the learning trajectories.

At the same time, learning continued (see Chamberlain et al., in press) and the institutional impact of schooling was not totally diminished during the COVID-19 school closures. On the contrary, teachers, in collaboration with the school authorities, quickly implemented digital forms of distance learning. It is possible that some teachers were incapable, unmotivated or both to develop pedagogically effective instruction in the digital space (see also Iivari et al., 2020). Research from the United States suggests that a concerning number of teachers lost contact with their pupils and did not interact with their pupils on a daily basis (Lieberman, 2020). This is consistent with models of instruction in which the *coordination* of pupils is the central challenge that teachers face. However, instruction continued, albeit under very different and sometimes quite demanding conditions (see also Basilaia & Kvavadze, 2020).

PRESENT STUDY

The unique situation of school closures related to the COVID-19 pandemic enables testing the effects of in-person versus distance learning in an unselected sample to estimate the potential loss in learning progress that could have occurred during this time. Two competing hypotheses can be scrutinised. On the one hand, one could assume that the lack of institutional schooling results in slower progress or even a decline in competence. This effect should be particularly pronounced for at-risk pupils, such as those from disadvantaged social backgrounds, those with learning difficulties or very young learners (see Cooper et al., 1996; Lee, 2020). On the other hand, one could argue that the institutional influence was not totally diminished and that learning took place in another form and maybe even at the same pace. Some studies even suggest that distance learning, if implemented well, can have advantages compared to traditional classes (e.g., Allen et al., 2006), although other studies report a negative impact of distance education (e.g., Ahn & McEachin, 2017). In the given emergency situation, one would not assume that distance learning would outperform in-person learning. However, advantages and disadvantages could be more or less balanced. The current pandemic situation offers a unique natural experiment to test these competing hypotheses by comparing learning progress before and during the school closures.

Even if the potentially negative effects of the school closures and the potentially positive effects of distance learning balance out or if the time intervals for observing any positive or negative effects are too small, one would nevertheless expect the variance of educational gains to significantly increase during school closures compared to regular in-person learning. As the institutional influence of school decreases, the effect of the family environment tends to increase, introducing another source of variance in learning. Furthermore, while some teachers and pupils might have found themselves completely unprepared for online distance learning, others might have already been using digital tools in their classrooms for some time. Teachers, on the one hand, not only differed in their experience but also in their motivation for using distance learning arrangements. Pupils, on the other hand, differed in their personal characteristics, such as age or self-regulatory learning skills, and found themselves in different contexts that made distance learning more or less effective. While some pupils were very well equipped at home and supported by their parents, others lacked basic material such as a working personal computer or their own desk or had to deal with dysfunctional family arrangements. Social background seems to play a pivotal role for support provided by schools, as surveys conducted with parents (Andrew et al., 2020) or teachers (Cullinane & Montacute, 2020) suggest. All these considerations are consistent with surveys from the UK that found, for instance, a huge variance in the proportion of students getting involved in school work at home as a function of the level of deprivation (Lucas et al., 2020). Taken together, one might assume that the variance in learning gains during the 8 weeks of the school closures would be larger than the variance in learning gains during the 8 weeks of regular in-person school attendance (see also van Lancker & Parolin, 2020).

We also hypothesised that the impact of the school closures would be more profound for younger compared to older pupils, as their capabilities for self-regulated distance learning are less developed, they require more cognitive scaffolding during learning and are probably more severely affected by the socioemotional strains that the COVID-19 pandemic brought about for families, as previous research on the effects of macrostructural disruption (e.g., Elder & Caspi, 1988) suggests. Also, limited physical activity due to restrictions requiring physical distancing, limited community interactions, as well as sports facilities, playground and park use (Moore et al., 2020) might be particularly detrimental for the younger children.

METHODS

Participants and procedure

All active users of the MINDSTEPS system who completed at least one teacher-generated assessment between 19 January 2020 and 11 May 2020 were included in the statistical analyses. MINDSTEPS is a computer-based formative feedback system developed at the Institute for Educational Evaluation in Zurich and deployed

TABLE 1
Number of assessments completed and number of pupils having completed these assessments

Assessments	*Before school closures*	*During school closures*	
Pupils	*Only before*	*Before and during*	*Only during*
Mathematics	13,816 (3,898/9,918)		50,760 (25,044/26,716)
	3,225 (983/2,242)	2,928 (1,197/1,731)	13,536 (7,337/6,199)
Reading	5,811 (1,893/3,918)		22,600 (12,401/10,199)
	2,350 (620/1,730)	1,340 (577/763)	9,364 (5,268/4,096)
Grammar	10,773 (3,451/7,322)		39,405 (18,931/20,747)
	3,036 (1,011/2,025)	2,984 (1,162/1,822)	12,997 (7,149/5,848)

Note: Total number of assessments and pupils on the left; respective numbers, divided into primary and secondary schools, on the right and in brackets.

to all pupils in the cantons of Aargau, Basel-Stadt, Basel-Landschaft and Solothurn and to many pupils from other German-speaking cantons of Switzerland. It serves pupils from grade 3 to 9 and covers the subjects of mathematics, German (the instructional language), English and French (the two foreign languages taught at schools), according to the official school curriculum. The system allows teachers to set up both linear and adaptive assessments as well as pupils to practice ad lib in these subjects. For the present analyses, we focused on teacher-generated assessments only in mathematics and German (in the reading comprehension and grammar domains) because instruction in these two subjects is most aligned between the single cantons. There were almost 15,000 items in mathematics and almost 10,000 items in German that could have been used to set up the single assessments. Although the system was originally developed to provide formative feedback to teachers and pupils, it can also be used to obtain ability estimates of pupils over time. Details of the system's theoretical rationale can be found in Tomasik et al. (2018). König et al. (2020) provide a hands-on demonstration of the system's capabilities. A total of $N = 28{,}685$ pupils ($N = 13{,}134$ in primary school and $N = 15{,}551$ in secondary school) were considered in the following analyses, although this sample size cannot be compared to a traditional study design, as the data are relatively sparse both between domains and over time. The numbers of pupils and assessments are summarised in Table 1. On average, pupils were $M = 9.20$ ($SD = 1.69$) years old in grade 3, $M = 10.20$ ($SD = 1.46$) years old in grade 4, $M = 11.21$ ($SD = 0.82$) years old in grade 5, $M = 12.20$ ($SD = 1.55$) years old in grade 6, $M = 13.17$ ($SD = 1.93$) years old in grade 7, $M = 14.49$ ($SD = .90$) years old in grade 8 and $M = 15.39$ ($SD = 1.63$) years old in grade 9. There were 50.3% boys and 30.8% non-native speakers in primary school as well as 49.1% boys and 32.7% non-native speakers in secondary school.

We divided the 8 weeks before the school closures and the 8 weeks during the school closures into eight intervals of 14 days. For each interval and each domain, we obtained the WLE ability estimates based on all items completed during that time. For WLE estimation, we used grade-by-grade one-parameter logistic models based on probabilistic measurement theory that were then vertically linked using the Stocking-Lord equation method. All the item parameters needed for this procedure were previously obtained from a much larger sample collected between September 2017 and February 2020 and fixed during WLE estimation. Berger et al. (2019) provide a justification of this procedure together with a validation of the item parameters obtained on the curriculum contents for mathematics.

Ethical Compliance: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. In line with the American Psychological Association's Ethical Principles and Code of Conduct, as well as with the Swiss Psychological Society's Ethical Guidelines, written informed consent from pupils and their parents was not required because this study was based on the assessment of normal educational practices and curricula in educational settings. The Institute for Educational Evaluation as a contractor of the cantonal educational authorities signed and committed to obeying the laws of the four cantons involved to ensure strict data confidentiality. In line with the laws of the four cantons, approval from an ethics committee was not required for this study.

Statistical modelling approach

School achievement was operationalised with eight latent factors, each representing one of the eight intervals, four before and four during the school closures. Within each interval, we used the respective three ability estimates in the three domains as manifest indicators of latent school achievement. Using these eight latent factors, we set up a multigroup second-order piecewise latent growth model (see Isiordia & Ferrer, 2016) with a joint intercept for the two periods under investigation but with separate estimates for primary school (grades 3–6) and secondary school (grades 7–9) pupils. The two central growth components we focused on in our analyses were

the *in-person learning slope* and the *distance learning slope*.

Before conducting the substantive analyses on these two, we first tested the measurement invariance properties of the latent factor model incorporating the eight single intervals. Reverting to a procedure suggested, for instance, by Ferrer et al. (2008), we started with a model representing configural invariance and subsequently tested models representing weak, strong and strict invariance across the eight intervals. Unlike Ferrer and colleagues, however, we parameterised weak invariance in terms of tau-equivalence by fixing all factor loadings to one. Following Chen (2007); as cited in Putnick & Bornstein, 2016), the criteria we used for assessing invariance were ΔCFI < .01, ΔRMSEA < .015 and ΔSRMR < .030 (for weak invariance only) or ΔSRMR < .015 (for strong and strict invariance). We used this model to statistically compare the two slopes and their associations with other variables by setting up equality constraints one by one.

All computations were conducted using the lavaan package for R, considering the nesting of pupils in school classes, employing the robust maximum likelihood estimator for parameter estimation and using full information maximum likelihood (FIML) for dealing with missing data. The decision for the latter, given the sparse data we faced, was made against the findings reported by Xiao and Bulut (2020), who found that FIML outperforms all other methods compared under most conditions.

RESULTS

Baseline model and measurement invariance

The baseline model with a common intercept across all eight intervals and with two different slopes for either in-person learning or distance learning fit the data well: $\chi^2(508) = 910.74$, $\chi^2/df = 1.79$, CFI = .972, RMSEA = .010, 90% CI_{RMSEA} = .009–.011, SRMR = .070 (all indices reported here and hereafter are robust estimates). We had to set the variance of some first-order factors to zero to prevent the estimation of an inadmissible solution but otherwise did not encounter any further computational problems despite the sparse data provided. Although the coverage of the variance–covariance matrix was very low, with most of the cells representing less than 5% of all cases, simulation studies suggest the effectiveness of structural equation modelling techniques under similar conditions (e.g., Willse et al., 2008).

Starting with this configural invariance model, we first established weak invariance in a tau-equivalent model by fixing all factor loadings to one. In addition, we needed to fix the variance of the distance learning slope factor to zero in the primary school group because this component was very close to zero ($\psi_{att} < .001$, $SE_\psi = .002$, $p = .90$), which would have resulted in inadmissible solutions in the following analyses. This weak invariance model did not fit the data significantly worse than the configural invariance model: ΔCFI = .004, ΔRMSEA = .000, ΔSRMR = .012.

To arrive at a strong invariance model, we first fixed the intercepts within the two groups and then also between them. With the first set of constraints, one can test whether invariance is given for each individual group and with the second whether the measurement property constrained is the same in both groups. Fixing the intercepts within groups did not significantly deteriorate the model fit compared to the weak invariance model: ΔCFI = .007, ΔRMSEA = .010, ΔSRMR = .002. Subsequently, fixing the intercepts between the two groups also did not harm the model fit: ΔCFI = .001, ΔRMSEA = .000, ΔSRMR = .001.

Finally, we fixed the error variances to be equal for the respective indicators, first within and then also between the two groups. In addition, we had to fix two error variances in the secondary school group to zero because otherwise they would have been estimated as negative. The within-groups strict invariance model did not fit the data significantly worse: ΔCFI = .002, ΔRMSEA = .000, ΔSRMR = .003. Compared to this model, the following between-group strict invariance model deteriorated in terms of ΔCFI = .011 but not in terms of ΔRMSEA = .001 and ΔSRMR = .004. We retained this model for further hypothesis testing, as model fit deterioration was only marginal and, and overall the model fit was satisfactory: $\chi^2(622) = 1317.94$ $\chi^2/df = 2.12$, CFI = .947, RMSEA = .012, 90% CI_{RMSEA} = .011–.013, SRMR = .092.

Means and variances of the learning slopes

In this strict invariance model, the mean of the learning slope for primary school pupils was estimated as $\nu_{inp} = .042$ ($SE_\nu = .007$, $p < .001$) for in-person learning and $\nu_{dis} = .018$ ($SE_\nu = .004$, $p < .01$) for distance learning. The learning progress of primary school pupils during in-person learning was more than twice as high as during school closures, and this difference was highly significant: $\Delta\chi^2(1) = 8.86$ ($p < .001$). For secondary school pupils, the in-person learning slope was $\nu_{inp} = .012$ ($SE_\nu = .005$, $p < .05$), and the distance learning slope was $\nu_{dis} = .008$ ($SE_\nu = .004$, $p = .05$). These slopes did not differ significantly from each other: $\Delta\chi^2(1) = 1.01$ ($p = .31$).

For primary school pupils, the variance of the in-person learning slope was deliberately set to $\psi_{inp} = .00$, and the variance of the distance learning slope was estimated as $\psi_{dis} = .009$ ($SE_\psi = .001$, $p < .001$), so that the two significantly differed. For secondary school pupils, the variance of the in-person learning slope was $\psi_{inp} = .002$ ($SE_\psi = .001$, $p = .27$), and the variance of the distance learning slope was $\psi_{dis} = .002$

($SE_{\psi} = .001$, $p < .05$). These two did not differ significantly: $\Delta\chi^2(1) = 0.99$ ($p = .32$). Clearly, the heterogeneity in learning progress only increased in primary school pupils and not in secondary school pupils.

Covariances of the learning slopes

Because for primary school pupils the in-person learning slope was set to zero, no covariance between this and other growth components was computed. In this group, we found a significant negative correlation between the intercept and the distance learning slope ($r = -.27$, $p < .001$). Primary school pupils generally achieving higher made slower learning progress during the school closures and vice versa.

We did not find this correlation to be significant for secondary school pupils, for whom the intercept and the distance learning slope were uncorrelated ($r = .00$, $p = .98$). By contrast, we found a strong positive correlation between the intercept and the in-person learning slope ($r = .57$, $p < .001$). The two slopes were not significantly correlated with each other ($r = .23$, $p = .75$).

DISCUSSION

Among the findings reported here, some are particularly noteworthy. First, the overall quality of our data, as indicated by the many insignificant measurement invariance tests performed and the appropriate modelling approach chosen as indicated by the excellent model fits, seems highly suitable for testing differences in learning progress between in-person and distance learning. The unique design of our study with unbiased observations before the school closures and continued observations during this event allowed us to draw strong causal conclusions about the two different types of learning. However, cautious interpretation is necessary because we were not comparing best in-person learning with best practice distance learning but rather usual in-person learning with ad hoc distance learning rapidly implemented in an emergency situation of high societal and individual uncertainty. Nevertheless, the natural experiment setup of this study in combination with large and unselected samples allows us to draw unique conclusions from the data.

We observed that the heterogeneity in learning progress significantly increased for primary school pupils during the school closures. In the 8 weeks before the school closures, learning in primary schools took place rather uniformly and with hardly observable differences between single pupils, but during the school closures, interindividual differences skyrocketed. These findings are compatible with those of parents' (Andrew et al., 2020) and teachers' survey (Cullinane & Montacute, 2020) that were conducted in the UK and found that pupils from the most affluent households were being offered active assistance (e.g., online tutoring) from their schools during the lockdown more frequently than pupils from the least affluent households. Although we have no data on the socio-economic status of pupils participating in this study, the very same social disparity might explain the growing heterogeneity in learning outcomes. At the same time, learning slowed down for this particular group. We found that primary school pupils learned more than twice as fast attending school in person compared to the distance setup. In contrast, secondary school pupils were not significantly affected in their learning pace by the school closures.

From a developmentalist perspective, the increased variance in and the decreased pace of learning progress in primary school pupils can probably be explained by cognitive, motivational and socio-emotional factors. The younger the pupils, the more they need to rely on cognitive scaffolding during instruction. In addition, their executive functioning and hence their capabilities for self-regulated learning might not yet be fully developed. Finally, younger pupils might be particularly vulnerable to the stress and strains related to the pandemic. Similar age-differential effects were reported for a completely different situation in the seminal work of Glen Elder (Elder & Caspi, 1988), who found that while older children gained in terms of autonomy and competence development during the great recession of the 1930s, younger children suffered more from the economic hardship of their families and more often became victims of marital discord or even family violence.

A final noteworthy effect is that the pace of learning during school closures could not be predicted by the pace of learning during in-person learning. The correlations between the intercept and the distance learning slope and between the two slopes were either not significant or negative. This could mean either that two completely different processes were measured before and during the school closures (which would be a validity issue) or that the same process was measured but the situations of in-person and distance learning were so different that learning progress was driven by different factors before and during school closures (which would be a substantial finding). Given the strict invariance measurement properties, the former explanation seems unlikely.

Potential long-term repercussions

Academic achievement can have cascading effects into other domains of life. While Masten et al. (2005) have demonstrated such "developmental cascades" on internalising and externalising problem behaviour, most existing studies usually focus on academic or achievement-related outcomes such as academic self-concept (Guay et al., 2003), the choice of a college major (Trautwein & Lüdtke, 2007), the choice of an

occupation (Heckhausen & Tomasik, 2002), or earnings later in life (Zax & Rees, 2002). Studies based on economic models are particularly interesting, as they provide a tangible description of the expected effect size. Taking such economic models as a foundation, Azevedo et al. (2020), for instance, estimated a life-time permanent loss in yearly earnings ranging from USD 355 to 1408 as a function of the duration of the school closures.

Although studies predicting more general, non-achievement related aspects of development by school achievement are not very common, there are good reasons to assume a positive association if one considers some basic concepts of human motivation. From a macro perspective, competence can be considered a fundamental need "to experience satisfaction in exercising and extending one's capabilities" (Levesque et al., 2004, p. 68) and its successful fulfilment is associated with intrinsic motivation, effective self-regulation, positive social development and well-being (Ryan & Deci, 2000). Following this reasoning, Tomasik et al. (2019) have demonstrated that steeper educational gains across compulsory schooling predicted successful development as indicated by notions of competence, confidence in oneself, strong character, caring for others and positive connections with others ("Five Cs Model"; see Lerner et al., 2015). Notably, these constructs in turn are predictive for contribution to society (Lerner et al., 2014).

This finding allows us to speculate about long-term repercussions for society as a whole. Besides the direct impact of the pandemic in terms of death toll or economic losses, one could almost certainly expect indirect consequences that are mediated by slower educational gains and lower academic attainment. These consequences comprise, but are not limited to, lasting effects on economic growth and tax revenues, lower job satisfaction, more prevalent health issues, higher crime rates and lower cohesion in society. Whether or not the 8 weeks of school closures are sufficient to produce measurable effects has to remain an open question here.

Theoretical implications

Silbereisen and Tomasik (2011) argue that circumstances and events that—in Bronfenbrennian terms—are located in the macro context of individual development become psychologically effective only insofar as they are able to translate into the most proximal micro contexts, such as the family or the school. Within these micro contexts, habits and routines are disturbed and require some form of adaptation. This is exactly the conceptual blueprint needed to understand why the outbreak of the COVID-19 pandemic could have had a measurable effect on pupils' learning.

Against this backdrop, at least three theoretical insights can be gained from the present analyses. First, the increased heterogeneity in the individual learning trajectories points to the existence of unobserved factors that might have moderated the impact of the macro-level transformation on individual adaptation and development. These moderating factors might either be conceptualised as "institutional filters" (see Schoon & Bynner, 2019), which prevent that events at the macro-level become manifest on the more subordinate levels. Or they might also be conceptualised at the level of individual or social resources that strengthen the resilience of children and their families (Tomasik, 2009). Finally, more or less effective coping strategies might be responsible for the large variance in developmental outcomes (Pinquart & Silbereisen, 2008). Which of these factors and in which combination were particularly relevant here needs to investigate in future research.

Second, our research demonstrated that macro-level events can impact specific outcomes (here: educational gains) also outside the specific developmental context that is most proximal to the specific outcome (here: educational institutions such as schools). We were able to present a case in point for a transmission mechanism for which already Bronfenbrenner (1979) has coined the term "meso context." Although we have no direct evidence, it is plausible to assume that features in the family context (such as the education of the parents or their occupational uncertainty during the pandemic) were at least in part responsible for the educational gains of children when the influence of schools was partly muted. Understanding and predicting such meso effects is an intriguing endeavour for future psychological and sociological work.

Third, our data provide convincing evidence that schools effectively attenuate social disparities in learning, at least in primary school pupils. During in-person schooling, we were able to observe a rather uniform learning progress, which is a finding that also seems to generalise across much longer periods of time (e.g., Helbling et al., 2019). This is not only good news for educational policy but also might help understanding how the participation in social contexts more generally shapes individual developmental trajectories.

Limitations of the study

Our study has some limitations, of which the questionable generalisability to other countries and educational settings is probably the least severe. Of course, we cannot say anything about the effects in other countries that were different in the organisation of their school system, in their cultural values, in their economic standing, or in the social and economic impact caused by the pandemic. Therefore, we refrain from speculating about any potential cross-cultural differences that might or might not be found in comparative studies, although there is some elaborated reasoning about the potentially differential impact

of the pandemic in low- and middle-income countries (Zar et al., 2020). Furthermore, data from international large-scale assessments suggest striking similarities in the factors and mechanisms associated with school achievement across countries with a quite diverse cultural and economic background (e.g., Lee, 2014). We do not want to argue for a universalist interpretation of our findings, but also do not see plausible reasons to assume that they would be completely different somewhere else (but see Guan et al., 2020, for a discussion of differential impact of the COVID-19 pandemic on career development from a cultural psychology perspective).

More problematic is the establishment of causal effects of distance learning, given the plethora of other factors that simultaneously comprised the situation during the school closures. The three probably most important to mention are the high level of strain for the families that could have undermined the teachers' educational efforts to provide good instruction, the teachers' lack of expertise and time for preparation and the lack of grading during and after the school closures that might have undermined extrinsic motivation in pupils. These and other factors threaten the internal validity of interpreting any differences as differences between in-person and distance learning. At the same time, the external validity of our study could not have been superior, and the results are informative for understanding both the short- and long-term causal effects of this specific historical event.

Another limitation of the present study is that it does not allow empirically answering the question whether the effect of reduced learning gains will translate into disadvantageous developmental trajectories within other domains of life in the short and in the long run. Not only does it not comprise developmental outcomes in these domains nor does it cover sufficiently long time spans to make substantial statements about such repercussions. We also cannot draw any conclusions about potentially aggravating or compensating factors in the other development context of these youth.

A final limitation of our study is that it did not cover any transition phase in the lives of the pupils affected by the school closures. None of our samples was entering school, transitioning between different school types, or graduating from school, although it is known that the timing of such events can make a huge difference for their effect (e.g., Schoon et al., 2002). Research conducted during German unification in the 1990s, for instance, suggests that those pupils who just graduated from school during the system change were much worse off as compared to those who have successfully entered the labour market or who had some "protected time" while still at school. We were not able to investigate these interesting effects in the present study because such transitions were not covered with the data available and because the duration of the school closures was way too short to produce meaningfully large comparison groups.

CONCLUSION

Empirical studies that could establish causal relations between the macro and the micro are extremely rare. For both practical and ethical research reasons, one has to rely on naturally occurring experimental designs and hence be in the right place at the right time to be able to study the links between societal events and individual adaptation and development. The COVID-19 pandemic provides an opportunity for such a natural experiment. With MINDSTEPS we were in the right place at the right time to provide solid evidence-based findings that can inform both developmental science and educational policy. The main message that these findings convey is that while older pupils are seemingly able to compensate for school closures in terms of a sustained learning progress, the effects are dramatically different for the younger ones. The learning gains of the younger children do not only slow down, with potential long-term repercussions for future development, but also become more heterogeneous. While some primary school pupils even seem to profit from school closures, others' school performance markedly deteriorates within a very short period of time. These children are at risk for loosing track in the academic domain and we may not leave them behind (see also Masonbrink & Hurley, 2020).

Manuscript received July 2020
Revised manuscript accepted September 2020
First published online November 2020

REFERENCES

Ahn, J., & McEachin, A. (2017). Student enrollment patterns and achievement in Ohio's online charter schools. *Educational Researcher*, *46*, 44–57.

Allen, M., Mabry, E., Mattrey, M., Burhis, J., Titsworth, S., & Burrell, N. (2006). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication*, *54*, 402–420.

Andrew, A., Cattan, S., Costa-Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). *Learning during the lockdown: Real-time data in children's experiences during home learning*. Institute for Fiscal Studies.

Aucejo, E. M., & Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55*, 70–87.

Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A., & Geven, K. (2020). *Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates (policy research working papers no. 9284)*. World Bank.

Basilaia, G., & Kvavadze, D. (2020). Transitions to online education in schools during a SARS-Cov-2 coronavirus (COVID-19) pandemic in Georgia. *Pedagogical Research*, *5*, em0060.

Berger, S., Verschoor, A. J., Eggen, T. J. H. M., & Moser, U. (2019). Development and validation of a vertical scale for formative assessment in mathematics. *Frontiers in Education*, *4*, 103.

Bronfenbrenner, U. (1979). *The ecology of human development*. Harvard University Press.

Burgess, S., & Sievertsen, H. H. (2020). *Schools, skills, and learning: The impact of COVID-19 on education*. VoxEU. https://voxeu.org/article/impact-covid-19-education

Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D.-O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, *97*, 533–547.

Chamberlain, L., Lacina, J., Bintz, W. P., Jimerson, J., Payne, K., & Zingale, R. (2020). Literacy in lockdown: Learning and teaching in COVID-19 school closures. *The Reading Teacher*.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*, 227–268.

Cullinane, C., & Montacute, R. (2020). *School shutdown (COVID-19 and social mobility brief no. 1)*. Sutton Trust.

Education Endowment Foundation. (2020). *Impact of school closures on the attainment gap: Rapid evidence assessment*. EEF.

Elder, G. H., Jr., & Caspi, A. (1988). Economic stress in lives: Developmental perspectives. *Journal of Social Issues*, *44*, 25–45.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, *4*, 22–46.

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are student absences worth the worry in US primary schools? *Education Finance and Policy*, *12*, 137–165.

Goodman, J. (2014). *Flaking out: Student absences and snow days as disruptions of instructional time (NBER working paper no. 20221)*. National Bureau of Economic Research.

Guan, Y., Deng, H., & Zhou, X. (2020). Understanding the impact of the COVID-19 pandemic on career development: Insights from cultural psychology. *Journal of Vocational Behavior*, *119*, 103438.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic performance: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*, 124–136.

Hansen, B. (2011). *School year length and student performance: Quasi-experimental evidence (SSRN Working Paper No. 2269846)*. New York: Social Science Research Network. Available at http://dx.doi.org/10.2139/ssrn.2269846

Heckhausen, J., & Tomasik, M. J. (2002). Get an apprenticeship before school is out: How German adolescents adjust vocational aspirations when getting close to a developmental deadline. *Journal of Vocational Behavior*, *60*, 199–219.

Helbling, L. A., Tomasik, M. J., & Moser, U. (2019). Long-term trajectories of school performance in the context of social disparities: Longitudinal findings from Switzerland. *Journal of Educational Psychology*, *111*, 1284–1299.

Iivari, N., Sharma, S., & Ventä-Olkkonen, L. (2020). Digital transformation of everyday life: How COVID-19 pandemic transformed the basic education of the young generation and why information management should care? *International Journal of Information Management*, *55*, 102–183.

Isiordia, M., & Ferrer, E. (2016). Curve of factors model: A latent growth modelling approach for educational research. *Educational and Psychological Measurement*, *78*, 203–231.

König, N., Tomasik, M. J., Berger, S., Giesinger, L., Helbling, L. A., & Moser, U. (2020, March). *An adaptive computer-based tool for formative student assessment*. Practitioner demonstration at the 2020 Learning Analytics and Knowledge Conference, Frankfurt am Main https://video01.uni-frankfurt.de/Mediasite/Showcase/lak2020/Presentation/efea148f6100411ea59edaf6b13df5a81d

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). *Projecting the potential impacts of COVID-19 school closures on academic achievement (EdWorkingPaper no. 20–226)*. Annenberg Institute.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *Economic Journal*, *125*, F397–F424.

Lee, J. (2014). Universal factors of student achievement in high-performing eastern and Western countries. *Journal of Educational Psychology*, *106*, 364–374.

Lee, J. (2020). Mental health effects of school closures during COVID-19. *The Lancet: Child and Adolescent Health*, *4*, P421.

Lerner, R. M., Lerner, J. V., Bowers, E. P., & Geldhof, G. J. (2015). Positive youth development and relational-developmental-systems. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science: Theory and method* (Vol. 1, 7th ed., pp. 1–45). Wiley.

Lerner, R. M., Wang, J., Chase, P. A., Gutierrez, A. S., Harris, E. M., Rubin, R. O., & Yalin, C. (2014). Using relational developmental systems theory to link program goals, activities, and outcomes: The sample case of the 4-H study of positive youth development. *New Directions for Youth Development*, *2014*(144), 17–28.

Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Educational Psychology*, *96*, 68–84.

Lieberman, M. (2020). *Taking attendance during coronavirus closures: Is it even worth it?* Retrieved from https://www.edweek.org/ew/articles/2020/04/17/taking-attendance-is-trickyduring-coronavirus-closures.html

Liu, J., Lee, M., & Gershenson, S. (2020). *The short- and long-run impacts of secondary school absences (EdWorkingPaper no. 20–125)*. Annenberg Institute.

Lucas, M., Nelson, J., & Sims, D. (2020). *Schools' responses to COVID-19: Pupil engagement in remote learning*. National Foundation for Educational Research.

Masonbrink, A. R., & Hurley, E. (2020). Advocating for children during the COVID-19 school closures. *Pediatrics*, *146*, e20201440.

Masten, A. S., Roismann, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., Boelke-Stennes, K., & Tellegen,

A. (2005). Developmental cascades: Linking academic performance and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, *41*, 733–746.

Moore, S. A., Faulkner, G., Rhodes, R. E., Brussoni, M., Chulak-Bozzer, T., Ferguson, L. J., Mitra, R., O'Reilly, N., Spence, J. C., Vanderloo, L. M., & Tremblay, M. S. (2020). Impact of the COVID-19 virus outbreak on movement and play behaviours of Canadian children and youth: A national survey. *International Journal of Behavioral Nutrition and Physical Activity*, *17*, 85.

Pinquart, M., & Silbereisen, R. K. (2008). Coping with increased uncertainty in the field of work and family life. *International Journal of Stress Management*, *15*, 209–221.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.

Sacerdote, B. (2012). When the saints go marching out: Long-term outcomes for student evacuees from hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, *4*, 109–135.

Schoon, I., & Bynner, J. (2019). Young people and the great recession: Variations in the school-to-work transition in Europe and the United States. *Longitudinal and Life Course Studies*, *10*, 153–173.

Schoon, I., Bynner, J., Joshi, H., Parsons, S., Wiggins, R. D., & Sacker, A. (2002). The influence of context, timing, and duration of risk experiences for the passage from childhood to midadulthood. *Child Development*, *73*, 1486–1504.

Serafini, G., Parmigiani, B., Amerio, A., Aguglia, A., Sher, L., & Amore, M. (2020). The psychological impact of COVID-19 on the mental health in the general population. *QJM: An International Journal of Medicine*, *113*, 531–537.

Silbereisen, R. K., & Tomasik, M. J. (2011). Psychosocial functioning in the context of social, economic, and political change. In X. Chen & K. H. Rubin (Eds.), *Socioemotional development in cultural context* (pp. 305–331). Guilford Press.

Tomasik, M. J., Berger, S., & Moser, U. (2018). On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues. *Frontiers in Psychology: Educational Psychology*, *9*, 2245.

Tomasik, M. J., Napolitano, C. M., & Moser, U. (2019). Trajectories of academic performance across compulsory schooling and thriving in young aduthood. *Child Development*, *90*, e745–e762.

Tomasik, M. J., & Silbereisen, R. K. (2009). Demands of social change as a function of the political context, institutional filters, and psychosocial resources. *Social Indicators Research*, *94*, 13–28.

Tomasik, M. J., & Silbereisen, R. K. (2016). Demands of social change across multiple domains of life and across time at the advent of the global financial crisis. *Research in Human Development*, *13*, 312–327.

Trautwein, U., & Lüdtke, O. (2007). Epistemological beliefs, school achievement, and college major: A large-scale longitudinal study on the impact of certainty beliefs. *Contemporary Educational Psychology*, *32*, 348–366.

United Nations Sustainable Development Group. (2020). *Policy brief: Education during COVID-19 and beyond*. United Nations.

van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and cild poverty: A social crisis. *The Lancet: Public Health*, *5*, E242–E244.

von Hippel, P. T., & Hamrock, C. (2019). Do test score gaps grow before, during, or between school years? Measurement artifacts and what we can know in spite of them? *Sociological Science*, *6*, 43–80.

Willse, J. T., Goodman, J. T., Allen, N., & Klaric, J. (2008). Using structural equation modelling to examine group differences in assessment booklet designs with sparse data. *Applied Measurement in Education*, *21*, 253–272.

Xiao, J., & Bulut, O. (2020). Evaluating the performance of missing data handling methods in ability estimation from sparse data. *Educational and Psychological Measurement*, *80*, 932–954.

Zar, H. J., Dawa, J., Fischer, G. B., & Castro-Rodriguez, J. A. (2020). Challenges of COVID-19 in children in low- and middle-income countries. *Paediatric Respiratory Reviews*, *35*, 70–74.

Zax, J. S., & Rees, D. I. (2002). IQ, academic performance, environment, and earnings. *Review of Economics and Statistics*, *84*, 600–616.