Impacts of the COVID-19 pandemic on school enrolments

6 October 2021[1]

SUMMARY

Evidence drawing from household data which is difficult to interpret points to almost half a million learners not being in school in 2021 because of the pandemic. The analysis that follows examines whether patterns seen in the official enrolment data of the schooling sector, including public and independent ordinary and special schools, confirm this. It is of course possible for a learner to be enrolled at the start of 2021, but then not to attend at all. This analysis is not aimed at investigating this. Yet enrolments on their own can provide an important indication of the degree to which the population of children may have disengaged from schooling as a result of the pandemic.

One possible problem is that the pandemic has made households reluctant to take children to school as first-time enrolled learners, in either Grade R or Grade 1. By comparing quarter 1 enrolment data from 2020 and 2021, the analysis finds evidence of this, with enrolment in 2021 among children aged 4.5 to 6.0 at the start of the year being around 27,000, or 2.3%, lower than what could be expected. Proportionally, this problem is greatest in Northern Cape and KwaZulu-Natal. (See section 7.)

A second possible problem is greater dropping out among learners who by law are subject to compulsory schooling. The data suggest the pandemic caused up to 19,000 additional drop-outs among such learners between 2020 and 2021. This 19,000 excludes deaths as estimated by Stats SA, and is concentrated among learners aged 8 to 12 in 2020. The 19,000 figure comes to 0.4% of learners of compulsory school-going age. (See section 6.)

A third possible problem is that dropping out worsened for learners aged 15 and above. This does not appear to have been a problem. Grade 12 enrolment increased by 20%, with age-specific increases in Grade 12 for ages 17 to 25 exceeding what Stats SA estimates to be population growth at these ages. For age 17 and above, enrolment numbers across all grades remained almost unchanged, while below age 17 they increased. It is possible that participation for age 17 and above *in grades below Grade 12* is lower in 2021 than it should be, but this depends in part on how one interprets Stats SA's population estimates. The worst indicator values are those for Northern Cape, but even in this province enrolment increases for ages 15 to 19 exceeded population growth slightly. (See section 8.)

The enrolment trends do not suggest that half a million learners have disengaged from the schooling system as a result of the pandemic. What is found is that around 27,000 young children have not enrolled as first-time learners in the way they should, while up to 19,000 learners at the compulsory school-going age dropped out. Without more historical analysis, it is difficult to conclude definitively that the 19,000 dropped out because of the effects of the pandemic. However, the historical analysis that exists makes it plausible that most of the 19,000 dropped out due to the pandemic. Whatever the cause, dropping out at the compulsory school-going age is especially concerning. What the enrolment data do not suggest is that the widely debated problem of dropping out above the compulsory ages has increased. In fact, the enrolment data indicate that the pandemic led to increased enrolment at that level, especially in Grade 12.

[1] Analysis by Martin Gustafsson (mgustafsson@sun.ac.za) for the Department of Basic Education.

1 Introduction

Enrolment and attendance in schools have been affected by the COVID-19 pandemic. The World Bank estimated in June 2020 that the pandemic would result in a global reduction in enrolment of around 0.5%, or seven million learners[2]. Given how the pandemic has unfolded since then, those figures can be considered an under-estimate.

South African data from the NIDS-CRAM household surveys, conducted telephonically during the pandemic, suggest that the number of learners returning to school in 2021 is around 400,000 to 500,000 lower than it would be without the pandemic[3]. However, as the analysts explain, the data are difficult to interpret given that responses are for the household as a whole, not for individual children, and given the context where schools were practising rotational attendance, and not daily attendance. The latter meant the question used, 'Are there any learners in your household who have not yet returned to school this year?', could be ambiguous to the respondent. Moreover, the fact that most households with children not returning to school in 2021 had *other* children in the same household who had returned, was difficult to interpret.

This report examines enrolment records from the start of the 2021 school year and compares these to earlier 2020 data in order to investigate the non-return problem, but also other issues. *Enrolment*, meaning official registration as a learner in a school, is not the same as *attendance*, yet enrolment, for which good data are available, is important to monitor. This is especially so in the context of the pandemic.

Work has begun on processing and analysing daily *attendance* data from schools, which exist at the national level, though the reliability of this varies across provinces. Attendance data from administrative systems would be the optimal data source to establish how daily attendance has been affected by the pandemic, and how various models of partial attendance to advance social distancing have been adopted across the system.

2 The data used

The data used are derived from schools-based management systems, which outside Western Cape is mostly SA-SAMS[4], designed and maintained by the Department of Basic Education (DBE). This means schools input the data as part of their regular operational processes. Investments in SA-SAMs, and plans for future enhancements to this this system, explain in part why it is now widely used. The 2017 School Monitoring Survey indicated that outside the Western Cape, usage of SA-SAMS by school principals has been high: at least 97% in six provinces, with a somewhat lower 94% in KwaZulu-Natal and 86% in Gauteng[5]. Western Cape has historically used its own system.

What is important about the functionality of SA-SAMS for the current report is how enrolments are captured at the start of each year. While the system can re-use learner information from the previous year, classifying a learner as being enrolled in the current year requires steps that prevent a simple roll-over from the previous year, which would raise the risk of learners being considered enrolled in the current year when they have not been present. This needs to be avoided for various reasons, including the fact that enrolments drive school funding and personnel provisioning. A key step is the assignment of each learner to a class in the new year. Each class has a code: for instance, in a school with three Grade 4 classes, these classes could be coded 4A, 4B and 4C. Even small schools with one class in a grade must

[2] Azevedo *et al*, 2020: 18.
[3] Shepherd and Mohohlwane, 2021.
[4] South African School Administration and Management System.
[5] Department of Basic Education, 2019b.

code that class. Both the 2020 and 2021 data used for the current report had the class code present for every learner. Further steps to prevent erroneous or fraudulent counting of learners as enrolled when they have left the school exist. These include functions linked to the monitoring of attendance within SA-SAMS.

For the two years, the data of quarter 1 of the school year was used, for public and independent ordinary and special schools. The 2020 data had been used for an earlier report examining the effects of the pandemic on the submission of assessment results[6], and were obtained in January 2021. The 2021 data were obtained in September 2021 specifically for the current report.

The 2021 data were 'raw' in the sense that they had recently been received from provinces, and had not been subjected to the full range of quality assurance steps in the DBE. However, the data were examined carefully in preparing the current report, and were found to be of a sufficient quality. Learner and school counts per province for the two datasets appear in Table 1 below.

Table 1: Learners and schools in the two datasets

	Learners		Schools	
	2020	2021	2020	2021
EC	1,853,270	1,858,043	5,469	5,371
FS	726,335	733,138	1,147	1,090
GP	2,566,165	2,626,425	3,076	3,115
KN	2,886,354	2,915,326	6,094	6,107
LP	1,768,019	1,807,735	3,914	3,896
MP	1,112,890	1,139,434	1,801	1,802
NC	306,183	306,361	596	596
NW	870,904	880,403	1,564	1,571
WC	1265219	1,287,394	1,819	1,835
SA	13,355,339	13,554,259	25,480	25,383

Forms to effect compliance with the Protection of Personal Information Act were signed by the producer of the current report.

3 Risk of duplicate learners

There are two risks that should be addressed as far as possible in making the comparison across the two years. The one risk is that there are duplicate learners in the data in each of the two years. The other risk, dealt with in section 4, is that learners are missing in the data.

One thing not found in the data is unrealistically large schools. In the 2021 data, there are 121 schools with 2,000 or more learners, the largest having 3,095 learners. The patterns were similar in the 2020 data. There is thus no 'bug' in the data in the form of a school with, say, 50,000 learners.

Based on figures presented in Table 2 below, in 2020 7.5% of learners had no unique 13-digit national ID number, either because the field was blank[7] or because the ID number was duplicated. The figure is a slightly higher 7.8% in 2021. In both years 0.9% of learners had a 13-digit ID number they shared with someone else. Over 99% of those sharing the ID number did so with just one other learner.

[6] 'Using EMIS assessment and class size data to inform the COVID-19 response', dated 31 May 2021.
[7] No data at all, or a series of one or more spaces and nothing else, were considered blank.

Table 2: Extent of national ID problems

Missing	Not unique	Has non-numeric text	Length is not 13	Does not agree with birth date	2020 learners	2021 learners
0	0	0	0	0	12,347,788	12,496,066
1	0	0	0	0	876,211	927,065
0	1	0	0	0	123,748	122,819
0	0	0	0	1	3,307	3,526
0	0	1	1	0	2,707	3,182
0	0	0	1	0	854	1,052
0	0	1	0	0	393	455
0	1	0	1	0	223	4
0	1	0	0	1	69	57
0	1	1	1	0	25	21
0	0	1	0	1	11	10
0	1	1	0	0	3	2
					13,355,339	13,554,259

A unique national ID number is not a guarantee that a record is not a duplicate. Analysis of learners with unique ID numbers revealed that it was far more likely to find someone else with exactly the same date of birth and surname and first name[8] in the same school as in the rest of the system. In fact, in the 2020 data it was around 1,600 times more likely to find someone else with the same birthday and names in the same school than would have been the case had such similar learners been distributed randomly across the system. This pattern is also found in the 2021 data. The explanation for most of this phenomenon would have to be that schools are entering learners more than once, possibly once with the correct ID number and a second time with the incorrect ID number.

A few important statistics were calculated. These were different versions of the probability of finding someone else in the data with the same birthday, surname and first name. For this analysis, learners with a missing birthday or either name missing were excluded. The details for 2020 are shown in Table 3. A normal level of birthday-and-names duplication is represented by the national 0.38% statistic. This is the percentage of learners in the data whose details were duplicated *in one or more other schools* (and possibly also in the same school, but duplication in another school had to exist for the learner to be counted). This 0.38% statistic is based on an analysis that uses fairly clean data, namely records from 2020 where the national ID number is unique. Provincial statistics shown in the second column of Table 3 are based on a search for duplicates across the entire country, not just the province. Clearly, there are regional factors. In KwaZulu-Natal, even with good data duplication is high, probably because IsiZulu names are common across the country. If one widens the criterion to include duplicates in the same school, the 0.38% statistic rises slightly to 0.39%. Without the concentration of duplicates within the same school discussed above, this statistic would essentially be the original 0.38%.

[8] First name and surname were condensed separately to facilitate processing. Spaces were removed, and then just the first 15 characters were considered. After this truncation, in the 2020 data only 0.1% of surnames and 2.1% of first names had 15 letters. This process was above all important to eliminate differences caused by typing errors such as that resulting in 'DE BEER' and 'DEBEER' for the same person. After the condensation, the latter would be seen.

Table 3: Probability of a birthday-names duplicate 2020

	Only other schools			Any school, including same		
	Original all	Original only unique IDs	After removal of duplicates	Original all	Original only unique IDs	After removal of duplicates
EC	1.30	0.04	0.11	1.35	0.05	0.11
FS	0.69	0.32	0.36	0.73	0.32	0.36
GP	1.26	0.34	0.45	1.35	0.34	0.45
KN	1.40	1.00	1.06	1.47	1.01	1.06
LP	0.63	0.15	0.19	0.66	0.15	0.19
MP	0.99	0.50	0.56	1.03	0.50	0.56
NC	0.56	0.04	0.06	0.57	0.04	0.06
NW	0.65	0.11	0.13	0.69	0.11	0.14
WC	1.53	0.09	0.14	1.63	0.11	0.16
SA	1.13	0.38	0.44	1.19	0.39	0.45

Note: Values are percentages.

Columns headed 'Original all' in Table 3 use all the 2020 data described in Table 1 above. Here the level of duplication is much higher, because of the presence of obviously duplicate records. Table 4 indicates that the situation is very similar in the 2021 data. What is striking is how very similar the figures in the second column are across Table 3 and Table 4. This suggests that the data used for these columns are relatively clean, and that the statistics in these columns can serve as benchmarks of what level of duplication to expect in the South African context.

Table 4: Probability of a birthday-names duplicate 2021

	Only other schools			Any school, including same		
	Original all	Original only unique IDs	After removal of duplicates	Original all	Original only unique IDs	After removal of duplicates
EC	1.10	0.03	0.15	1.12	0.03	0.06
FS	0.74	0.32	0.46	0.79	0.32	0.38
GP	1.09	0.33	0.55	1.13	0.34	0.40
KN	1.34	1.00	1.18	1.39	1.00	1.05
LP	0.69	0.15	0.30	0.73	0.16	0.19
MP	1.12	0.50	0.72	1.18	0.50	0.57
NC	0.67	0.04	0.12	0.70	0.05	0.07
NW	0.75	0.11	0.26	0.80	0.11	0.15
WC	1.60	0.11	0.42	1.81	0.14	0.19
SA	1.09	0.38	0.43	1.15	0.39	0.43

Columns headed 'After removal of duplicates' in the previous two tables use data after a systematic removal of duplicates, a process described below[9].

Inspection of the data revealed that among the learner records with an existing ID number duplicated across records, around a third were a case of the combination of the surname and first name being different, while around two-thirds were a case of the surname and first name being shared, nearly always across two, not more, records with the same ID. This situation was very similar in 2020 and 2021. Nearly all cases of shared names involved a record in one province, and a second record in another province. This kind of duplication is spread roughly in proportion to enrolments across provinces, though Eastern Cape and Western Cape have exceptionally high levels of this type of duplication.

[9] The reason why, say, the third column of Table 3 displays a higher national statistic than the second column – 0.44 against 0.38 – is that there are more records in the data used for the former. The more learner records there are, the greater the probability that someone else will share the same birthday, first name, and surname.

It should be noted that duplication of learners across provinces is common in many reported enrolment statistics, including statistics submitted to National Treasury to determine equitable share formula weights. Where there are good reasons to believe that two provinces can legitimately claim to have the same learner, because the learner has been enrolled in two separate provinces while migration from one to the other is under way, allowing the learner to be counted in each for funding reasons is tolerated. This is in part because of the difficulty of establishing the direction in which the learner is moving before the deadline for submitting enrolment figures to Treasury.

A first round of removal of duplicates resulted in around 35,000 records being removed from each year's data. This process focussed just on records with duplicate 13-digit ID numbers. Where the ID number and the two name fields were duplicated, it was assumed that just one record should remain. In removing records, a provincial rank field was introduced. This carried assumptions around what province a learner was actually in where duplication across provinces occurred. The ranking was as follows: GP, WC, KN, FS, NW, MP, NC, LP, EC. This meant, for instance, that if a learner had a record for both GP and NW, the GP record was retained. The provincial ranking is roughly in line with what is understood to be the inter-provincial migration patterns. Apart from this provincial ranking approach, other fields such as grade were used to decide which records to remove in the very few cases where duplication occurred within the same province, meaning the removal was never a random process.

Following this process, duplicate ID numbers still remained in the data, specifically where names were *not* also duplicated.

The steps just described were not applied where names were missing. No surnames were missing[10], but nationally 2.0% of first name fields were blank in both years. This was worst in Eastern Cape and Western Cape – around 3.0% missing – and best in North West – 0.5% missing.

A second round of removal of duplicates focussed on records where the ID number was missing. Here three fields were considered to approximate unique identifiers of learners when combined: date of birth, surname, and first name. In the 2020 data, 97% of records where the ID number was missing had a combination of names and date of birth which was unique across the entire 2020 dataset. These records could virtually all be considered to represent unique learners. The question was what to do with the remaining 3%, around 23,000 records in the case of the 2020 data.

Eyeballing of the data, and considering the analysis discussed above, led to the decision to remove two categories of suspected duplicates from the records where the ID number was missing and the combination of names and date of birth was not unique across the entire dataset. Firstly, learners with the same non-missing birthday[11], surname and first name in the same school were considered duplicated, and just one record was left after the removal of duplicates. Secondly, where birthday, surname, first name and also grade were duplicated in the dataset as a whole, records with a blank ID number were considered duplicates, and were removed[12]. This process resulted in just over 10,000 records being removed from each of the two years' datasets. The 2021 duplication problems were very similar to those of 2020.

[10] To be precise, one was missing in the 2021 data.
[11] In the 2020 data, 149 records had no date of birth, and a further 1,196 had a birthday which produced an age outside the range of 0 to 30. In the 2021 data, 51 values were missing, and a further 1,315 records had birthdays producing out-of-range ages.
[12] If the duplication occurred only among records with blank ID numbers, one remained after the removal.

Table 5 reflects the total enrolment in each year found after the two rounds of removing duplicates. Nationally, 0.4% and 0.3% of the original learner records were dropped for 2020 and 2021 respectively.

Table 5: Learners after removal of duplicates

	2020		2021	
	Learners	% of original	Learners	% of original
EC	1,851,189	99.9	1,857,378	100.0
FS	725,193	99.8	731,650	99.8
GP	2,547,202	99.3	2,609,686	99.4
KN	2,880,447	99.8	2,910,145	99.8
LP	1,766,920	99.9	1,806,453	99.9
MP	1,110,871	99.8	1,136,651	99.8
NC	306,019	99.9	306,109	99.9
NW	868,853	99.8	877,831	99.7
WC	1,249,856	98.8	1,271,321	98.8
SA	13,306,550	99.6	13,507,224	99.7

The removal of duplicates made no difference to the 2020 school count, but did result in the disappearance of one very small Limpopo school in the 2021 data.

4 Risk of missing learners and schools

If whole schools which exist 'on the ground' are missing in either of the two datasets, comparisons across the two years will be compromised. Table 6 indicates that the extent of unmatched schools is low: for instance, schools in the 2021 data not found in the 2020 data account for just 0.4% of learners in 2021. Table 6 uses the original data before the removal of duplicates. The extent of unlinked schools seems on the whole compatible with year-on-year closing of certain schools and the opening of other schools.

Table 6: Schools not linked across 2020 and 2021

	2020		2021	
	% of 2020 schools not linked	% of 2020 learners in these schools	% of 2021 schools not linked	% of 2021 learners in these schools
EC	2.0	0.3	0.2	0.1
FS	5.4	0.3	0.2	0.3
GP	1.4	0.4	2.7	0.7
KN	0.5	0.3	0.8	0.4
LP	0.9	0.1	0.5	0.2
MP	1.1	0.2	1.1	0.7
NC	0.3	0.0	0.3	0.3
NW	0.5	0.1	1.0	0.4
WC	0.2	0.0	0.9	0.2
SA	1.2	0.2	0.9	0.4

5 Individual learners found in 2020 but not 2021

Though it may be ideal to gauge possible declines in enrolment in 2021 by linking individual learners across 2020 and 2021, this is not practical with the available data because of problems with national ID numbers, in particular missing IDs, as discussed above. The extent of the problem can be assessed by focussing on learners aged 7 to 13 on 1 January 2021, or 6 to 12 on 1 January 2020. In terms of the law, these learners are too old to be moving into the schooling system for the first time, and too young to be dropping out. Household data make it clear that there has been virtually no dropping out for learners in this age bracket, at least not in the years preceding the pandemic – see Department of Basic Education (2019a: 6).

Here and in the remainder of the report, three different versions of the data will be analysed: (1) the original data reflected in Table 1; (2) the data after the removal of duplicates reflected in Table 5; (3) as for (2), but limited to schools found in the datasets of both years. Most of the analysis will use (2), but the other options will be explored too in order to establish whether taking certain data issues into account impacts substantively on the findings.

Table 7 draws from the data after the removal of duplicates. Especially in Gauteng, linking of learners one would expect to be almost all present in both years is not practical, largely because of missing ID problems.

Table 7: Learners not linked to the other year using national ID

	% of 2020 learners aged 6 to 12 on 1 January 2020	% of 2021 learners aged 7 to 13 on 1 January 2021
EC	6.9	7.0
FS	6.9	6.5
GP	18.4	18.7
KN	8.0	8.1
LP	5.6	5.7
MP	8.5	9.0
NC	5.1	3.5
NW	6.9	6.7
WC	9.4	9.4
SA	9.6	9.7

Creating pseudo-IDs is not an easy option. Such IDs were created using a combination of national ID number (which could be blank), date of birth, surname, first name, gender and race. These new IDs lowered the 2020 national value of 9.6% seen in Table 7 to 8.1%. The gain was thus small. Further work could improve the degree of linking. However, this route was not pursued for the current report. As indicated below, alternative strategies, while not ideal, can shed considerable light on the impacts of the pandemic on enrolment patterns.

6 Comparing birth cohorts to detect compulsory-age dropping out

Figure 1 provides a picture of the trends for learners aged 6 to 12 on 1 January 2020, using the three different versions of the data discussed above. Note 'birth cohort' here describes *learners* born in a particular year, not children in the population. One would expect the red 2021 curves to lie very close to the corresponding green 2020 curves. Even without dropping out, the match would not be perfect as learners leave and enter the general population due to migration and mortality. The fact that learners aged 6 on 1 January 2020 were more likely to be in school in 2021 than in 2020 can be attributed to children entering school later than they should. What is concerning about Figure 1 is the drop in enrolment on the right-hand side of the graph, where children are older.

Figure 1: Trends for compulsory-aged birth cohorts

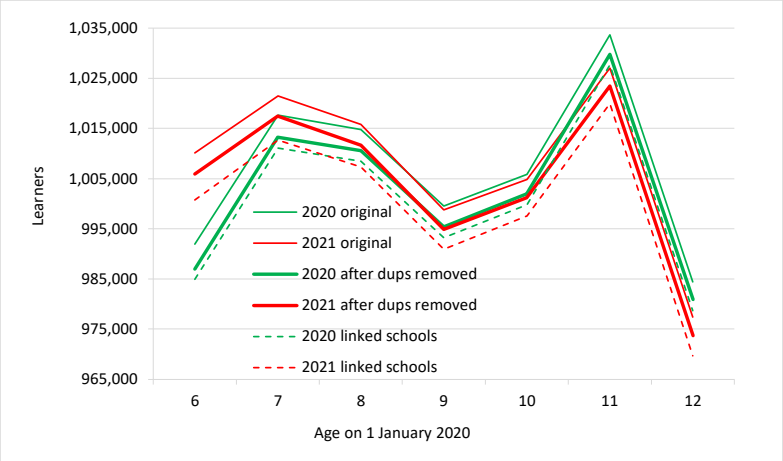


Table 8 provides figures for the gap between the red and green in Figure 1, considering only those ages where the 2021 value is lower than the 2020 value. Percentages in Table 8 use as the denominator only the 2020 ages where there was a decline. The declines were seen for ages 9 to 12 using the first two versions of the the data, and for ages 8 to 12 using the linked schools data.

Table 8: Enrolment declines at the compulsory ages

Version of the data	Decline ages 8 to 12 (in 2020)	% of relevant learner birth cohorts 2020
Original data	15,441	0.4
After duplicates removed	14,729	0.4
Linked schools	22,331	0.4

Is a decline of 0.4% problematic, given migration and mortality? There are apparently no age-specific migration data to assist here, but international migration during the pandemic would have been very low. With regard to mortality, Stats SA points to around 3,000 children aged 8 to 12 dying each year in recent years[13]. This would barely have been affected by the pandemic. During the first 100 days of the pandemic, around four children aged 8 to 12 died of COVID, in a context of 3,088 COVID deaths in total[14]. The Stats SA report thus suggests the figures from Table 8 minus 3,000 would provide an estimate of a loss of learners across the seven age cohorts of Figure 1 who did not return in 2021 as they should. Whether this loss is worse than that of previous years could be assessed by running a similar comparison of 2019 to 2020 enrolment data. However, the required 2019 data were not readily available.

To conclude, up to around 19,000 learners across seven age cohorts who should be in the system in 2020 and 2021 were not enrolled in 2021. This remains 0.4% of the learner population of 2020 aged 8 to 12. A key factor could be the pandemic, specifically parent fears around sending children to school, and worsening poverty in the household which may have affected the ability of parents to cover transport and school uniform costs.

7 Enrolment levels at the start of the schooling cycle

Figure 2 below illustrates trends with respect to *age* cohorts, meaning age in one year is compared to the same age the next year. The aim here is to assess whether fewer learners entered the schooling system for the first time as a result of the economic and social shock of

[13] Statistics South Africa, 2017: 8.
[14] Pillay-Van Wyk *et al* (2020).

the pandemic. For Figure 2, the version of the data after the removal of duplicates was used. The top red curve is below the top green curve in several places on the left-hand side of the graph, indicating that enrolment levels declined somewhat between the two years. (The peaks at 6.3 and 5.3 years in the graph reflect the peak in births in September seen in South Africa.)

Figure 2: Grades R and 1 trends

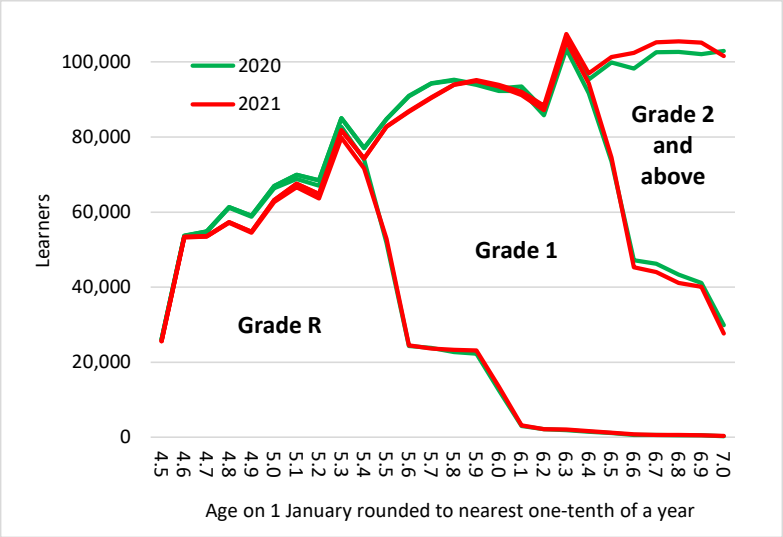


Table 9 quantifies the decline for learners aged 6.0 and below. It is clear that below age 6 on 1 January, gaps between 2020 and 2021 appear more concerning. The decline figures in Table 9 represent total enrolment age 6.0 and below in 2020 minus total enrolment age 6.0 and below in 2021.

Table 9: Enrolment decline age 6.0 and below

Version of the data	Decline	Decline as a % of 2020 enrolment age 6 and below
Original data	35,097	3.0
After duplicates removed	34,873	3.0
Linked schools	38,236	3.3

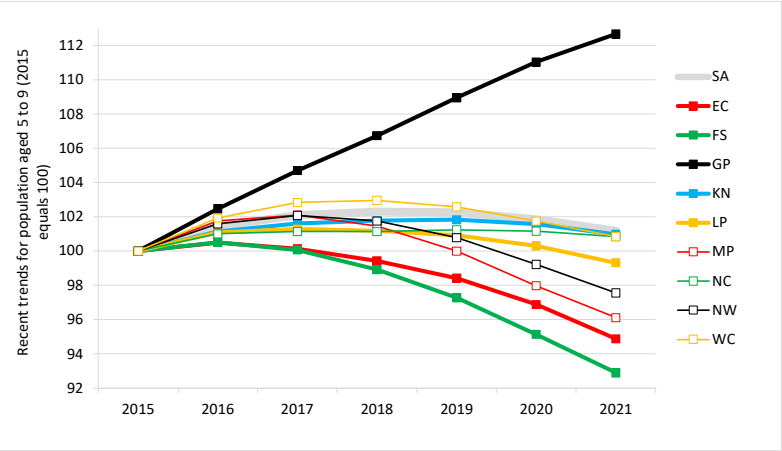
A breakdown of the Table 9 statistics by province is provided in Table 10. Here inter-provincial migration plays a role, as clearly reflected in the increases seen in Gauteng and Western Cape.

Table 10: Enrolment decline age 6.0 and below by province

	Original data		After duplicates removed		Linked schools	
	Decline	%	Decline	%	Decline	%
EC	8,416	5.0	8,184	4.9	7,810	4.7
FS	2,949	5.3	2,951	5.3	3,130	5.7
GP	-3,424	-1.7	-3,656	-1.9	-2,392	-1.2
KN	12,828	4.5	12,738	4.5	13,749	4.8
LP	6,589	3.4	6,615	3.4	6,907	3.6
MP	3,622	3.6	3,718	3.7	4,162	4.1
NC	1,906	8.1	1,911	8.1	2,035	8.6
NW	2,773	3.7	2,871	3.8	3,185	4.2
WC	-562	-0.7	-459	-0.6	-350	-0.5
SA	35,097	3.0	34,873	3.0	38,236	3.3

National and provincial population trends for children need to be considered. Stats SA figures released with the 2021 mid-year population estimates (MYPE) point to an ongoing and accelerating decline in the national population aged 5 to 9 in recent years, the decline being 0.7% between 2020 and 2021. Provincial trends vary greatly. The following graph uses a background Excel file published online by Stats SA, together with the 2021 MYPE report[15].

Figure 3: Population trends for ages 5 to 9



Source: Stats SA mid-year population estimates for 2021.

Table 11 takes Table 10 values and adjusts them, using population trends. Nationally, enrolment of children aged 6.0 and below was 27,000, or 2.3%, lower at the start of 2021 than one might expect. Declines in first-time enrolment emerge as particularly problematic in Northern Cape, with 2021 enrolments for children aged 6.0 and below on 1 January being 7.8% lower than what one might expect. Less serious, yet concerning declines are seen in other provinces too, particularly KwaZulu-Natal.

Table 11: Declines age 6.0 and below with demographic adjustments

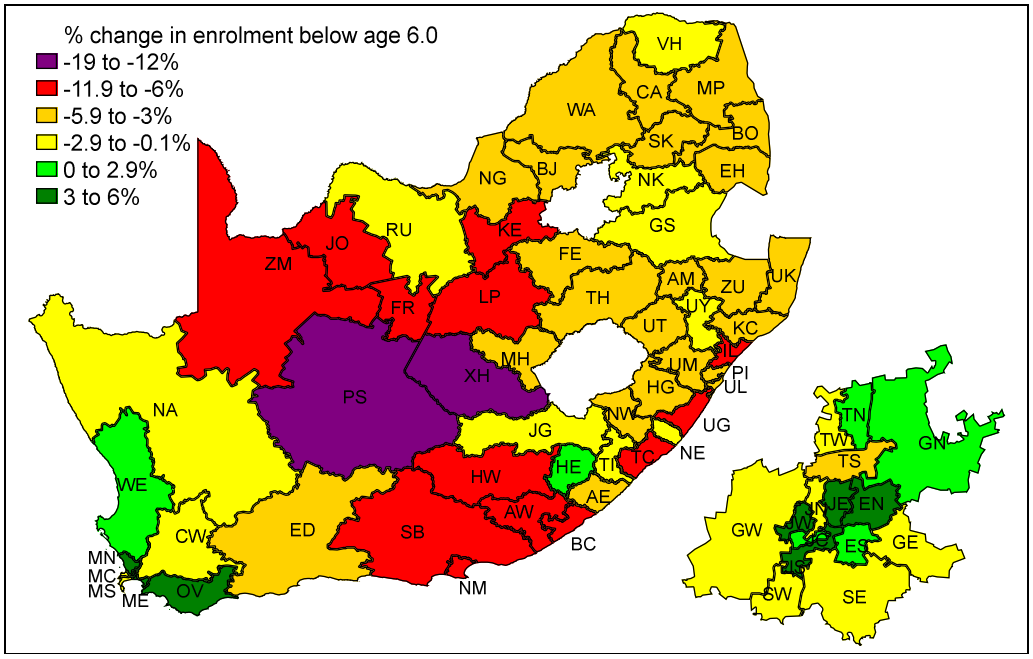
	2020 to 2021 % change in pop. ages 5 to 9	'After duplicates' % from Table 10 with adjustment (positive is decline)	New 'After adjustments' estimate of decline
EC	-2.1	2.8	4,703
FS	-2.4	3.0	1,634
GP	1.5	-0.4	-805
KN	-0.6	3.9	11,076
LP	-1.0	2.4	4,694
MP	-1.9	1.8	1,778
NC	-0.3	7.8	1,834
NW	-1.7	2.1	1,589
WC	-0.9	-1.5	-1,139
SA	-0.7	2.3	26,956

Figure 4, which draws only from data where the same schools had learners below 1 January age 6.0 in both years, provides a picture of the distribution of the problem of low first-time enrolment by education district. The map does not take into account population trends by district, as data on such trends are not readily available. Yet enrolment declines of 6% or more, as seen in a band of districts running through the middle of the country, from Nelson Mandela (NM) to ZF Mgcawu (ZM) in Northern Cape, are very likely to be greater than

[15] File name *Provincial projection by sex and age (2002-2021)_web.xlsx*.

declines in the population of young children. Put differently, the red and especially purple districts are districts where special attention should be focussed on recovering levels of first-time enrolment.

Figure 4: Declines age 6.0 and below by district



The concerning declines in KwaZulu-Natal are concentrated in one relatively urbanised district to the north of Durban, Ilembe (IL), and the KwaZulu-Natal south coast.

8 Enrolment levels beyond compulsory schooling

The following four graphs indicate that at the secondary level, and beyond compulsory schooling, national participation in schooling did not decrease, but rather increased fairly substantially. Using figures from Figure 5, it can be seen that enrolment increased by 6% in Grade 9, 10% in Grade 11, and a whole 20% in Grade 12. In Grade 8 it remained almost the same, while in Grade 10 it declined by 2%. As will be discussed, this decline is probably not problematic, and it should be noted that Grade 10 enrolment remained higher than Grade 9 enrolment in both years. The graph uses the data after the removal of duplicate records. The other two versions of the data are not displayed because they produce virtually indistinguishable trends. This suggests that in key respects the original data were of a better quality for secondary-level learners than for primary-level learners[16].

Calculating grade repetition is difficult when learner identifiers are often missing, but trends in the average age per grade provide an idea of whether the proportion of repeaters has increased or decreased. This is what Figure 6 does. The large declines in the average age in grades 10 and 11, in the context of the enrolment trends by grade and (as will be seen) by age, could only be caused by lower levels of grade repetition. This would largely explain the decline in enrolment in Grade 10 seen in Figure 5. A decline in grade repetition in Grade 10 is arguably a positive development, given that this grade has historically displayed the highest

[16] This is with respect to the absence of duplicates, and consistency in the set of schools. However, with respect to the linkability of 13-digit ID numbers, the secondary level performs no better than the primary level.

levels of repetition of all grades. A recent study concluded that in 2018, 27% of Grade 10 learners were repeaters[17]. What is striking in Figure 6 is the sharp *increase* in the average age of Grade 12 learners, pointing to an increase in grade repetition here. This is likely to be because the system has decided to give more learners than usual a second chance, given the difficulties experienced by Grade 12 learners during the disruptive 2020 school year.

Figure 5: Enrolment by grade at the secondary level

Figure 6: Trends in the average age per grade

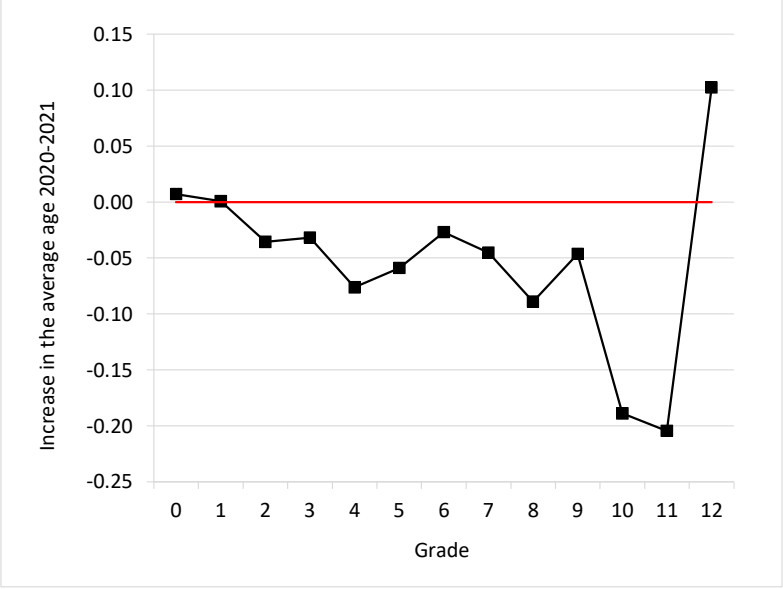


Figure 7 points to enrolment by age increasing, by up to 10% for age 16. For ages 13 to 18, the increase was 3.7%. This should be seen against the fact that the population aged 15 to 19 increased by 2.8% between 2020 and 2021, according to Stats SA[18]. Thus any enrolment increase of more than 2.8% represents an increase in the participation rate. From age 17, enrolment by age is remarkably constant. If the red 2021 curve is shifted to the left by one year, the result is Figure 8 which, like the earlier Figure 1, aligns the same birth cohort. The enrolment declines are low for the youngest, reaching 2.3% for those aged 15 at the start of 2020. This amounts to a decline of around 20,000 learners for this birth cohort, of which only

[17] Van der Berg *et al*, 2021.
[18] Same source as for Figure 3.

about 1,500 would be accounted for by mortality[19]. Above age 15 in 2020, dropping out is substantial. For instance, among those aged 17 at the start of 2020, 37% do not 'survive' to the next school year. Up to age 16, these levels of dropping out would not have increased between 2020 and 2021, given the age-specific enrolment increases seen in Figure 7, increases which exceed population growth. For age 17 and above, dropping out appears to remain as in the past if one ignores population growth, but if a population growth adjustment is applied, then a decline in participation rates for ages 17 is found, equal to around the population growth rate of 2.8%. However, this should be seen against the clear fact that participation in Grade 12 increased in absolute terms by 20%. Moreover, age-specific enrolment in Grade 12 increased for ages 17 to 25 by more than the 2.8% rate of increase in the population – this would be the gap between the thin red and green curves in Figure 7.

Figure 7: Enrolment by age at the secondary level

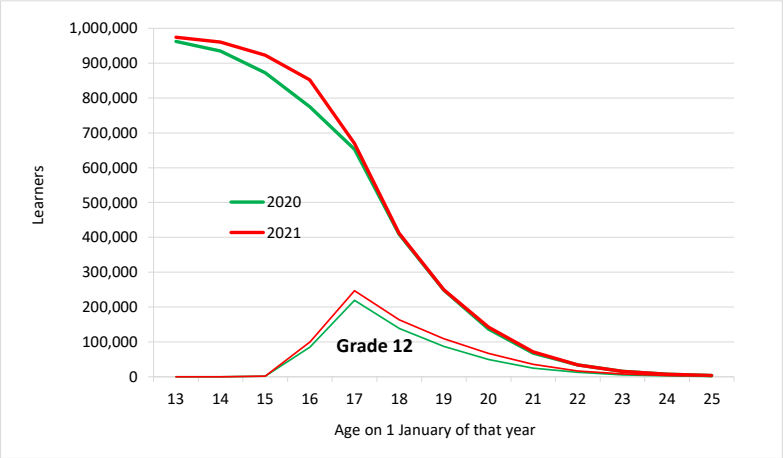
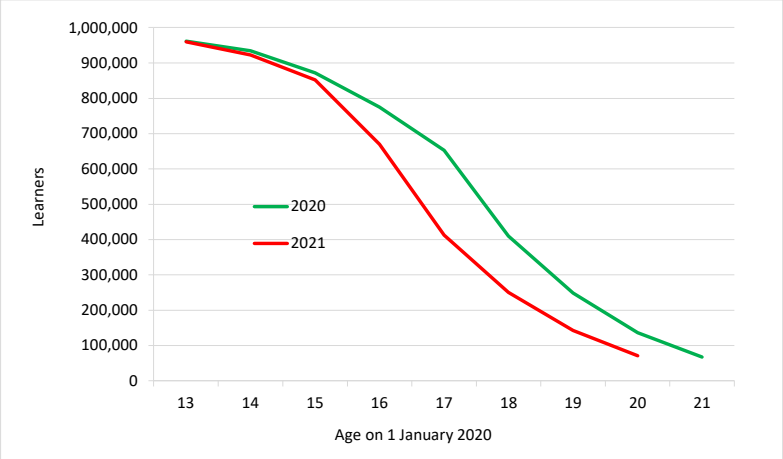


Figure 8: Trends for secondary-level birth cohorts



The following two graphs provide provincial statistics on enrolment changes by grade and age. Though several provinces experienced declines in Grade 8 enrolment, enrolment by age for the relevant ages, around 13 to 14, grew for almost all provinces – see Figure 10. The only decline was that for Northern Cape, where the number of learners aged 13 at the start of the year declined by 1%.

[19] Statistics South Africa, 2017: 8.

Figure 9: Provincial changes in enrolment by grade

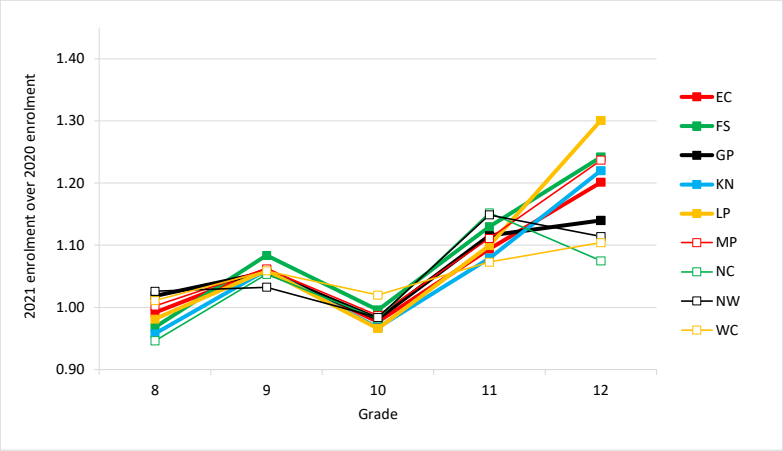
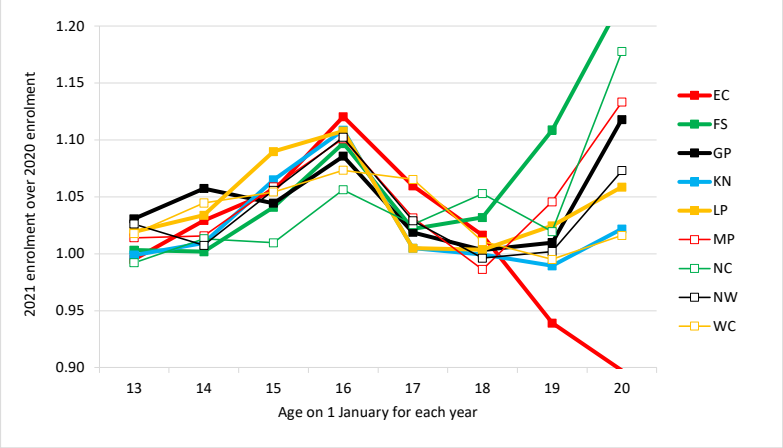


Figure 10: Provincial changes in enrolment per by age



Again, changes in enrolment must be compared to changes in the size of the population at certain ages. Table 12 shows the relevant population change statistics, using the same source as was used above. It is noteworthy that despite the declines for eight of nine provinces in the earlier age 5 to 9 range – see Table 11 above – the population is estimated to increase in the age 15 to 19 range, across all provinces, according to Stats SA[20]. Even if one takes these increases account, the conclusion is that participation in schooling at the secondary level grew between 2020 and 2021. In all provinces, enrolment growth outpaced growth in the population.

[20] It should be noted that official enrolment trends and official population trends have not always agreed in the past. Arguably, enrolment statistics are the more reliable of the two given the more intensive data collection processes, and almost no reliance on statistical modelling to produce the statistics. See for instance Gustafsson (2018).

Table 12: Declines ages 15 to 19 with demographic adjustments

	2020 to 2021 % change in pop. ages 15 to 19	% change in enrolment ages 15 to 19	% after adjustment	Enrolment increase ages 15 to 19 after adjustment
EC	4.0	5.7	1.7	6,767
FS	2.8	5.6	2.8	4,815
GP	1.3	4.3	3.0	16,723
KN	2.6	4.7	2.1	13,461
LP	4.4	5.5	1.1	4,364
MP	2.6	5.1	2.5	6,520
NC	2.7	3.2	0.5	354
NW	3.7	4.8	1.1	2,205
WC	2.4	5.5	3.1	8,267
SA	2.8	5.0	2.2	65,115

References

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*. Washington: World Bank.

Department of Basic Education (2019a). *General Household Survey (GHS): Focus on schooling 2018*. Pretoria.

Department of Basic Education (2019b). 2017 *School Monitoring Survey: Quantitative survey main report*. Pretoria.

Gustafsson, M. (2018). *Understanding the sharp primary level enrolment increases beginning in 2011*. Stellenbosch: University of Stellenbosch.

Pillay-Van Wyk, V., Bradshaw, D., Groenewald, P., Seocharan, I. *et al* (2020). COVID-19 deaths in South Africa: 99 days since South Africa's first death. *South African Medical Journal*, 110(11): 1093-1099.

Shepherd, D. & Mohohlwane, M. (2021). *The impact of COVID-19 in education - more than a year of disruption*. Stellenbosch: NIDS-CRAM.

Statistics South Africa (2017). *Mortality and causes of death in South Africa, 2015: Findings from death notification*. Pretoria.

Van der Berg, S., Van Wyk, C., Selkirk, R., Hofmeyr, H. (2021). *Learner flows through schools: Using high quality administrative data to understand education system performance.* Stellenbosch: University of Stellenbosch.