RESEP
Research on Socio-Economic Policy

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

EDUCATIONAL ISSUES AND THE IMPACT OF COVID-19:

What education data reveal

Authors
Servaas van der Berg, Chris van Wyk, Rebecca Selkirk,
Herman Meyer, Heleen Hofmeyr, Eldridge Moses and Joel Gondwe

Michael & Susan Dell
FOUNDATION

We gratefully acknowledge support from the
Michael & Susan Dell Foundation

Authors:
Servaas van der Berg, Chris van Wyk, Rebecca Selkirk, Herman Meyer, Heleen Hofmeyr, Eldridge Moses and Joel Gondwe

Design and layout:
Ink Design Publishing Solutions

December 2022

We gratefully acknowledge support from the Michael & Susan Dell Foundation

Michael & Susan Dell
FOUNDATION

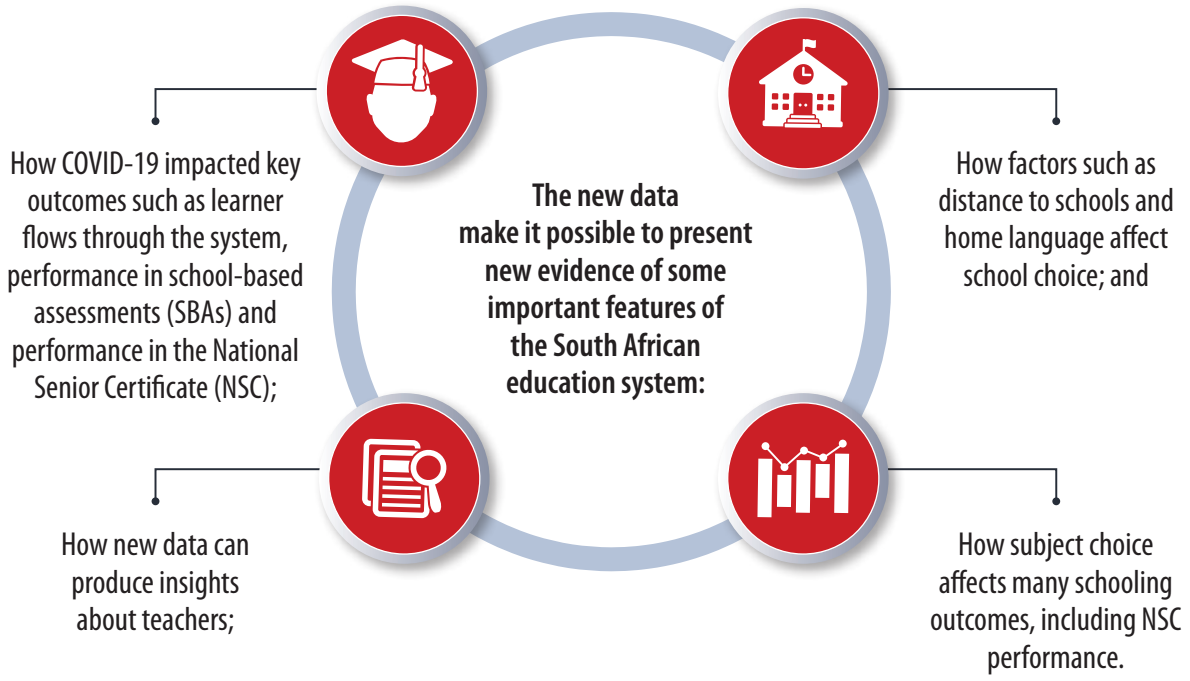
Contents

1 Introduction

1.1 Background

COVID-19 caused disruptions to schooling on an unprecedented scale. Emerging evidence indicates that these disruptions impacted schooling in South Africa in many ways, from shortened curricula to significant learning losses occurring over the two years of the pandemic (Hoadley, 2020; Ardington, Wills and Kotze, 2021; Van der Berg *et al*., 2022). This report constitutes the first attempt at investigating how these outcomes affected broader system performance in terms of key outcomes such as learner flows through the system, matric results, and performance in school-based assessments (SBAs). The report also sheds light on some important other education issues.

This research project follows from a previous project, also funded by the Michael & Susan Dell Foundation, that aimed to use high-quality administrative data to understand the performance of South Africa's schooling system (Van der Berg *et al*., 2021). Due to the continued efforts of the Data-Driven Dashboards (DDD) initiative, the Department of Basic Education (DBE), provincial education departments and schools, more data have become available which make it possible to extend the analysis presented in that report and to provide some new evidence of some important features of the South African education system:

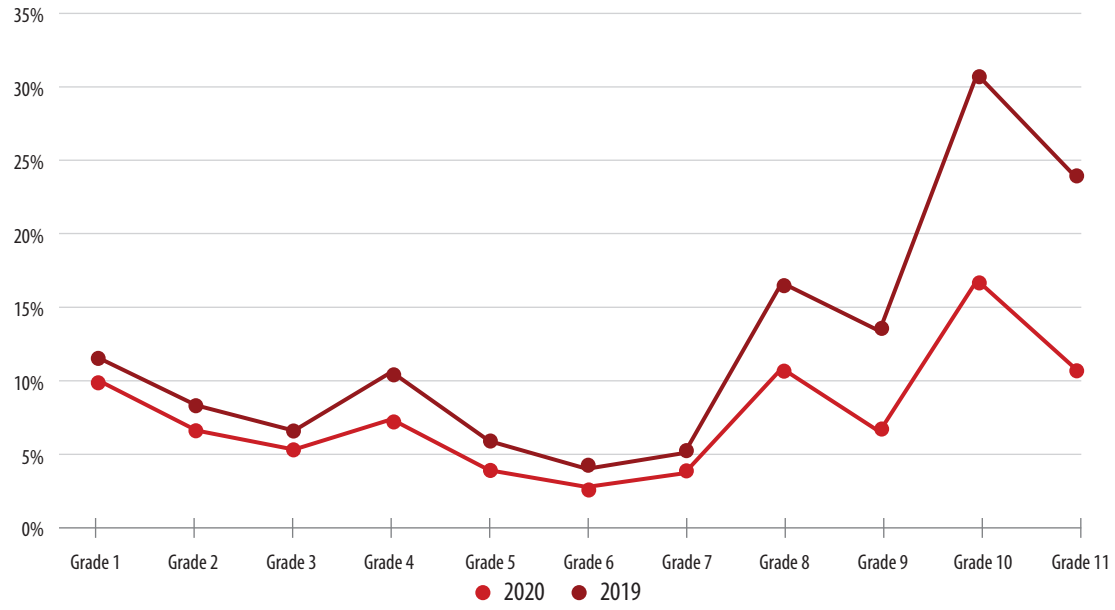


In addition to contributing to our understanding of the education system's performance, this report aims to show what kind of analysis is possible with the increasing amounts of high-quality administrative data that are becoming available. We hope that our analysis will inform future efforts to use this data for policy and practice.

1.2 The education context in the post-COVID situation

Most learners missed at least three-quarters of a school year over the course of 2020 and 2021, as a result of lockdowns, school closures and rotational timetables that were introduced to maintain social distancing in classrooms (Van der Berg *et al.*, 2022). As a response to these disruptions, the DBE introduced shortened curricula, which had major implications for how much learning occurred during the pandemic (Hoadley, 2020). Learning losses were indeed substantial: Based on results from the Western Cape Education Department (WCED) Systemic Tests, Van der Berg *et al.,* (2022) estimate that learners in 2021 were on average about a full year of learning behind in Mathematics and somewhat less in Language, compared to learners who were in the same grade in 2019. In addition to learning losses, incomplete curriculum coverage and shortened academic years significantly affected patterns of enrolment, grade progression, repetition and dropout[1] throughout the schooling system. A major finding of this report is that there has been a dramatic decrease in repetition throughout the system during the COVID-19 period. This can be seen in Figure 1, which shows the proportion of learners that repeated a year in each grade in 2019 and 2020 (based on Lurits data). The figure clearly shows that repetition decreased in all grades between 2019 and 2020, but especially in the high school grades. For example, repetition in Grades 10 and 11 almost halved between 2019 and 2020. These changes in learner flows, in combination with learning losses, have implications for enrolment numbers and class sizes in different grades, learner preparedness to take on the curriculum, and within-class heterogeneity. The evidence in this report helps to inform our understanding of these issues so that policymakers can respond timeously to the unfolding reality in South African schools.

Figure 1 Average South African repetition rates by grade before and during COVID



Source: LURITS 2019–2021

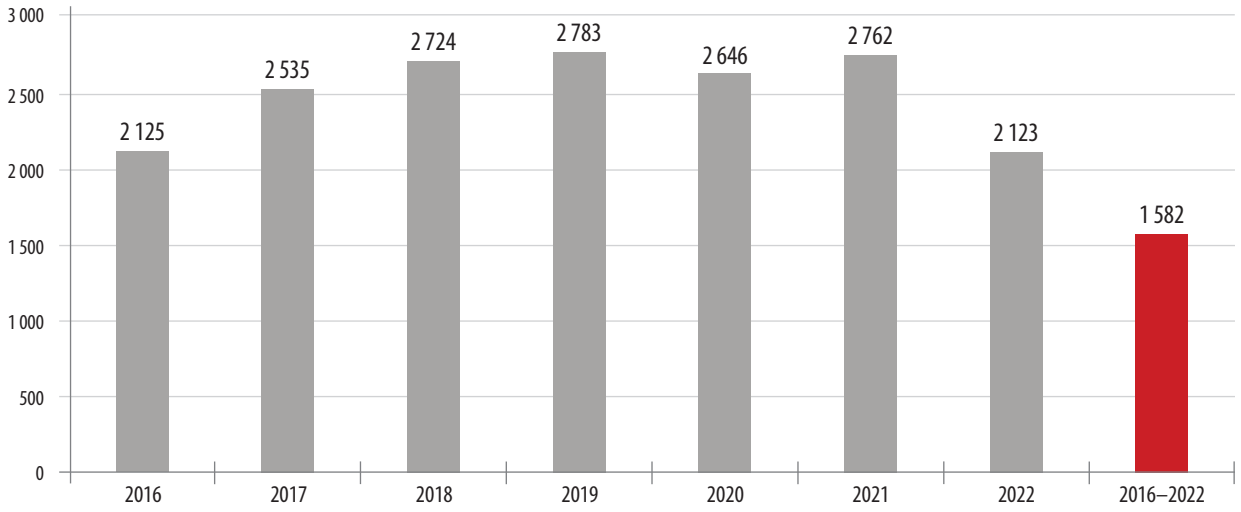
1 Dropout in this report refers to the unaccounted learners who were not identified in the datasets (SA-SAMS, DDD and LURITS) in a particular year after having been observed in the previous year. Because matching of learners across years is still imperfect, some learners are observed in a particular school and grade in one year, but in the next year they are not accounted for. This could mean they dropped out of the education system, but also that they could be some of the learners newly found in the next year, i.e. drop-ins. Improving the unique identifiers of learners across years, provinces and schools remains a major need in order to enhance the accuracy of analysis.

Data availability and the increased generation and linking of education data

The analysis in this report is based on four datasets, namely the Learner Unit Record and Tracking System (Lurits) data, the SA-SAMS database, data from the Data-Driven Districts (DDD) initiative, and National Senior Certificate (NSC) examination data. The DDD data are SA-SAMS data regularly collected from schools as part of the Data Driven Districts Programme that is coordinated by an NGO, the New Leaders Foundation, in agreement with the Department of Basic Education. This data are then used to populate Data Driven District Dashboards, a management tool. For this report, DDD data from different years were matched and linked over time to create a longitudinal dataset, and then anonymised and made available to Resep for this analysis. The DDD and the SA-SAMS data available for this research covered only the Eastern Cape, Limpopo and Gauteng. While these provinces are not fully representative of the situation nationally, they give a good picture of what is likely happening at the national level, as they represent 45 percent of all enrolments and 48 percent of the public schools in the country, according to the DBE's Masterlist of Schools for 2021 (DBE, 2022).

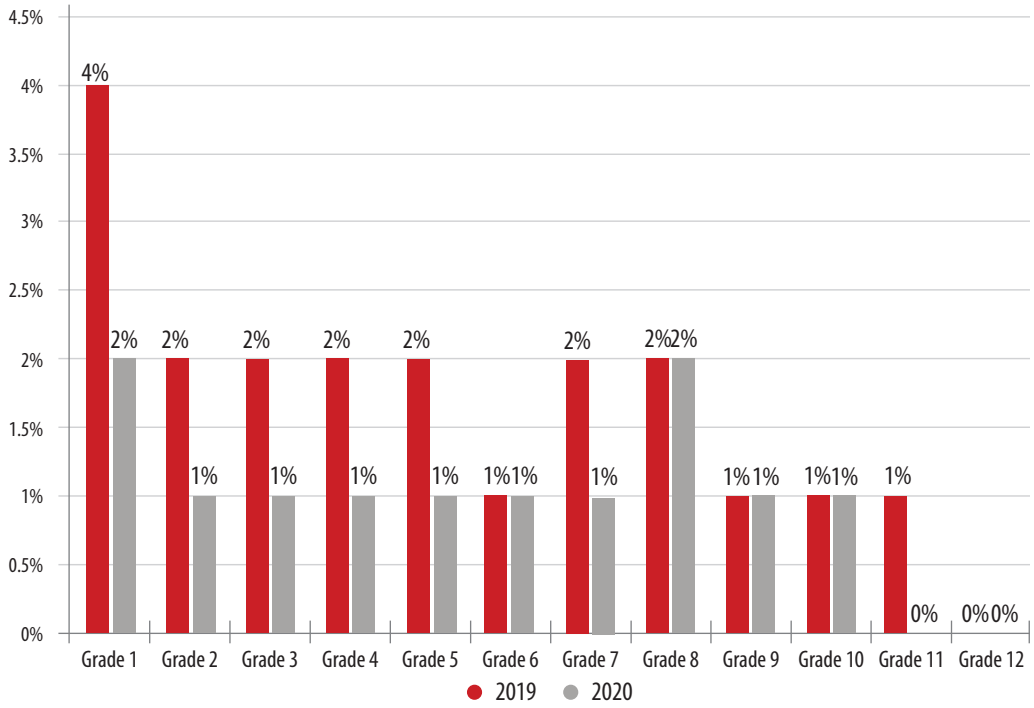
While the quality of the data stored in SA-SAMS has improved considerably in recent years, there are still various data quality issues that could be improved. Firstly, the available SA-SAMS data obtained for the DDD still have some inconsistencies, as the number of schools that submit their data every term varies. Since cohort analysis requires data from the same schools over time, only data from schools that submitted every year can be used to conduct such analysis. Figure 2 shows that although many schools in Gauteng submitted learner-level data each year for the period 2016–2022, the number that submitted varied greatly between years, with low levels of submission in the earlier years as well as 2022. Only 1 582 schools submitted each year for the entire period. Similar patterns are observable for Limpopo and the Eastern Cape. Analysis of longitudinal data presented in this report is, unless stated otherwise, only based on the sub-sample of schools that submitted data each year in the three provinces considered.

Figure 2 Number of Gauteng schools that submitted data by year



The quality and comprehensiveness of the data have improved over the period 2019 to 2022. For instance, the proportion of learner records collected for Gauteng that were duplicate records decreased substantially between 2019 and 2022, as Figure 3 shows. Similar improvements have occurred in Limpopo and the Eastern Cape. This indicates that information was captured with increased accuracy, resulting in improved data quality which can also enhance the accuracy of analysis.

Figure 3 Proportion duplicate learner records in Gauteng, 2019 versus 2022



Source: DDD data

Around the world, education management information systems (EMIS) data are often under-utilised. The purpose of collecting such administrative data is to provide information that can assist education policymaking. This is why this report emphasises the importance and quality of unit-level data and the power of longitudinal data.

Table 1 Description of datasets

DATASETS	PURPOSE	COMMENT
DDD	**Cohort Analysis**	*Linked learner unit-level data from schools that submitted every year from 2016 to 2022*
DDD	**Enrolment Patterns (Repetition, Dropout)**	*Linked learner unit-level data from schools that submitted every year from 2016 to 2023*
DDD	**School-Based Assessments (SBAs)**	*Learner unit level subject data from 2016 to 2021. (Linking SBA performance over time makes it possible to start investigating SBAs across schools, and also how such data relate to learner outcomes and learner flows.)*
SA-SAMS for EC LP and GT	**Age Distribution and Overage**	*Learner unit-level data from the Learner table in the SA-SAMS database of the current year*
SA-SAMS for EC LP and GT	**Teacher Details**	*Individual teacher data in the Educator table in the SA-SAMS database of the current year*
SA-SAMS for EC LP and GT	**Subjects Taught by Teachers**	*Individual teacher data: Linked current Educator table with the Educator Subject Taught table in the SA-SAMS database*
SA-SAMS GT	**Feeder Schools**	*Learner unit level data: Current Gr8 learners linked to Gr7 learners of the previous year from the Learner tables in the SA-SAMS database*
NSC data from DDD	**SBA data predict NSC outcomes**	*Individual NSC (matric) examination data matched to the DDD data through a unique identifier (anonymised SA ID)*
NSC from DBE	**NSC Performance over time**	*NSC data from 2008 to 2021*
Masterlist of Schools for 2021	**Integration of datasets and providing relevant details of schools**	*Uniquely identify each school in the country through a school identifier, generally called the "EMIS number"*

1.3 A preview of this publication

A major conclusion of this report is that the COVID-19 pandemic gave rise to clear discontinuities in South Africa's education system performance. This is evident in changes in repetition, dropout and progression patterns, as well as in performance in school-based assessments (SBAs) and even the NSC (matric) examinations. In analysing the effects of COVID-19 on key outcomes, the report also demonstrates how high-quality administrative data on individual learners can be used to understand important education questions, such as school choice and access to quality schooling. We also show how newly available data on teachers can bring new insights.

IN ANALYSING THE EFFECTS OF COVID-19 ON KEY OUTCOMES, THE REPORT ALSO DEMONSTRATES HOW HIGH-QUALITY ADMINISTRATIVE DATA ON INDIVIDUAL LEARNERS CAN BE USED TO UNDERSTAND IMPORTANT EDUCATION QUESTIONS, SUCH AS SCHOOL CHOICE AND ACCESS TO QUALITY SCHOOLING.

The report begins with an analysis of patterns of enrolment, repetition and progression using DDD data. Next, National Senior Certificate (NSC) data are analysed to investigate changes in the number of NSC candidates, passes and bachelor passes before and after COVID-19. Analysis of new teacher data stored in SA-SAMS is presented in Section 3. Section 4 presents some results using SA-SAMS data to investigate feeder school dynamics and access to high-quality schools in Gauteng. This is followed by an analysis of performance in school-based assessments (SBAs) obtained from DDD to investigate questions such as how subject choice is related to matric results, and how performance in the SBAs changed during the pandemic period (Section 5). Section 6 concludes.

The analysis presented in this report, in combination with findings from previous research, leads to the following conclusions regarding the impact of COVID-19 on education system performance in South Africa:

- First, a large amount of learning time was lost due to lockdowns and school closures (Ardington, Wills and Kotze, 2021).
- Secondly, lost learning time resulted in major learning losses over the two years of the pandemic (Van der Berg *et al*., 2022). Learning losses in the Western Cape Systemic Tests over the period 2019–21 for Grades 3, 6 and 9 amounted to 40–70% of a year in Language and 95–106% in Mathematics. Even the more affluent Quintile 5 schools experienced smaller yet still significant learning losses. Learning losses in other provinces are likely to have been at least of the same magnitude.

IN GRADES 4 TO 9, THE REQUIREMENT FOR HOME LANGUAGE WAS REDUCED FROM 50% TO 45%, THAT FOR THE (USUALLY ENGLISH) FIRST ADDITIONAL LANGUAGE FROM 40% TO 35%. FOR MATHEMATICS, THE PROMOTION REQUIREMENT OF 40% WAS DROPPED ALTOGETHER …

- Thirdly, to be fair to learners who had missed so much school through no fault of their own, the education authorities applied more lenient promotion rules:
  - For Grade 10 and Grade 11, the DBE increased the SBA component of the promotion requirements from 25% to 60%. Moreover, controlled tests replaced examinations at the end of the year, and it was required that these tests *"should only be set on content taught, content not taught cannot be assessed"* (Circular S7, *Revised promotion requirements for Grades 10 and 11 for the 2020 year*).
  - Adjustments were also made to the assessment rules for Grades 4 to 9. The requirement for Home Language was reduced from 50% to 45%, that for the (usually English) First Additional Language from 40% to 35%. For Mathematics, the promotion requirement of 40% was dropped altogether, meaning that if all other requirements are met, Mathematics performance would be condoned, irrespective of performance. Moreover, it was explicitly stated that learners whose Mathematics Grade 9 mark has been condoned would still be able to choose to continue with Mathematics in Grade 10, i.e. they would not be compelled rather to take Mathematics Literacy (*National Assessment Circular 5 of 2022*).

Greater leniency is observed in some of the SBA results for the pandemic years that are discussed later in this report.

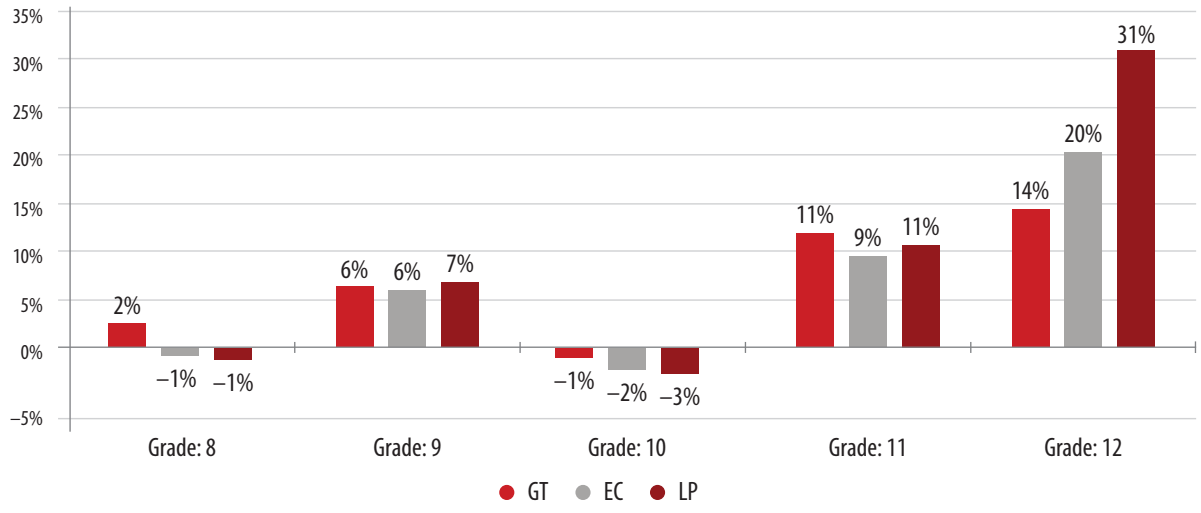
- Fourthly, despite the learning losses, learner flows through the system accelerated, presumably because assessments had become so much more lenient.
- Fifthly, as a consequence, enrolment in matric and the number of candidates writing the NSC examination increased by very large proportions, while pass rates and bachelor pass rates remained virtually unchanged, implying that the number of passes and bachelor passes increased greatly. This has the potential to seriously undermine the integrity of the NSC in the eyes of the public, of universities and of employers.

2 Patterns of enrolment, progression and repetition

2.1 Enrolment patterns

A clear trend in the data is that enrolment patterns changed substantially as a result of the pandemic. The high school grades, in particular, saw significant changes to usual enrolment patterns. This can be seen in Figure 4, which shows the percentage change in enrolments for Grades 8 to 12 in Gauteng, Limpopo and the Eastern Cape. A clear trend in all three provinces is that enrolment in Grades 8 and 10 decreased, while it increased in Grades 9, 11 and 12, with particularly large increases observed in Grades 11 and 12. This trend indicates that one of the results of lower repetition rates in Grade 10 is that fewer learners were enrolled in Grade 10 in 2021 (as a result of more learners being promoted to Grade 11, where enrolment increased). The data therefore suggest that the pre-pandemic pattern whereby learners were often held back in Grade 10 as a form of "gatekeeping" (reportedly as many schools attempted to protect their matric performance) occurred to a lesser extent than before. Analysis of NSC data presented later in this report shows that these changes in enrolment patterns were also reflected in the NSC data, with many more learners writing the matric exams in 2021 than had been the case before the pandemic.

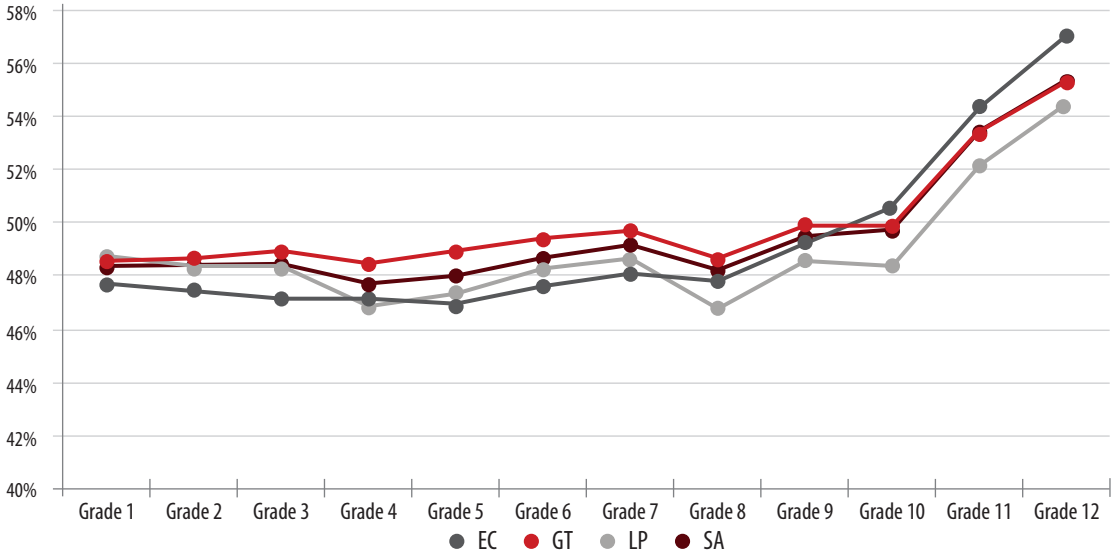
Figure 4 Percentage change in enrolment numbers in three provinces between 2020 and 2021



Source: LURITS

Figure 5 shows how enrolment patterns differ by gender in the three provinces considered. The figure shows the proportion of female learners enrolled in each grade in each of the three provinces. Enrolment patterns by grade and gender are roughly similar across the three provinces, with slightly fewer females than males being enrolled in the primary school grades (in part because more boys repeated. After this, the pattern reverses, with higher proportions of females than males being enrolled in the FET phase. For example, while females in the Eastern Cape made up roughly 48% of enrolments in Grade 1, this proportion was 57% in Grade 12. This confirms our previous finding that boys are more likely to drop out of school (Van der Berg *et al.*, 2021).

Figure 5 Female enrolment share in three provinces by grade, 2021

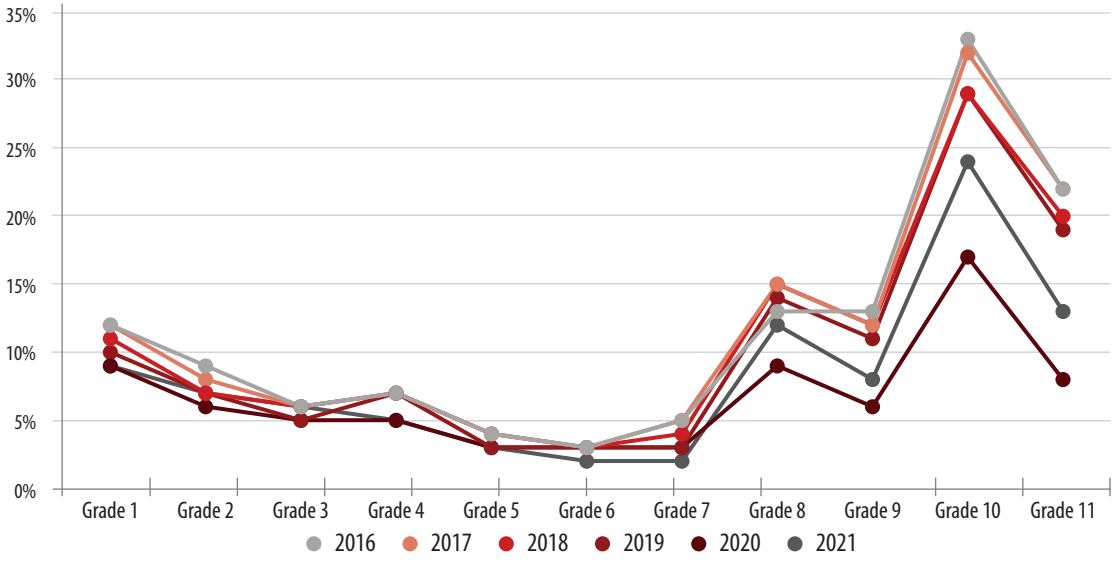


Source: Lurits (2021) and DDD (2021)

2.2 Progression, dropout and repetition: What the data can tell us

As mentioned, one of the results of the pandemic was that repetition rates decreased substantially throughout the system. This can be seen in Figure 6, which shows grade repetition rates for Gauteng for the period 2016 to 2021. The high school grades saw particularly large decreases in repetition rates, especially in Grades 10 and 11. For example, in Gauteng the proportion of learners repeating Grade 10 roughly halved (from 33% to 17%) between 2016 and 2020. Although repetition rates increased somewhat in 2021, they were still far below pre-pandemic levels. Similar patterns hold for Limpopo and the Eastern Cape.

Figure 6 Gauteng repetition rates by grade and year, 2016–21



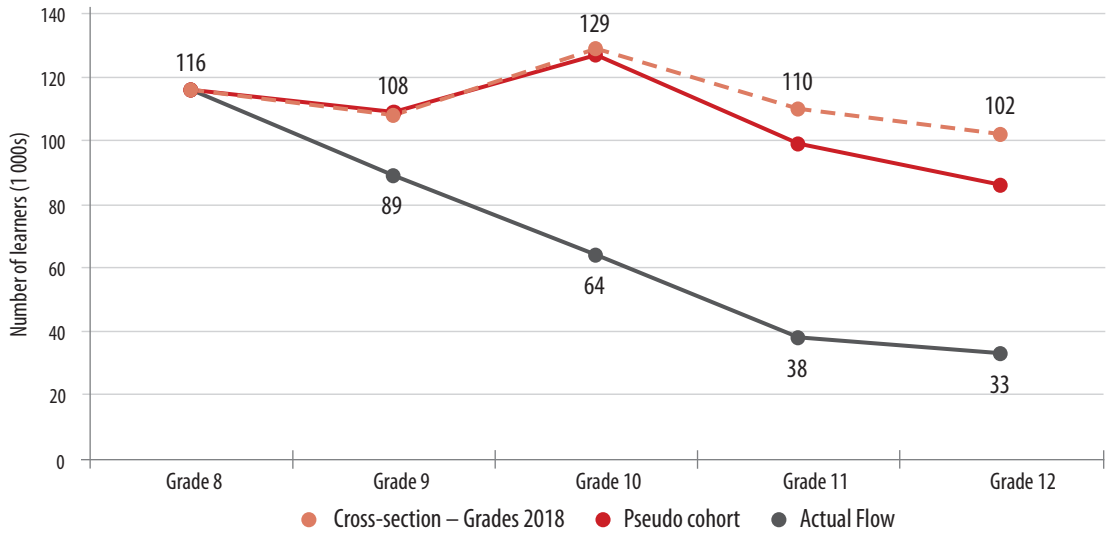
Source: DDD data

2.3 Cohort analysis of learner flows

While cross-sectional data can provide a snapshot of enrolment patterns per grade and year, the data made available by the DDD initiative can be used to match individual learners across years to analyse the flows of learners through the system. Learner-level records are especially useful for determining patterns of on-track flows, that is, learners progressing through the system without repeating a grade and being in the appropriate grade for their age at any point in time.

Figure 7 shows how learner-level records can provide a much more accurate reflection of flows than data constructed from repeated cross-sections. The figure compares the progression of a true cohort of learners (the black line) who were in Grade 8 in 2017 with cross-sectional enrolment data (the dotted line) and a pseudo-cohort (the red line) over the period 2017 to 2021 for the Eastern Cape. The figure shows how cross-sectional data overestimate progression, since repeaters from a previous year are included in enrolment numbers in a given grade. Following the red lines, it is clear that many learners were held back (and/or dropped out or could not be accounted for in the sample) each year between Grades 8 and 12. As a result, only a relatively small proportion of the true cohorts that were in Grade 8 in 2017 were in Grade 12 by 2021. Similar patterns are observable in Gauteng and Limpopo. The figure, therefore, shows how considering true cohorts of learners (constructed from learner-level panel data) provides a much more accurate picture of learner flows through the system. Collecting quality learner-level data that can be linked over time thus enables us to paint a much more accurate picture of flows through the schooling system.

Figure 7 Number of Eastern Cape learners (1 000s) by grade and learner flow type, 2017–21

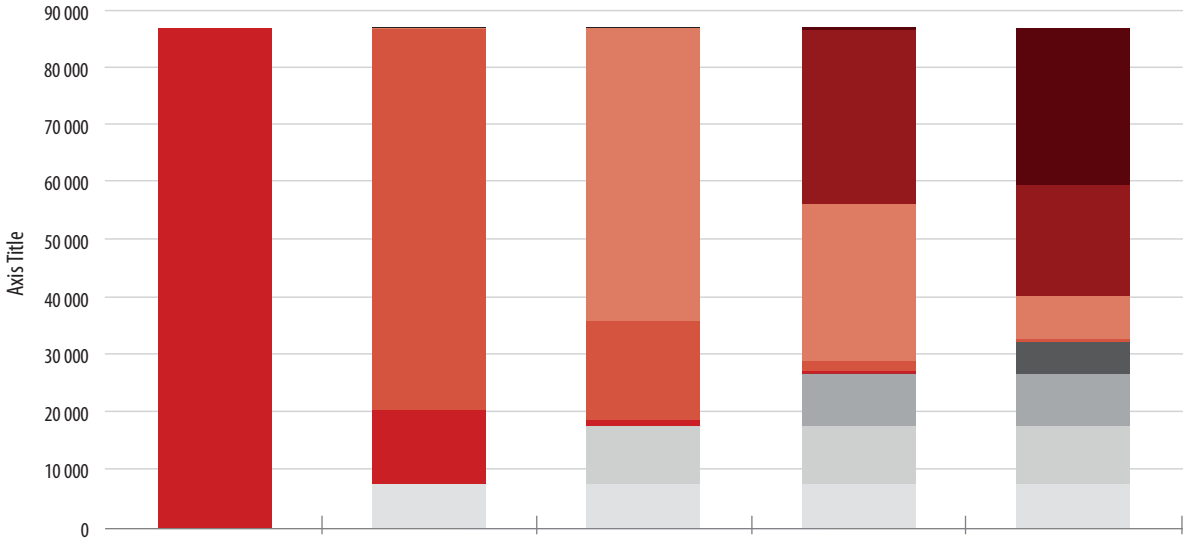


Source: DDD

Figure 8 shows what happened to the 2017 Gauteng Grade 8 cohort of learners. As indicated before, the analysis is restricted to those schools which have submitted data to the DDD Programme for all the years covered. Of the 116 306 learners in Grade 18 in 2017, which include some who were already repeating Grade 8 in 2017, 88 625 progressed to Grade 9 in 2018 16 083 repeated and remained in Grade 8 (hence they retain the same colour), and

11 598 were no longer in the balanced panel of Gauteng schools on which this analysis is applied – they are thus unaccounted for, which could mean they left to other schools or other provinces, dropped out (i.e. left school altogether), or a small number may even have died or moved to another country. It is conventional to assume that this group largely constitutes drop-outs, but not all of them are. In 2019, some of the cohort that had reached Grade 9 progressed again, some repeated Grade 9, and some even repeated Grade 8 again, which is contrary to the repetition policy that states that they may only repeat once in a school phase. A new group could now not be accounted for, again probably largely but not exclusively drop-outs. This situation is repeated in Grades 10 and 11, so that in 2021, we can classify the full 2017 cohort into different groups: Those who have reached matric without delay in the 4 years under review, those who had repeated Grade 8, Grade 9, Grade 10 or Grade 11, those that could not be accounted for already in 2018 (Grade 9), those not accounted for in 2019 (Grade 10), those not accounted for in 2020 (Grade 11), and those not accounted for in 2021 (Grade 12). It is clear from this picture that only about 29% of the cohort reached matric in the minimum period.[2]

Figure 8 Grade enrolment, progression and dropout/unaccounted for among the 2017 Gauteng Grade 8 cohort



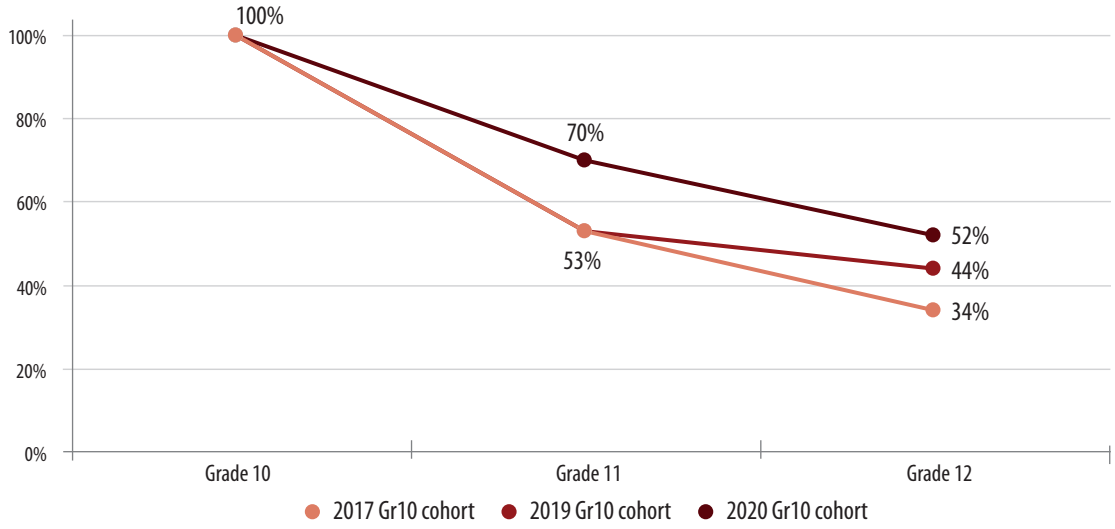
	2017	2018	2019	2020	2021
Dropout/unaccounted Gr 8	0	7 671	7 671	7 671	7 671
Dropout/unaccounted Gr 9	0	0	10 033	10 033	10 033
Dropout/unaccounted Gr 10	0	0	0	9 126	9 126
Dropout/unaccounted Gr 11	0	0	0	0	5 444
Gr 8	87 053	12 835	1 163	372	98
Gr 9	0	66 342	17 155	1 883	401
Gr 10	0	133	50 901	27 283	7 533
Gr 11	0	67	108	30 325	19 340
Gr 12	0	5	22	360	27 407

Source: DDD data. This is a balanced panel, covering only those schools for which data were submitted in all the years concerned (Dropout here refers to the unaccounted learners who were not identified in the dataset).

2 Although here one should add that it is possible that some of those unaccounted for could have entered another part of the school system, and then perhaps even joined the small group that progressed without interruption.

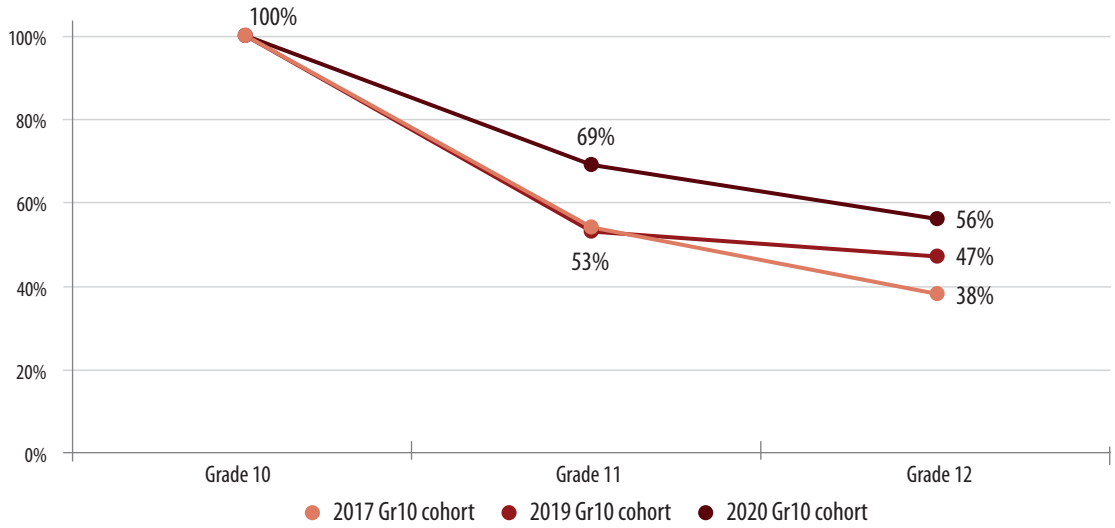
The effect of the reduced repetition rates in Grades 10 and 12 on flows to matric can be seen in Figure 9 for the Eastern Cape. It shows the percentage of the Grade 10 cohort (including earlier repeaters) that reached Grade 11 and then Grade 12 without any further repetition, i.e. in two years. This percentage rose from 34% for the 2017 Grade 10 cohort (similar to most of the pre-pandemic cohorts) to 44% for the 2019 cohort, which experienced the more lenient promotion pattern only when they had reached in Grade 11 in 2020, and 52% for the 2020 cohort, which experienced such greater leniency in both grade 10 in 2020 and grade 11 in 2021. Thus the direct flow without repeating of Grade 10 cohorts to matric in two years increased by more than a third compared with the pre-pandemic cohort. For the case of Gauteng (Figure 10), the 2020 cohort saw an almost 50% increase in this flow through, from 38% to 56%.

Figure 9 Eastern Cape flow through patterns to Grade 12 for the 2017, 2019 and 2020 Grade 10 cohorts



Note: Calculated from a balanced panel of schools in the DDD data. The 'flow through' measured here is the proportion reaching the grade in question without any (further) repetition.

Figure 10 Gauteng flow through patterns to Grade 12 for the 2017, 2019 and 2020 Grade 10 cohorts

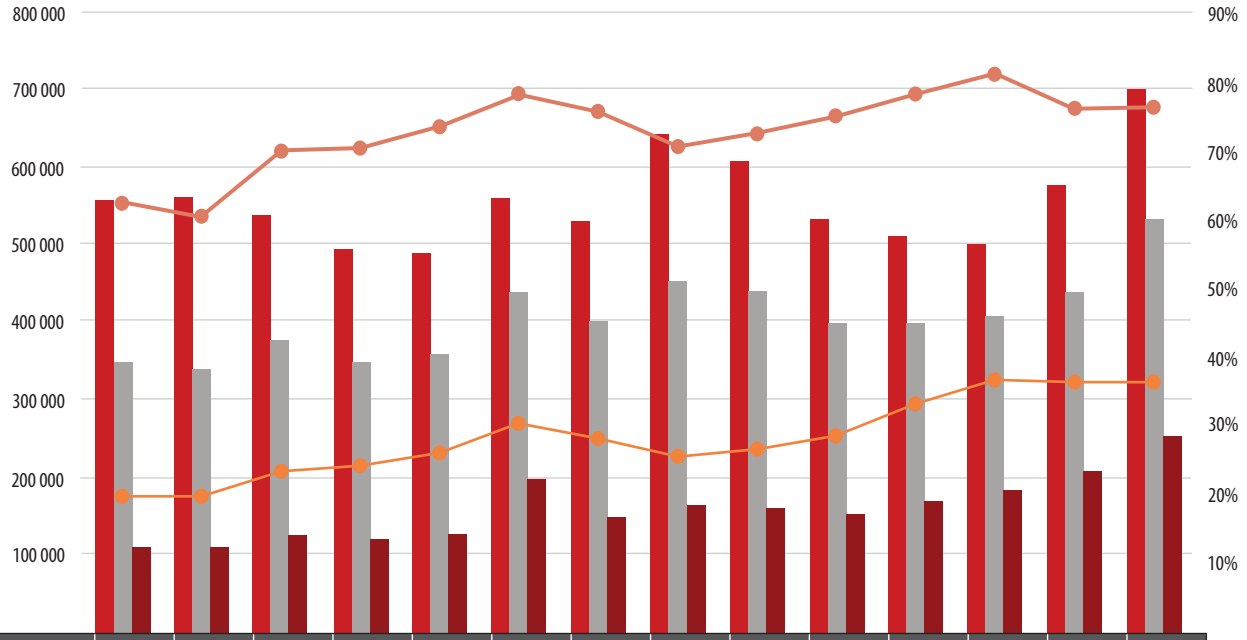


Note: Calculated from a balanced panel of schools in the DDD data. The 'flow through' measured here is the proportion reaching the grade in question without any (further) repetition.

2.4 Patterns in NSC results

Given the evidence of much higher flow-through rates through the system in 2020 and 2021, it is instructive to consider how these changes translated into National Senior Certificate (NSC) results. Figure 11 shows national NSC results for this period. There was a very large increase of about 200 000 matric candidates from 2019 (before the pandemic) to 2021 (after two pandemic years). This largely reflects how more learners were promoted to Grade 12 during 2020 and 2021 than had been the case before the pandemic, but is also influenced by more matric candidates being given a second chance to write this examination. The number of NSC and bachelor passes increased similarly during the pandemic period: There were roughly 130 000 more matric passes in 2021 than in 2019, and roughly 70 000 more bachelor passes in 2021 than in 2019. The fact that more learners achieved these matric outcomes during 2020 and 2021, despite the lost learning time and resultant learning losses due to the COVID-19 pandemic, can potentially endanger the integrity of the NSC examinations.

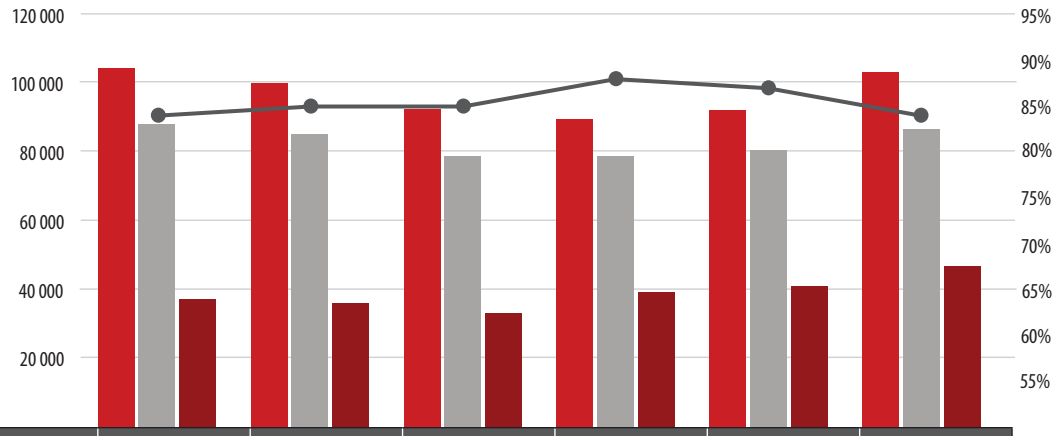
Figure 11 South African NSC results, 2008–21



	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Wrote	558 743	562 671	539 672	496 047	491 401	532 629	532 629	644 316	608 909	534 420	512 735	504 315	578 468	703 936
Passed	349 663	341 198	377 850	349 743	361 234	439 943	403 720	455 788	442 418	401 381	400 761	409 913	440 702	537 626
Bachelors	111 539	111 442	127 346	121 305	128 976	150 752	150 752	166 268	162 325	153 640	172 043	186 060	210 820	256 028
Pass Rate	63%	61%	70%	71%	74%	78%	76%	71%	73%	75%	78%	81%	76%	76%
Bachelors Rate	20%	20%	24%	24%	26%	31%	28%	26%	27%	29%	34%	37%	36%	36%

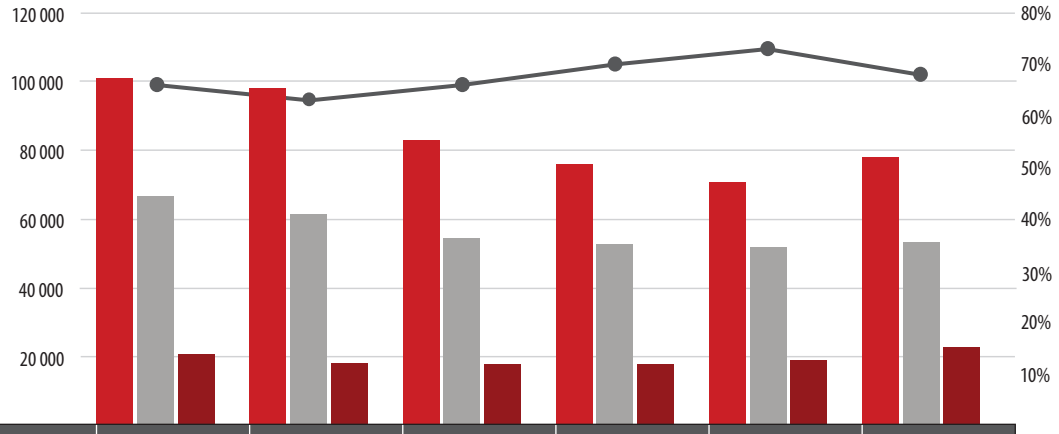
Figure 12, Figure 13 and Figure 14 show NSC results obtained from DDD for Gauteng, Limpopo and the Eastern Cape, respectively. As was the case nationally, the number of matric candidates increased between 2019 and 2020 for all three provinces. As observed nationally, the pass *rate* decreased slightly in these three provinces, but all three provinces showed a large increase in the *number* that passed as well as the *number* of bachelor passes.

Figure 12 Gauteng NSC results, 2015–20



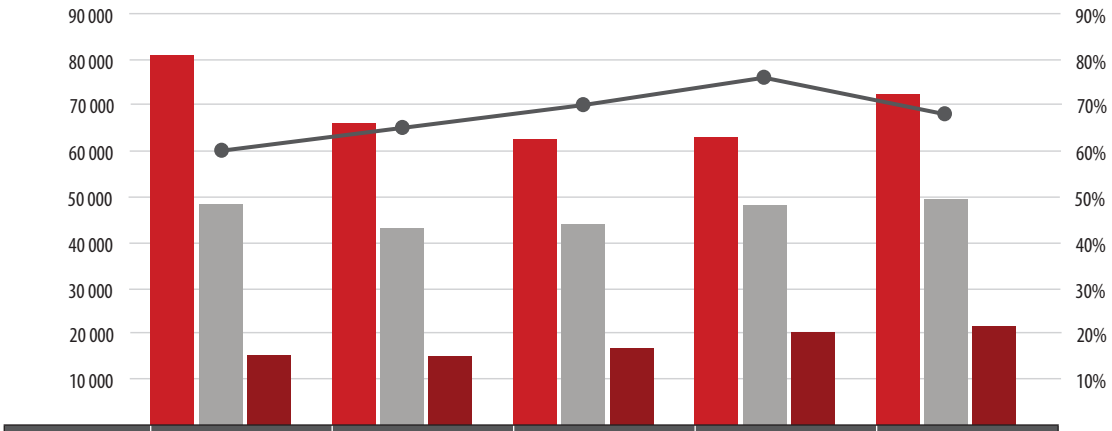
	2015	2016	2017	2018	2019	2020
Wrote	104 359	99 980	92 322	89 426	92 177	103 198
Passed	88 063	85 041	78 762	78 796	80 454	86 636
Bachelors	37 291	36 045	33 357	39 171	41 062	46 770
Pass Rate	84%	85%	85%	88%	87%	84%

Figure 13 Limpopo NSC results, 2015–20



	2015	2016	2017	2018	2019	2020
Wrote	100 680	97 892	82 697	75 874	70 552	77 935
Passed	66 415	61 269	54 295	52 742	51 611	53 097
Bachelors	20 832	18 017	17 672	17 814	18 886	22 649
Pass Rate	66%	63%	66%	70%	73%	68%

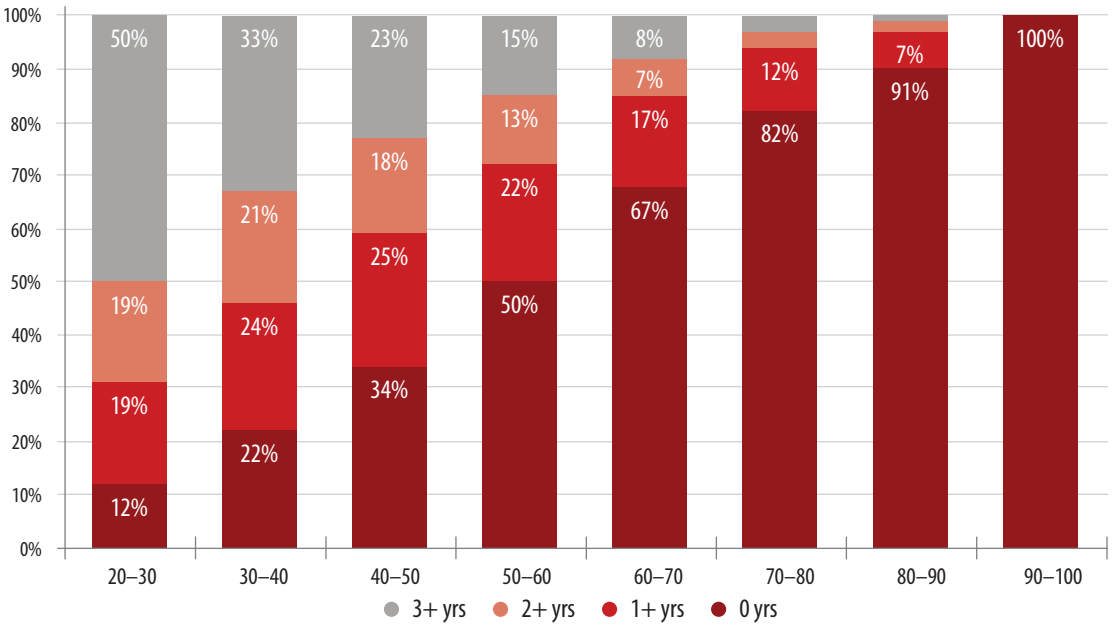
Figure 14 Eastern Cape NSC results, 2015–20



	2016	2017	2018	2019	2020
Wrote	80 675	65 881	62 297	62 871	72 257
Passed	48 141	42 928	43 912	48 067	49 266
Bachelors	15 376	15 084	16 966	20 329	21 727
Pass Rate	60%	65%	70%	76%	68%

Since many learners are over-aged for their grades, it is instructive to consider how being over-aged is related to NSC results. Figure 15 shows NSC results obtained from DDD for learners in the Eastern Cape who are on track (0 years over-age), 1 year, 2 years or 3 or more years over-age. Over-age learners perform worse on average than learners who are in the appropriate age-for-grade, with learners who are three or more years over-age performing particularly poorly. Similar patterns hold in Gauteng and Limpopo. Cape.

Figure 15 Eastern Cape NSC results by and age and performance categories, 2019



3 Some information about teachers

The SA-SAMS database contains two tables on teachers, which have not been analysed before. The one contains details about teachers' demography and the nature of their appointment. The second table contains information on the subjects that teachers teach and the grade at which they teach each subject. While the first table is more complete than the second, both allow us to get some evidence about teachers which has not been available in the past. So, for instance, we can see that the impending retirement wave of public school teachers affects teachers paid by school governing bodies (SGBs) far less than teachers paid by the state; Figure 16 contrasts the age distribution of teachers paid by the state in all three the provinces investigated (figure on the left) and SGB-paid teachers in public schools (figure on the right). While many teachers in public schools are nearing retirement age (teachers in public schools can retire from age 55), SGB-paid teachers are more likely to be below age 35. The figure on the left also shows that Gauteng has many more young state-paid teachers than the other two provinces and fewer that would retire soon.

Figure 16 Age distribution of state-paid and SGB-paid teachers in three provinces, 2022

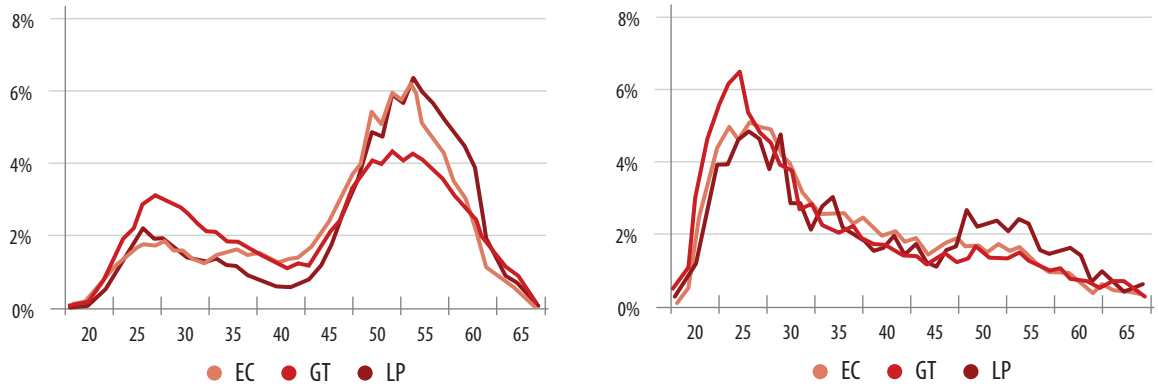


Table 2 shows that a much larger proportion of state-paid teachers are 50 years or older than is the case for public school teachers paid by SGBs, particularly in Limpopo.

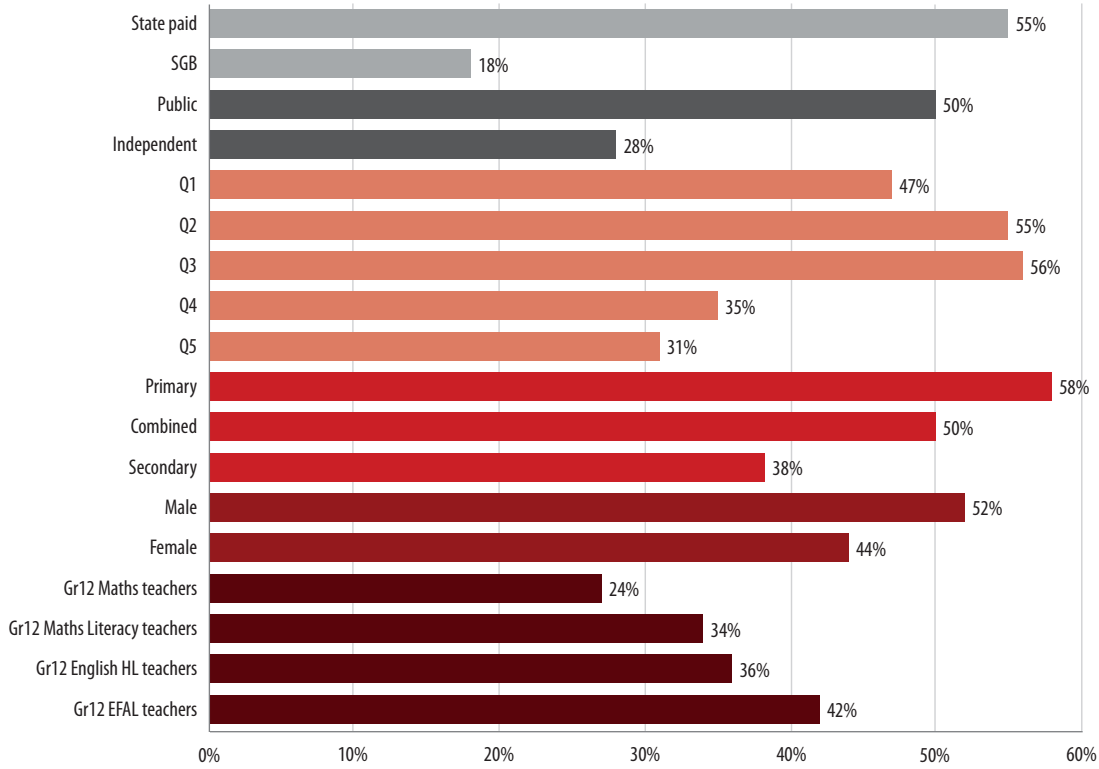
Table 2 Percentage of teachers in public schools in three provinces that are 50 years or older, 2022

	STATE-PAID	SGB-PAID
Eastern Cape	49%	17%
Gauteng	42%	17%
Limpopo	58%	25%

Source: SA-SAMS 2022

The teacher information in SA-SAMS in the Eastern Cape even allows some analysis of which teachers teach which subjects (Figure 17). This figure shows, for the Eastern Cape, what categories of teachers are more likely to be 50 years or older. The teacher groups that stand out as being younger are those in independent schools, and, within public schools, those paid by SGBs, in Quintile 4 or 5, or secondary schools. Among teachers in public schools teaching in Grade 12, those teaching Mathematics and, to a lesser extent, Mathematical Literacy tend to be younger.

Figure 17 Percentage of Eastern Cape teachers aged 50 years or more, 2022



Source: SA-SAMS 2022

Table 3 also utilises information from the Eastern Cape on subjects taught. This shows the proportion of all public school teachers teaching the three subjects shown in Grade 12 that are SGB-paid. It is surprising how common SGB teachers are even in the lowest quintile of schools. In both poor and wealthy schools, SGB teachers make an especially significant contribution to teaching the so-called gateway subjects of Mathematics and Physical Science.

Table 3 Percentage of Eastern Cape public school teachers teaching three gateway subjects in Grade 12 that are SGB-paid, 2022

	Q1	Q2	Q3	Q4	Q5	ALL
Maths	12%	8%	7%	30%	41%	**13%**
Physical Science	13%	9%	9%	20%	36%	**13%**
EFAL	7%	6%	5%	15%	49%	**7%**

Source: SA-SAMS 2022

A surprisingly large proportion – about one in eight – of state-paid teachers in these three provinces are temporary teachers (Table 4). Also interesting is that around 62% of SGB-paid teachers hold permanent positions.

Table 4 Nature of appointment of teachers in public schools in three provinces, 2022

	SGB-PAID	STATE-PAID	TOTAL
Permanent	11 874	140 390	152 264
Substitute	558	2 830	3 388
Temporary	6 750	19 147	25 897
Total	**19 182**	**162 367**	**181 549**

Source: SA-SAMS 2022

4 Feeder schools in Gauteng

4.1 What we can tell about feeder schools from the available data

This section summarises the results of the analysis that will also be presented in more detail in a separate working paper (Moses, forthcoming) on what the available data can tell us about feeder schools. Given South Africa's bimodal education system whereby there are large quality differences between different parts of the education system (Fleisch, 2008; Spaull, 2013), school choice is a crucial component of equalising opportunity.

Various barriers constrain school choice. The most visible barrier to entry to high-performing schools is the fees they charge, which are unaffordable to most South Africans. A second barrier to the diversification of learner bodies in high-quality schools is feeder zone rules, whereby some schools and provinces give preference in admission to learners who live within relative proximity to a given school. As South African neighbourhoods are often not racially diverse, these types of constraints may reinforce pre-existing racial segregation between schools, particularly in urban areas.

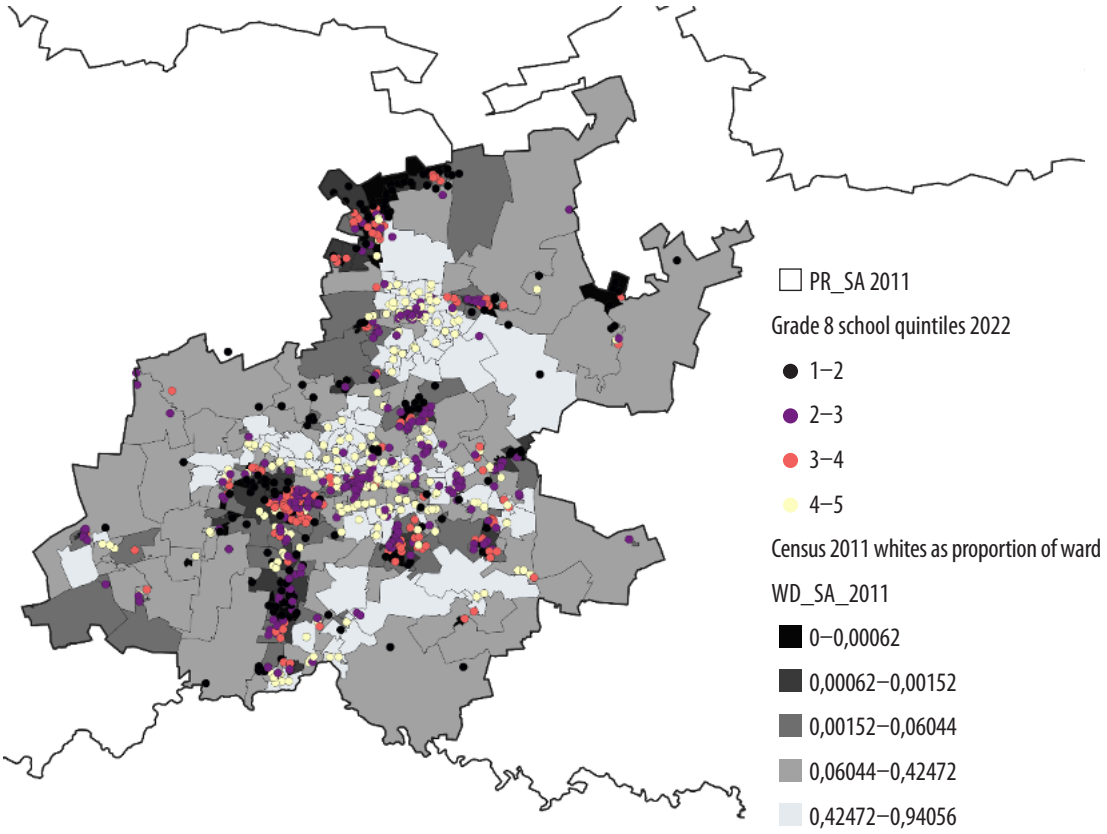
Analysing the extent of geographical mobility between primary and high schools in Gauteng given the feeder zone rule can reveal whether good quality schools are indeed more racially and socio-economically diverse than the neighbourhoods they are located in. This analysis is confined to Gauteng, as this issue is especially pertinent in metropolitan areas with potentially more possibilities for school choice. The focus falls on the transition from primary to secondary schools.

For this purpose, the DDD data from SA-SAMS are coupled with 2019 NSC results. The DDD data provide information on the location of a learner's primary school in 2021, the new high school that the learner moves to in 2022, as well as the school language of instruction and the learner's home language, age, race and gender. The data do not contain information about the learner's home address; therefore we can only analyse the movement from primary to high school. The implicit assumption is therefore that learners' primary school is usually located close to their homes. The NSC results of 2019 are used to group Gauteng high schools according to performance for the analysis.

4.2 Geographic and social mobility

The apartheid legacy of school segregation is evidenced by the distribution of schools by school quintile in 2022. Figure 18 shows where Gauteng's high schools are located, the school quintile assigned to them by the Department of Basic Education (a measure of the wealth of the community they serve) and the proportion of people who are white in each ward (obtained from Census 2011 data). White areas below show wards where at least 42% of the resident population is white. Most Quintile 5 high schools (the most affluent) schools are located in these largely white areas. In contrast, the poorest high schools (Quintiles 1 and 2) are predominantly located in Gauteng's townships and close to the provincial borders in the north of the province.

Figure 18 Gauteng high schools by school quintile and white population ratio, 2022

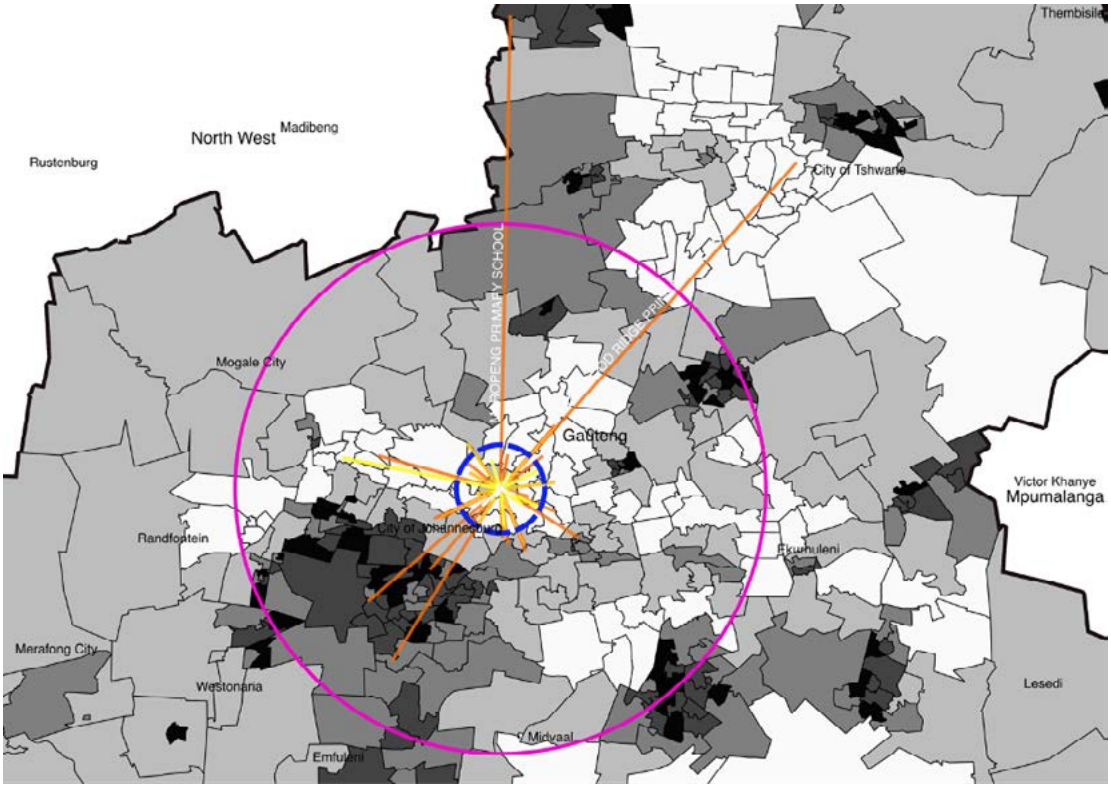


Source: Moses (forthcoming) based on SA-SAMS data (Data Driven Dashboards, 2022).

In response to the persistent inequalities in access to good schools, the Gauteng Department of Education increased the feeder zone radius from 5 km to 30 km in the 2020 school year to diversify the learner population in affluent schools, as the larger radius around such schools would now include neighbourhoods of colour. An example of the potential effect of the increased catchment/feeder area is illustrated for a high-performing high school in Figure 19. The former 5 km radius around the school is shown in blue, while the new 30 km radius feeder zone introduced in 2020 is shown in purple. Learner movements between primary schools and the high school are shown in orange to yellow, with lighter shades indicating larger movements of learners from specific primary schools to the high-performing high school. As can be seen, the change to a 30 km proximity rule included many children residing

in mostly black, coloured and Indian/Asian areas, yet the learner movement for this high-performing high school shows that most learners are still drawn from a relatively tight radius of 5 km around the school, coming from mostly white neighbourhoods. It could be that the other major barrier to entry, school fees, operated to maintain the pattern of entry relatively intact.

Figure 19 Feeder primary schools for a high-performing Gauteng high school's Grade 8 intake, 2022



Source: Moses (forthcoming) based on SA-SAMS data (Data Driven Dashboards, 2021 and 2022) and NSC 2019 data (Department of Basic Education, 2020).

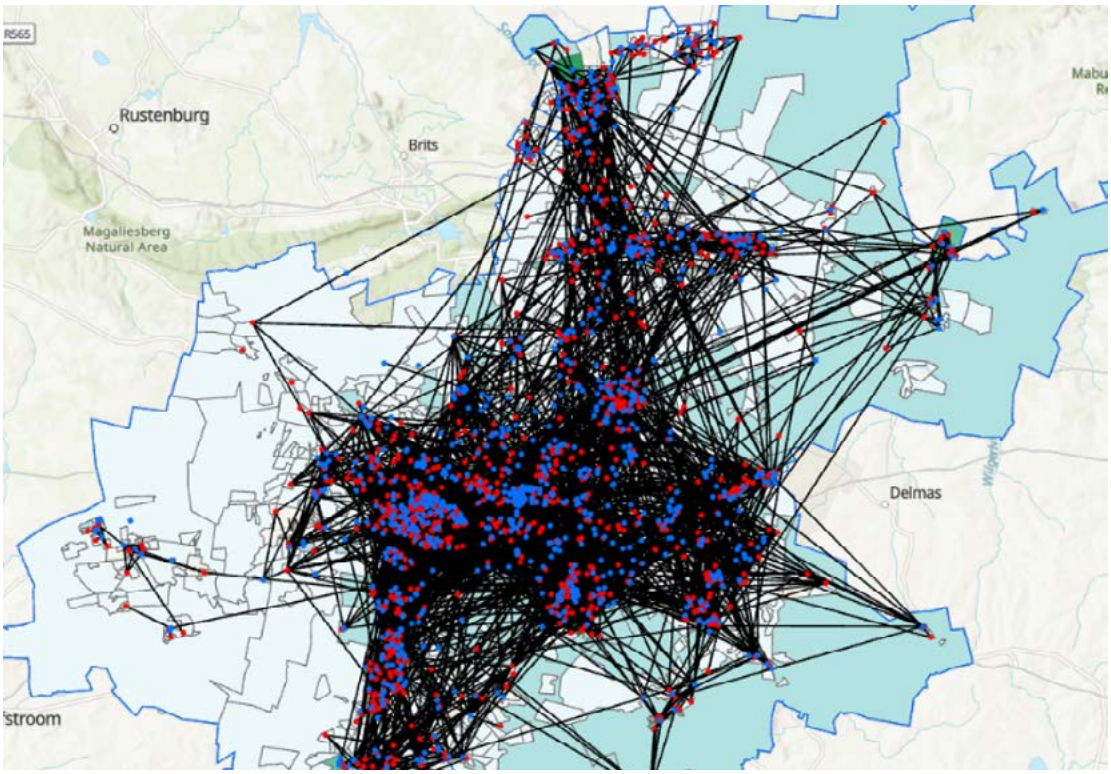
This is not an isolated case. Almost 92% of learners in Grade 8 are drawn from primary schools within 10 km of the receiving high school. Children enrolled in primary schools that are close to a good-quality high school therefore have a better chance of getting into such a school. It appears that increasing the feeder zone radius from 5 km to 30 km has had a limited effect on reversing some of the apartheid-era racial concentrations within schools.

Given that the distances between feeder primary schools and receiving high schools are quite small, and because good high schools are located in mostly white neighbourhoods, it is perhaps unsurprising that the racial composition of the best-performing high schools is largely white. Gauteng schools that performed nearer the bottom end of the NSC distribution are almost exclusively made up of black, coloured and Indian/Asian learners. In contrast, in performance decile 10, white and Indian/Asian students make up 60% and 10% of the student body, respectively.

4.3 School choice and language in South Africa

Another factor that also influences school choice is language. The SA-SAMS data allow one to investigate how language preference in terms of the home language subject that is offered in high school may affect school choice. Figure 20 shows the movements between primary and high schools for isiZulu-speaking learners – the numerically dominant group (see De Kadt et al., 2021) who had attended English LOLT primary schools.[3] The black lines show these learners' migration paths from primary to high school, while the red and blue dots represent primary (feeder) schools and high schools. The green shaded areas show the proportion isiZulu-speaking residents by ward (from Census 2011 data), with darker shades representing larger proportions isiZulu-speaking population. As the figure shows, most of the high schools these learners end up in are concentrated around Soweto, Katlehong, Johannesburg, Tembisa, and Garankuwa close to the Northwest province border.

Figure 20 Movements between primary and high schools for isiZulu-speaking learners in Gauteng from English LOLT primary schools, 2021/22



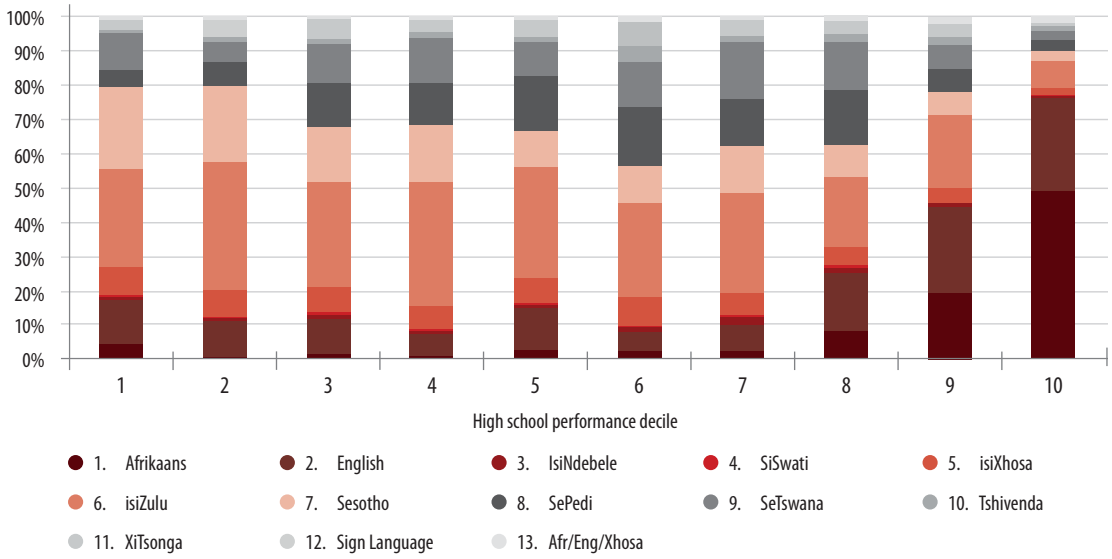
Source: Own calculations based on DDD data for 2021 and 2022 and geographic locations of schools from the Masterlist of Schools.

3 In the higher primary grades, all except Afrikaans school have English as LOLT.

The data show that many learners attend high schools that are far from their primary schools. This may be reflective of some school choice. On the other hand, the concentration in high schools in certain areas may suggest more limited choice, which may inter alia also reflect home language subject preferences.

Regardless of the reasons for choosing a school, the outcomes manifest themselves in relatively predictable patterns of access to high-quality education. Figure 21shows the percentages of Grade 8 learners speaking one of 13 home languages[4], by school performance decile (derived from the NSC 2019 results). IsiZulu, Sesotho and Sepedi home language learners dominate the bottom 80% of Gauteng schools in the NSC results, while the top 20% of schools by performance are dominated by Afrikaans and English speakers.

Figure 21 Distribution of home language by Gauteng schools' matric average decile, 2019



Source: Moses (forthcoming) based on SA-SAMS data (Data Driven Dashboards, 2022) and NSC 2019 data (Department of Basic Education, 2020).

4 These include the 11 official languages, plus 2 further categories for sign language and a mix of Afrikaans, English and Xhosa.

5 School-based assessments and examinations

5.1 Using school-based assessment data from DDD

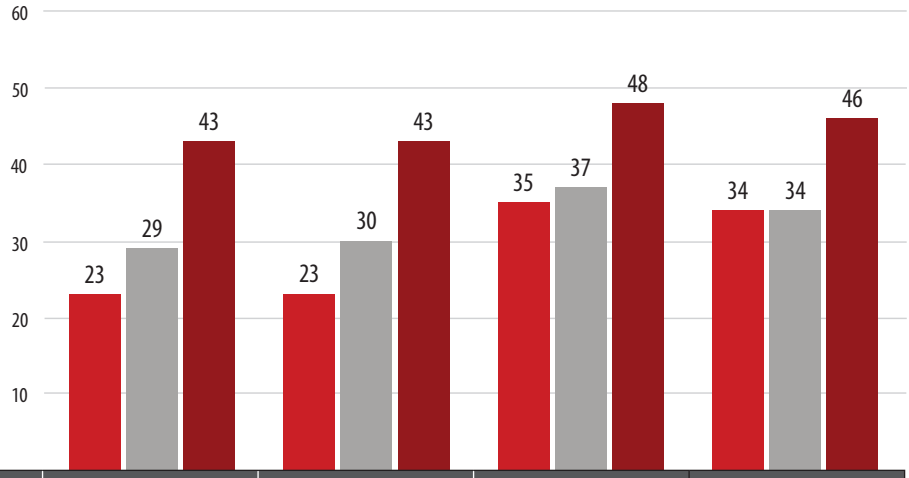
As shown in Section 2, one of the results of the COVID pandemic was that more learners were promoted to the next grade than before the pandemic. This could be because assessments were less challenging than in previous years due to the trimmed curriculum that covered less content, as teachers were told to focus assessments only on those parts of the work that have been covered. There were also changes to the weight given to end-of-year examinations versus term marks: While end-of-year examinations had counted 75% towards learners' final marks before 2020, this changed to only 40% during the pandemic period. An important question, therefore, is how SBA marks changed as a result of these changes to the curriculum and examination rules.

5.2 Patterns of school-based assessments in a sample of schools

School-level longitudinal data make it possible to investigate how SBA marks changed during the pandemic period. Using DDD data for the Eastern Cape, it was possible to calculate aggregate SBA marks at the school level for the period 2018 to 2021. This analysis for the Eastern Cape provides a picture of how COVID affected system performance.

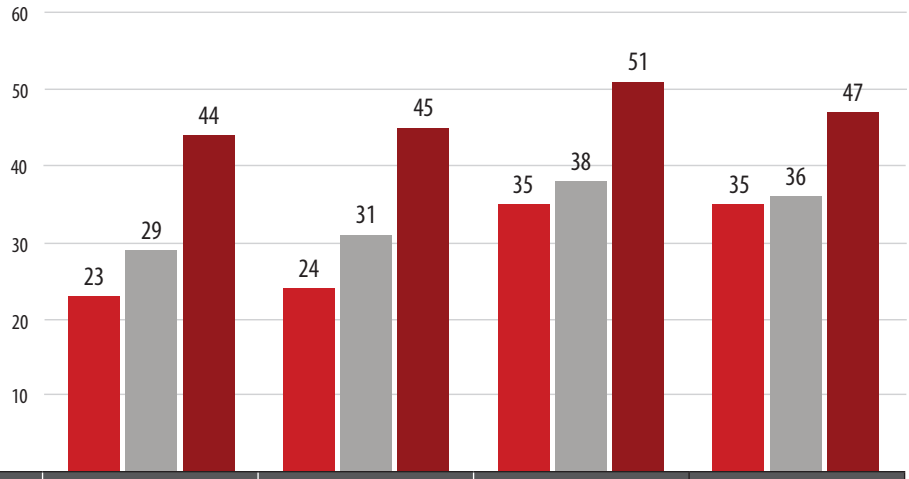
Figure 22 shows the average end-of-year Grade 10 SBA marks in Mathematics, Mathematical Literacy, and English as First Additional Language (EFAL) for the period 2018 to 2021. SBA marks changed little between 2018 and 2019, but in all cases, this was followed by a noteworthy increase in marks in 2020 and 2021 in all three subjects. The rise in Mathematics marks was particularly large, from an average of 23% in 2019 to 35% in 2020. A similar pattern is observed in Figure 23, which shows SBA marks for the same three subjects for Grade 11.

Figure 22 Eastern Cape Grade 10 SBA marks (in percentages) in three subjects, 2018–21



	2018	2019	2020	2021
Mathematics	23	23	35	34
Mathematics Literacy	29	30	37	34
EFAL	43	43	48	46

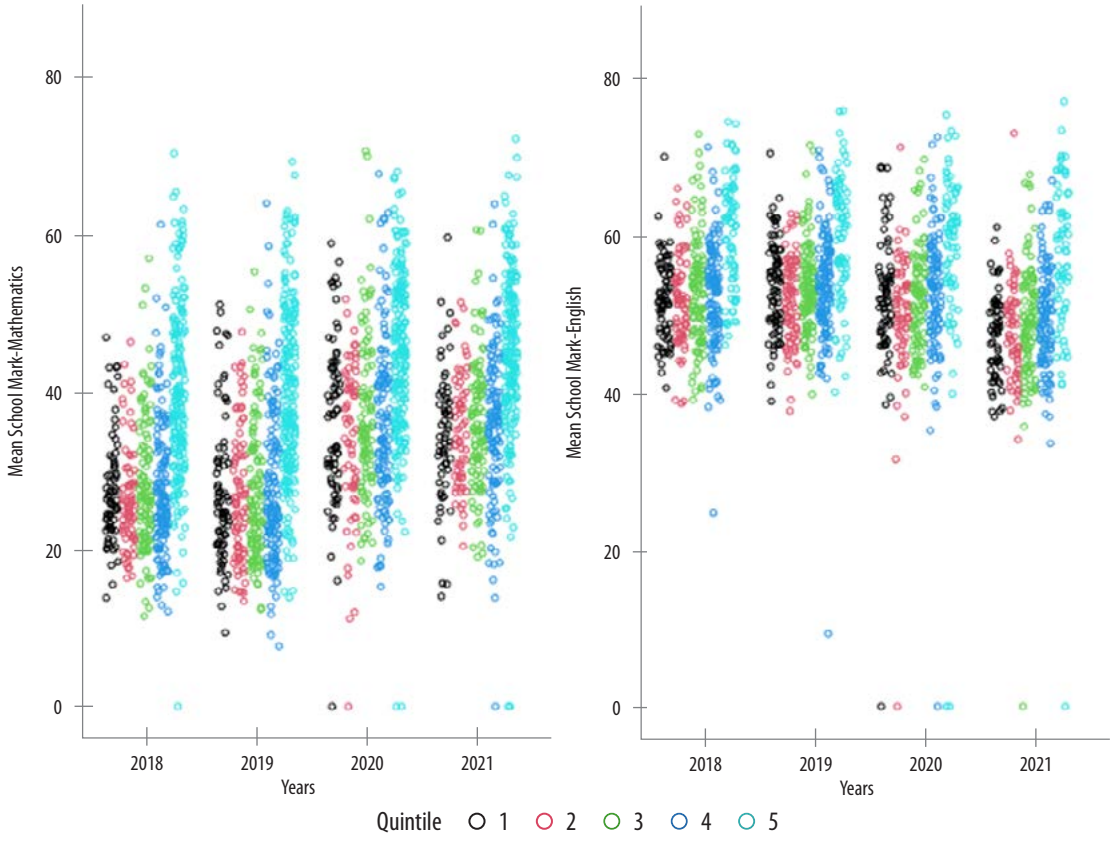
Figure 23 Eastern Cape Grade 11 SBA marks (in percentages) in three subjects, 2018–21



	2018	2019	2020	2021
Mathematics	23	24	35	35
Mathematics Literacy	29	31	38	36
EFAL	44	45	51	47

The increase in SBA marks during the pandemic is also shown graphically in Figure 24. The figure shows how SBA marks for Mathematics and EFAL were distributed across school quintiles in Gauteng for the period 2018 to 2021. Each dot shows the average mark for one school, and the quintiles are represented by different colours, arranged from the left (Quintile 1, the poorest schools) to the right (Quintile 5) for each subject and year. The first thing to note is that schools consistently performed much better in EFAL than in Mathematics. In EFAL, performance remained high but did not show clear change across the years. For Mathematics, this figure too shows relatively similar performance in 2018 and 2019, but performance increased markedly for all school quintiles in 2020 and 2021, as can be seen especially in the smaller frequency of weak performance in Mathematics. In both subjects, Quintile 5 schools seemed to show more outlier schools at the top end, reflecting better performance, even though SBA assessments are largely set at the school level and assessments in weaker schools are often less demanding.

Figure 24 Gauteng Grade 10 Mathematics and EFAL marks in a balanced sample of schools, 2018–21



The evidence presented here confirms that presented earlier in this report that SBA marks improved during the pandemic period, while repetition declined. Given previous evidence of large learning losses occurring during the pandemic years (Ardington, Wills and Kotze, 2021; Van der Berg *et al*., 2022), this means that marks and grade promotion rates increased at a time that major learning losses were occurring. That is, despite large learning losses as a result of lost learning time, learners were assessed more leniently and fewer learners had to repeat. This increased leniency in assessment as a result of the pandemic has serious

implications for learner preparedness to take on the curriculum in the new grades they are promoted to and is also likely to increase within-class heterogeneity. This is especially concerning when one considers that varying levels of learner preparedness within the same classroom were already a problem before the pandemic (Wills and Hofmeyr, 2019).

5.3 How can school-based assessment be used to advise learners about Mathematics subject choice in Grade 10?

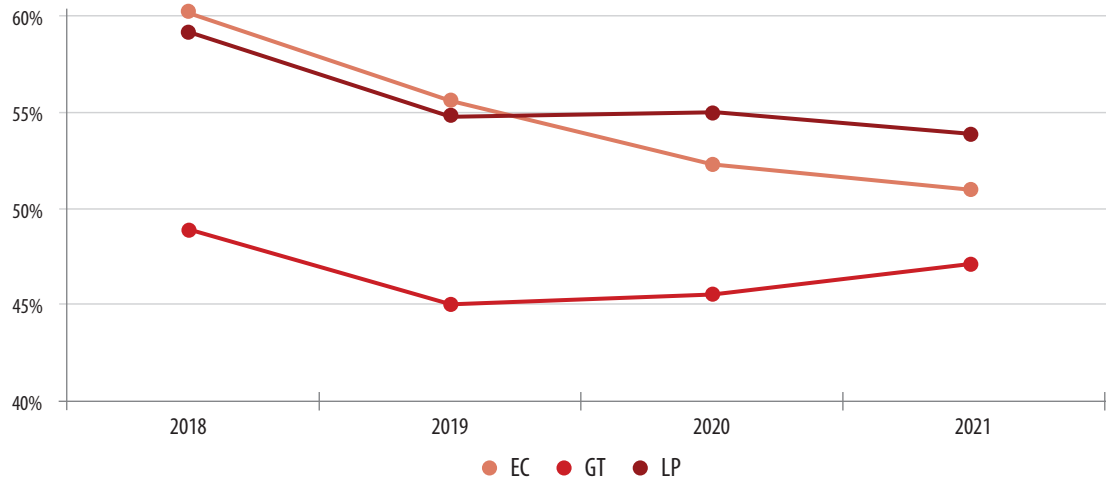
The DDD data, which contain information about individual learners over time, can be used to investigate patterns in achievement in certain key subjects, such as Mathematics, and how this achievement is related to later performance.

An important decision faced by all Grade 10 learners is whether to take Mathematics or Mathematical Literacy in the FET phase. This choice may have important implications for their future, for instance for options for later studies at universities, as well as impacting their flows through the school system and their probability of passing matric. Because of the importance of this choice, it is instructive to consider which learners choose Mathematics and how this choice affects later achievement.

To do so, the DDD data from Gauteng, Limpopo and the Eastern Cape were used to model the decision to take Mathematics. The regression results can be found in Appendix Table A1. The first question considered was how predictive Grade 9 SBA marks in Mathematics are of the choice to take Mathematics in Grade 10. As expected, the results indicate that controlling for other factors, better achievement in Grade 9 Mathematics is associated with a higher probability of taking Mathematics in Grade 10. The regression results also show that girls and learners who are in the appropriate age-for-grade are more likely to take Mathematics, controlling for other factors.

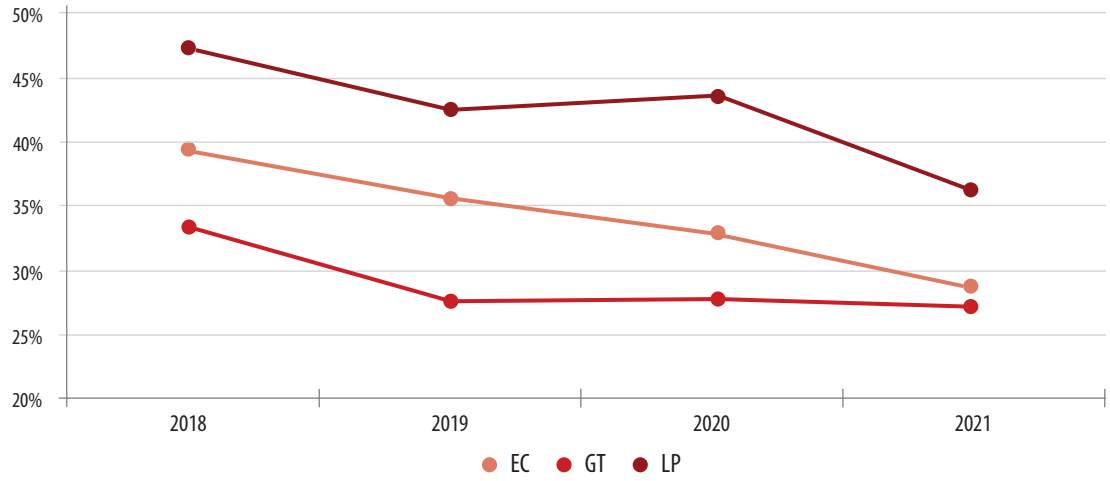
Given that the three provinces considered have DDD data spanning several years, it is also possible to use the data to investigate trends in the decision to take Mathematics over time. Figure 25 shows the percentage of new Grade 10 learners taking Mathematics each year for the period 2018 to 2021, and points to a decrease in the proportion of learners taking Mathematics in all three provinces over this period. For example, while 60% of Grade 10 learners in the Eastern Cape chose Mathematics in 2018, this proportion was only 51% in 2021. A similar trend is observed in the other two provinces. Interestingly, while Gauteng saw a slight increase in the percentage of learners taking Mathematics after 2019, Limpopo and the Eastern Cape saw a slight decline in the proportion of learners taking Mathematics during this period.

Figure 25 Percentage of new Grade 10 learners in three provinces electing to take Mathematics, 2018–21



An interesting observation that emerges from interrogating the DDD in this way is that many learners who perform relatively poorly in Mathematics in Grade 9 nonetheless go on to choose Mathematics in Grade 10. This can be seen in Figure 26, which shows the proportion of low-achieving learners[5] choosing Mathematics in Grade 10. The figure shows that in Limpopo, especially, a large proportion of weak achievers (40% or less) in Grade 9 Mathematics decided to take Mathematics in Grade 10 – between 36% (2021) and 47% (2018).

Figure 26 Percentage of low-achieving Grade 9 learners electing to take Mathematics in Grade 10 in three provinces, 2018–21

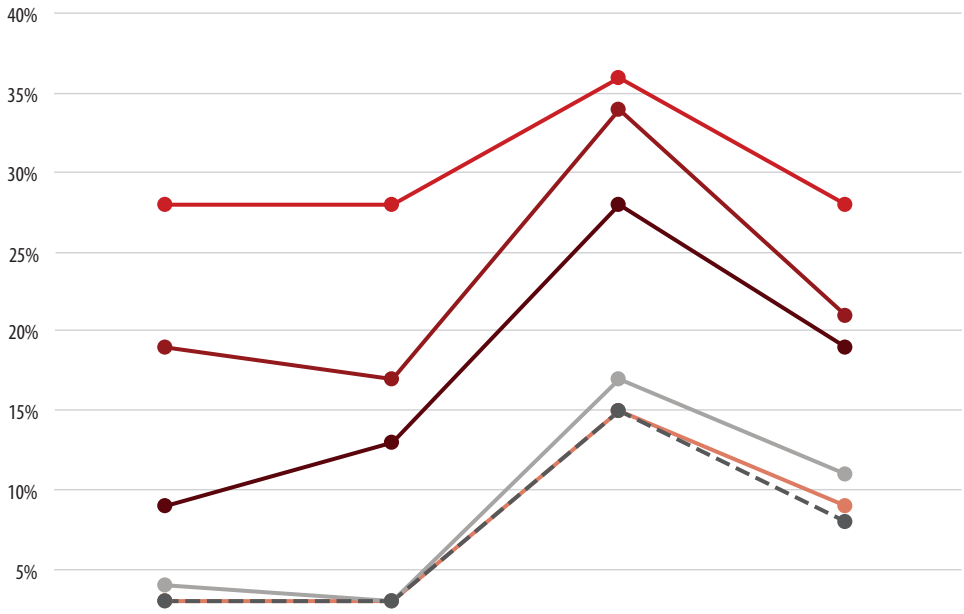


Note: 'Low achieving' is here defined as a mark of 40% or less in Grade 9 Mathematics

5 Low-achieving learners are defined as those who achieved 40% or less in Mathematics in Grade 9.

The learner-level DDD data make it possible to determine what happened to these learners over time: Figure 27 shows that low-achieving learners who chose Mathematical Literacy (the solid lines in the figure) in Grade 10 performed much better than their peers who chose Mathematics (the dotted lines in the figure) in all three provinces. Considering individual years, it is clear that low-achieving learners performed better in Grade 10 Mathematics during the pandemic than before. For example, while only 3% of low-achieving learners in Grade 9 went on to achieve 40% or more in Grade 10 Mathematics in 2018 and 2019, this proportion rose to 15% and 8% in 2020 and 2021, respectively.

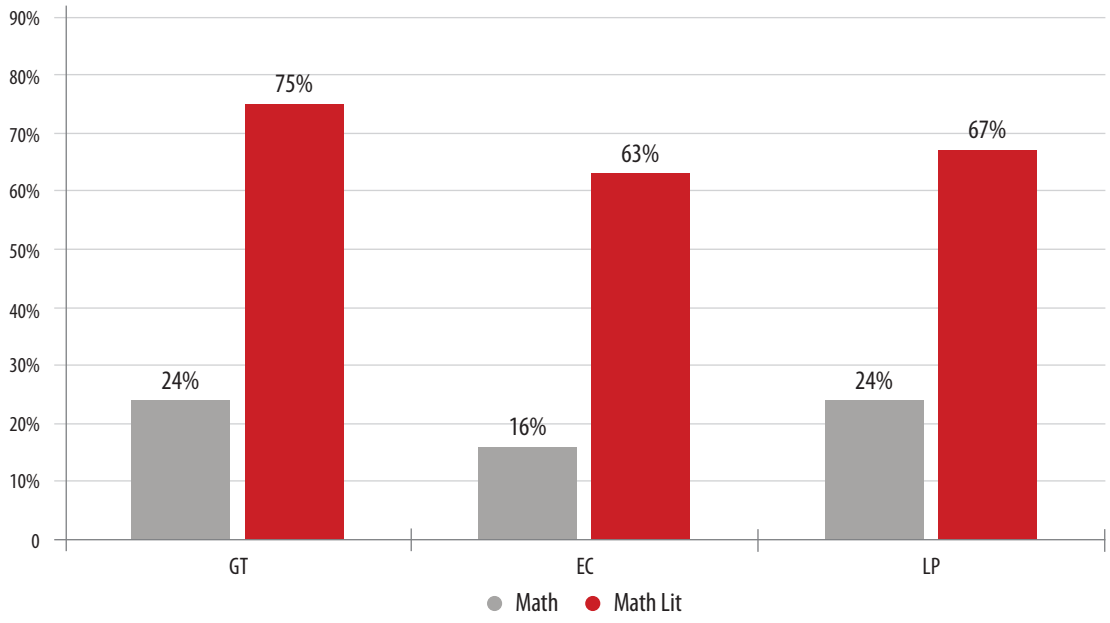
Figure 27 Percentage of low-achieving Grade 9 learners in three provinces achieving more than 40% in Mathematics or Mathematical Literacy in Grade 10, 2018–21



	2018	2019	2020	2021
GT Math	3%	3%	15%	8%
EC Math	4%	3%	17%	11%
LP Math	3%	3%	15%	9%
GT Mlit	28%	28%	36%	28%
EC Mlit	19%	17%	34%	21%
LP Mlit	9%	13%	28%	19%

This evidence indicates that the choice of Mathematics or Mathematical Literacy has important implications for later academic performance. We now consider the *NSC achievement* of learners who took Mathematics versus those who took Mathematical Literacy. This analysis is possible due to the longitudinal nature of the data, which allows one to track learners from Grade 9 to Grade 12, and to compare their achievement on the NSC examinations to their Grade 9 Mathematics achievement. It must be noted, however, that due to the limited number of years for which this data are available, this can only be done for the 2017 Grade 9 cohort, and only for the learners who reached Grade 12 in 2020, implying that these learners did not repeat or drop out between 2017 and 2020. In this sense, the sample considered is a selected sample of learners who progressed uninterrupted from Grade 9 in 2017 to Grade 12 in 2020. Figure 28 shows the NSC performance of low-achieving Grade 9 learners and indicates that of these learners, those who chose Mathematical Literacy were much more likely to achieve a mark of more than 40% in Mathematical Literacy in the NSC examinations. For example, in Gauteng, only 24% of low-achieving Grade 9 learners went on to achieve an average of over 40% in the NSC Mathematics examinations, compared with 75% of learners who chose Mathematical Literacy. Detailed results of this analysis are presented in Appendix Table A2.

Figure 28 Percentage of low achieving 2017 Grade 9 learners from three provinces reaching matric in 2020 and scoring more than 40% in Mathematics or Mathematical Literacy in the NSC examinations



LEARNERS PERFORMING POORLY IN GRADE 9 MATHEMATICS WOULD ESPECIALLY FACE MUCH BETTER PROSPECTS, IN TERMS OF BOTH REACHING AND PASSING MATRIC, IF THEY CHOSE MATHEMATICAL LITERACY INSTEAD OF MATHEMATICS.

Aside from tracking learners from Grade 9 to matric and observing their performance in the final examinations, it is possible to use regression analysis to examine how subject choice (between Mathematics and Mathematical Literacy) is related to learners' progression through school from Grade 9 to 12 and whether they pass matric once reaching it (regardless of whether they pass Mathematics or Mathematical Literacy). This could only be done for the 2017 Grade 9 cohort, as matric results linked to earlier SBA performance were available only up to 2020. The regression results reported in Appendix Table A2 show that in all three provinces, choosing Mathematics in Grade 10 is associated with a decrease in the probability of both reaching matric without interruption and passing matric. This result, considered in conjunction with the preceding results presented in this section, suggests that many learners (at least in these three provinces) may have performed better in matric, had they chosen Mathematical Literacy instead of Mathematics in Grade 10. Learners performing poorly in Grade 9 Mathematics would especially face much better prospects, in terms of both reaching and passing matric, if they chose Mathematical Literacy instead of Mathematics.

Another topic for consideration regarding the choice between Mathematics and Mathematical Literacy is how many learners switch between the two subjects over time. Anecdotally, some believe that initially choosing Mathematics in Grade 10 and then later switching to Mathematical Literacy will result in improved Mathematical Literacy achievement on the NSC examinations compared to what would have been achieved if a learner had chosen Mathematical Literacy from Grade 10. While there appears to be some superficial empirical support for such a view, further investigation is needed to evaluate if that is indeed the case and if any benefit (should it exist) outweighs negative effects such as the reduced likelihood of reaching matric without repeating or dropping out.

The NSC data can also be used to investigate how the choice of Mathematics is linked to choosing Physical Science in Grade 12. The majority of learners who wrote Mathematics wrote Physical Science too, although a smaller number of learners did combine Physical Science and Mathematical Literacy. Multivariate regression analysis shows a strong correlation between achievement in these two subjects, indicating that a large part of achievement in Physical Science is statistically 'explained' by Mathematics performance.

5.4 How do school-based assessments affect repetition?

To confirm previous findings regarding the associations between SBA and repetition, Grade 10 repetition at the end of 2019 was studied using regression analysis. The regression results found in Appendix Table A3 point to the expected association between SBA marks and repetition, whereby Grade 10 learners with higher achievement in Mathematics (or Mathematical Literacy) and English First Additional Language (EFAL) in Term 4 are less likely to repeat Grade 10. Interestingly, the analysis also shows that female learners who chose Mathematical Literacy were less likely to repeat Grade 10 than their male counterparts, while over-age learners were more likely to repeat regardless of subject choice.

The regression analysis also shows that learners who had already repeated Grade 10 were much less likely to repeat Grade 10 again. This provides some evidence of the repetition policy being enforced, whereby learners are not allowed to repeat more than once per phase (Department of Basic Education, 2013). However, there are nevertheless still large numbers of learners who repeat Grade 10 more than once, according to the data. Interestingly, the regression results further show that attending a Quintile 3–5 school is positively associated with repetition (compared to a Quintile 1 or 2 school), indicating that for a given level of achievement, learners are more likely to repeat Grade 10 if they are in higher quintile (wealthier) schools. Another way of considering this may be that good performance in the poorest schools is more likely to be sustained.

INTERESTINGLY, THE ANALYSIS ALSO SHOWS THAT FEMALE LEARNERS WHO CHOSE MATHEMATICAL LITERACY WERE LESS LIKELY TO REPEAT GRADE 10 THAN THEIR MALE COUNTERPARTS, WHILE OVER-AGE LEARNERS WERE MORE LIKELY TO REPEAT REGARDLESS OF SUBJECT CHOICE.

6 Conclusion

The aims of the analyses presented in this report were twofold: (i) To show the impact of COVID-19 on key indicators of education system performance, such as learner flows through the system and performance in school-based assessments; and (ii) To show the kind of analysis that becomes possible with the increasing amount of high-quality administrative data that is becoming available in South Africa.

In addressing the first aim, we showed how COVID-19 caused severe disruptions to the system, which can be seen in discontinuities in several key indicators of education system performance. Although the COVID-19 pandemic gave rise to severe learning losses, learner flows through the system accelerated and learner marks improved. This confirms that the pandemic period gave rise to much greater leniency in assessment, as DBE policy indeed intended, as discussed earlier. As a consequence, the COVID-19 pandemic may also endanger the integrity of the NSC examinations, given that many more learners passed matric and achieved bachelor passes in 2020 and 2021, despite the lost learning time and resultant learning losses that occurred.

To address the second aim, we showed how new data on teachers, school location and subject choice make it possible to highlight important dynamics in the system,. This further strengthens the case for enhanced efforts to improve the availability and quality of administrative data for policy-making and administration in South Africa's school system.

7 References

Ardington, C., Wills, G. and Kotze, J. (2021) 'COVID-19 learning losses: Early grade reading in South Africa', *International Journal of Educational Development*. Elsevier Ltd, 86(August), p. 102480. doi: 10.1016/j.ijedudev.2021.102480.

Department of Basic Education (2022) Masterlist of Schools for 2021, Retrieved from: https://www.education.gov.za/Programmes/EMIS/EMISDownloads.aspx, December 2022

Department of Basic Education (2013) National policy pertaining to the programme and promotion requirements of the National Curriculum Statement, grades R–12, *Government Gazette*, p. 74.

De Kadt, J., Hamann, C., Mkhize, S.P, & Parker, A. (Eds.).(2021). *Quality of Life Survey 6 (2020/21): Overview Report*. Johannesburg: Gauteng City-Region Observatory (GCRO).

Fleisch, B. (2008). *Primary education in crisis: Why South African schoolchildren underachieve in reading and mathematics*. Johannesburg. Juta.

Hoadley, U. (2020) 'Schools in the Time of COVID-19: Impacts of the Pandemic on Curriculum', *Resep Non-Economic Working Paper*, (November), p. 29. Available at: https://www.researchgate.net/publication/345310078_Schools_in_the_Time_of_COVID-19_Impacts_of_the_Pandemic_on_Curriculum.

Spaull, N. (2013) 'Poverty & privilege: Primary school inequality in South Africa', *International Journal of Educational Development* 33(5). doi: 10.1016/j.ijedudev.2012.09.009.

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*. Resep Report. Stellenbosch. https://www.ssrn.com/abstract=4009654

Van der Berg, S., Hoadley, U., Galant, J., Van Wyk, C., & Böhmer, B. (2022). *Learning Losses From COVID-19 in the Western Cape: Evidence from Systemic Tests*. Resep. Stellenbosch. https://www.ssrn.com/abstract=4212977

Wills, G. and Hofmeyr, H. (2019) 'Academic resilience in challenging contexts: Evidence from township and rural primary schools in South Africa', *International Journal of Educational Research*, 98(June), 192–205. doi: 10.1016/j.ijer.2019.08.001.

8 Appendix A

Table A1 Modelling the predictors of performance in Grade 10 Mathematics in three provinces in 2018, 2020 and 2021

	(1)	(2)	(3)
	Gr10 math 2018	**Gr10 math 2020**	**Gr10 math 2021**
Province (base: Gauteng)			
Eastern Cape	0.043***	0.017***	0.032***
	(0.003)	(0.002)	(0.002)
Limpopo	0.068***	0.106***	0.081***
	(0.003)	(0.002)	(0.002)
Grade 9 Term 4 Math mark	0.013***	0.013***	0.011***
	(0.000)	(0.000)	(0.000)
Gr10 quintile (base: Q1)			
Q2	−0.030***	−0.014***	−0.023***
	(0.003)	(0.003)	(0.003)
Q3	−0.075***	−0.062***	−0.065***
	(0.003)	(0.003)	(0.002)
Q4	−0.055***	−0.032***	−0.010***
	(0.004)	(0.004)	(0.003)
Q5	−0.107***	-−0.045***	0.029***
	(0.003)	(0.003)	(0.003)
Missing Quintile	−0.030***	0.001	0.050***
	(0.005)	(0.004)	(0.004)
Female	0.021***	0.012***	0.018***
	(0.002)	(0.002)	(0.002)
Grade 9 age (base: correct age)			
Over-age	−0.145***	−0.133***	−0.153***
	(0.002)	(0.002)	(0.002)
Underage	0.050***	0.049***	0.051***
	(0.002)	(0.002)	(0.002)
Constant	0.107***	0.058***	0.045***
	(0.004)	(0.003)	(0.003)
Observations	237 271	271 569	311 970
R-squared	0.217	0.235	0.200

Notes: The outcome variable of interest is performance in Mathematics in Grade 10. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table A2 Modelling the probability of the 2017 Grade 9 cohort both progressing to matric without interruption and passing matric in Gauteng, Eastern Cape and Limpopo

	(1)	(2)	(3)
	Gauteng	**Eastern Cape**	**Limpopo**
Took Gr10 math	−0.034***	−0.029***	−0.009***
	(0.004)	(0.003)	(0.003)
Gr9 math mark	0.008***	0.007***	0.009***
	(0.000)	(0.000)	(0.000)
Female	0.048***	0.005*	−0.001
	(0.004)	(0.003)	(0.003)
Age: (base: correct age)			
Over-age	−0.180***	−0.198***	−0.183***
	(0.004)	(0.004)	(0.003)
Underage	0.039***	0.091***	0.074***
	(0.005)	(0.005)	(0.004)
Gr9 quintile (base: Q1)			
Q2	−0.002	0.045***	−0.002
	(0.007)	(0.004)	(0.004)
Q3	0.000	0.092***	0.017***
	(0.007)	(0.004)	(0.004)
Q4	−0.004	0.383***	0.142***
	(0.007)	(0.009)	(0.011)
Q5	0.210***	0.523***	0.306***
	(0.007)	(0.007)	(0.012)
Missing quintile	0.005	0.229***	0.121***
	(0.009)	(0.010)	(0.011)
Constant	0.106***	0.037***	−0.102***
	(0.010)	(0.007)	(0.038)
Controls	Home Language	Home Language	Home Language
Observations	59 609	76 467	76 854
R-squared	0.218	0.226	0.216

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A3 Modelling the probability of grade repetition among learners taking Mathematics (Model 1) and learners taking Mathematical Literacy (Model 2) in three provinces, 2019

	(1)	(2)
	Math2019	**MLit2019**
Math/Mlit mark	−0.009***	−0.014***
	(0.000)	(0.000)
EFAL mark	−0.013***	−0.013***
	(0.000)	(0.000)
Female	−0.002	−0.025***
	(0.003)	(0.003)
Repeated prev year	−0.285***	−0.277***
	(0.006)	(0.005)
Years over-age (base:0)		
1 year	0.040***	0.042***
	(0.004)	(0.003)
2 years	0.073***	0.055***
	(0.006)	(0.004)
3 or more years	0.088***	0.051***
	(0.007)	(0.005)
Quintile (base: Q1)		
Q2	0.023**	0.019*
	(0.012)	(0.010)
Q3	0.056***	0.028***
	(0.012)	(0.010)
Q4	0.060***	0.046***
	(0.017)	(0.015)
Q5	0.069***	0.064**
	(0.026)	(0.025)
Missing quintile	0.064***	0.016
	(0.021)	(0.023)
Province (base: GT)		
EC	0.001	−0.037***
	(0.018)	(0.014)
LP	0.055***	0.081***
	(0.014)	(0.012)
Constant	1.139***	1.394***
	(0.024)	(0.018)
Controls	Home Language	Home Language
Observations	144 294	162 769
R-squared	0.336	0.422

Notes: The outcome variable of interest is a dummy variable indicating whether a learner repeated. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

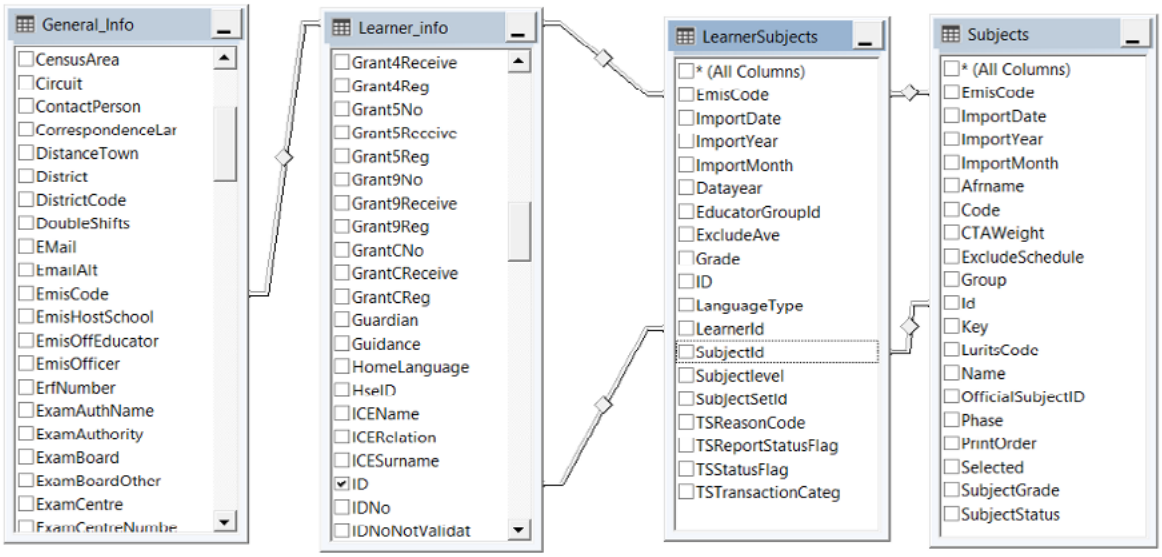
9 Appendix B

The challenges of analysing learner-level data

Various major challenges occur when initially dealing with big data. The volume of this learner unit record data is the main challenge in managing a dataset of this size. A further challenge is the speed of processing and managing the data. Big data also increase the difficulty of data integration. Data integration in this context refers to linking the data from different years to follow individual learners over time. Therefore, getting the data in a format and structure that is manageable remains one of the biggest challenges when analysing learner record administrative data.

It was possible to format and analyse the data because of the availability of a detailed entity relationship diagram (ERD) which we obtained from the database developers. An ERD is a type of flow chart that provides a snapshot of how the entities (schools, learners, subjects, etc) relate to each other. It is the blueprint that gives a visual representation of the relationships between the different sets of data (entities). An ERD is essential for modelling the data stored in a database. It is the basic design upon which a database is built and should therefore be easily available and accessible to the data users. The ERDs in Figure B1 provides an example of an ERD and how to link a learner to their school and their subjects in SA-SAMS database systems.

Figure B1 Entity relationship diagram for SA-SAMS



RESEP
Research on Socio-Economic Policy

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Michael & Susan Dell
FOUNDATION

We gratefully acknowledge support from the
Michael & Susan Dell Foundation