Learner flows through schools:

Using high quality administrative data to understand education system performance

Servaas van der Berg
Chris van Wyk
Rebecca Selkirk
Heleen Hofmeyr

August 2021

RESEP
Research on Socio-Economic Policy

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918 · 2018

Learner flows through schools:

Using high quality administrative data to understand education system performance

Servaas van der Berg, Chris van Wyk, Rebecca Selkirk & Heleen Hofmeyr

1. Introduction

South Africa has, like many countries in the world, invested large sums of funds to develop and maintain longitudinal data systems for the education sector. There are a number of reasons why countries invest in the development of such data systems, ranging from purely administrative purposes (for example, budgeting for education expenditure) to supporting evidence-based policy-making (Figlio, Karbownik and Salvanes, 2015). The availability of national administrative education data also creates opportunities for expanding our knowledge of the functioning of national education departments. For example, enrolment data can provide useful information about how learners progress through the system, which provides a good indication of how successful the system is at developing in learners the capabilities and skills required to be successfully promoted to the next grade, and ultimately complete grade 12 (matric) and obtain a National Senior Certificate qualification.

This report presents a picture of the kind of data analysis that is possible with high quality education data that is increasingly becoming available through initiatives undertaken to improve South Africa's major school administrative data system, known as SA-SAMS (the South African School Administration and Management System). The results presented in this report illustrate the crucial importance of high-quality administrative data for informing education policy and practice, but it also highlights the importance of further steps to improve the collection, quality and analysis of such data. This report is largely based on new analysis, but also draws from previous work based on analysis of some of the data that will be discussed.

This report is structured as follows: Section 2 explains the background to and motivation for the study. Information Box A provides details of the different datasets used to inform the analysis. Sections 3, 5, and 5 show how national learner-level administrative data can be used to understand national patterns of enrolment, repetition, and dropout, respectively. Section 6 makes use of longitudinal data from the Western Cape over a thirteen-year period to show how such longitudinal data can be used to provide a more detailed picture of learner flows through the schooling system that lie beneath the patterns observed in the national data. Results regarding the association between marks from school-based assessment (SBA)[1] and grade progression, repetition, and dropout in Limpopo are presented in Section 7. Section 8 presents a summary of the main findings of the report, and discusses potential policy implications. Section 9 concludes.

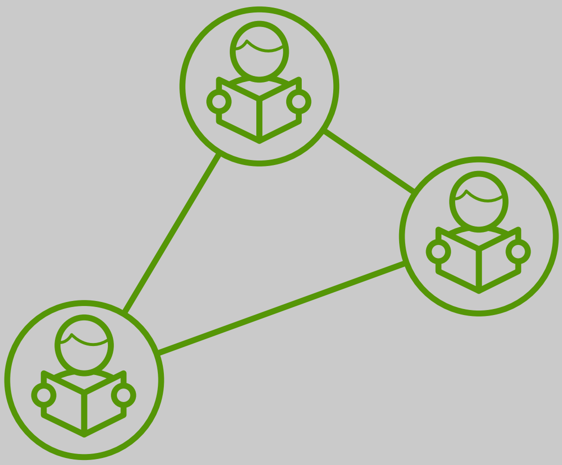
1 Tests written within individual schools and not forming part of standardised assessments.

2. Background and motivation

Since 2013, the New Leaders Foundation, a local not-for-profit company in collaboration with the Michael and Susan Dell Foundation, has been working with the Department of Basic Education (DBE) in seven provinces to improve quality of the SA-SAMS data system. Part of this effort has been the creation of data dashboards (based on the SA-SAMS data) for each school that can be used for management purposes by schools, districts and other provincial and national education authorities. This has come to be referred to as the Data Driven Districts (DDD) initiative. This collection process has improved over time, with the result that data from about 85% of all schools is now collected. To investigate the possible usefulness of the underlying data that had been collected for policy purposes, Resep (Research on Socio-Economic Policy) at the University of Stellenbosch analysed this data on behalf of the Michael and Susan Dell Foundation. Resep has produced three papers (Van der Berg, Van Wyk, Selkirk, Rich & Deghaye, 2019; Van der Berg, Wills, ~~et al,~~ 2019; Van Wyk, 2021) that analyse the SA-SAMS (and the Central Education Management Information System (Cemis) in the Western Cape) data in line with this goal.

This report, also supported by the MSDF, has been compiled with the intention of drawing from the strength of available datasets to present a picture of learner performance and flows through the South African education system. The analysis presented in this paper makes two main contributions to our understanding of the efficiency of the country's education system: Firstly, by triangulating across the different available administrative education datasets in the country, we are able to synthesise the information regarding learner flows in one place, thus allowing us to present a picture of what kind of analysis is possible with high-quality administrative education data. Secondly, we present new evidence of the links between performance in school-based assessment (SBA) and learner flows, which allows us to investigate what lies beneath the grade progression, repetition, and dropout patterns that we observe in the administrative data, in terms of performance in school assessments[2]. This, too, is intended to illustrate how administrative data can inform our understanding of learner performance on a continuous basis. Ultimately, this report shows high-quality and timely administrative data can inform education policy and practice at all levels, from the school to the national department.

Triangulating across administrative **datasets allows us to synthesise information** on learner flows



2 A previous report for the MSDF (Van der Berg, Van Wyk, et al., 2019) also investigated the association between SBA marks and learner flows through the schooling system. The analysis in this report builds on this existing evidence by investigating this association for a much larger sample of schools in Limpopo than was possible in the previous report

Information Box A: Data description and limitations

We were able to obtain three datasets that contain learner-level data, namely DDD and Learner Unit Record and Tracking System (Lurits) data, which are housed within the SA-SAMS data repository, and Cemis data, which is housed within the Western Cape's data repository. To comply with the Protection of Personal Information Act (POPIA) we obtained the data in an anonymised format (no personal information, e.g. names, surnames, etc.). These datasets contain longitudinal data that allows for tracking individual learners through the education system from year to year.

DDD data was available for the period 2015 to 2020 for one province (Limpopo). The major advantage of this dataset is that it contains learners' marks in the SBA's. Building on previous work on the association between SBA marks and learner flows among samples of schools in Limpopo, Gauteng, and the Eastern Cape, we present further evidence of this association, this time across multiple years[3]. Unfortunately, there are some inconsistencies in this DDD data that limit how the data can be analysed. Specifically, not all the schools in Limpopo that currently submit SA-SAMS data have been doing so since 2015. To make the most of the longitudinal aspect of the data, it is best to work with the same schools over time. To remedy this problem, we used the same schools that submitted every year from 2015 to 2020. The result is that we could not include all public schools in Limpopo in the analysis. Furthermore, the cohort we followed from 2015 is much smaller than if we followed a cohort beginning from 2016, but then the period will be shorter (only 4 years). Restricting our analysis of Limpopo's DDD data in this way naturally introduces a selection effect whereby only schools that were meticulous in submitting their data in both 2015 and 2020 are included.

The Cemis data has the advantage that it has been in existence long enough (2007 to 2019) to follow individual learners through the entire education cycle, thus allowing a detailed picture of learner flows through the system. However, Cemis is limited to the Western Cape, which differs from other provinces in a number of important respects – for example, the Western Cape has fewer poor schools and outperforms the other provinces in almost all the international educational assessments South Africa participates in (Howie, Combrinck, Tshele, Roux, McLeod Palane & Mokoena, 2016; Reddy, Winnaar, Juan, Arends, Harvey, Hannan, Namome, Sekhejane & Zulu, 2019) and there is evidence to suggest that the province has better functioning administration than most other provincial education departments (Wills, Shepherd & Kotze, 2016). These factors limit the extent to which trends in learner flows observed for the Western Cape can be considered indicative of trends in the rest of the country.

Lurits data is available for the period 2017 to 2019. To comply with POPIA, the data was made available in an anonymised format (i.e. no learner names or surnames). As part of a different research project aimed at understanding how reliable the Lurits data is at the learner level, the DBE created unique learner identification numbers in anonymised format which were used for the analysis presented in this report. While attempts were made throughout the report to conduct analysis at the learner level as far as possible, the aggregated Lurits data (at the school level) seemed to be of better quality than unit-level (learner-level) data. One major source of inconsistency in the Lurits data (as was the case with Limpopo's DDD data) was that the same number of schools did not submit their data in all years. For example, 22 782 public schools submitted their data in 2017, while 22 899 schools submitted in 2018, and 23 069 schools submitted in 2019. This poses a challenge for analysing trends in learner enrolment between years. However, a comparison of learner enrolment between years shows that the quality of the data is of an acceptable standard, especially in 2018 and 2019. Figure 1 and Figure 2 show enrolment per school in 2017 versus 2018 and 2018 versus 2019, respectively, captured in the Lurits data. The graphs show that enrolment numbers are relatively consistent between these years, especially between 2018 and 2019, indicating that the data is of acceptable quality to make meaningful comparisons between the two years. Therefore analysis of Lurits data over time is limited to the period 2018 to 2019 throughout this report.

3 Resep's previous analysis of the association between SBA marks and learner flows, although conducted in three provinces, was limited to a shorter timespan as a result of data quality concerns. For this report, we were able to obtain more reliable DDD data from Limpopo with an additional year of data included (2015 to 2019), which makes it possible to analyse the association between SBA marks and learner flows for a longer period.

Figure 1: Comparison of 2017 versus 2018, Lurits enrolment data by school

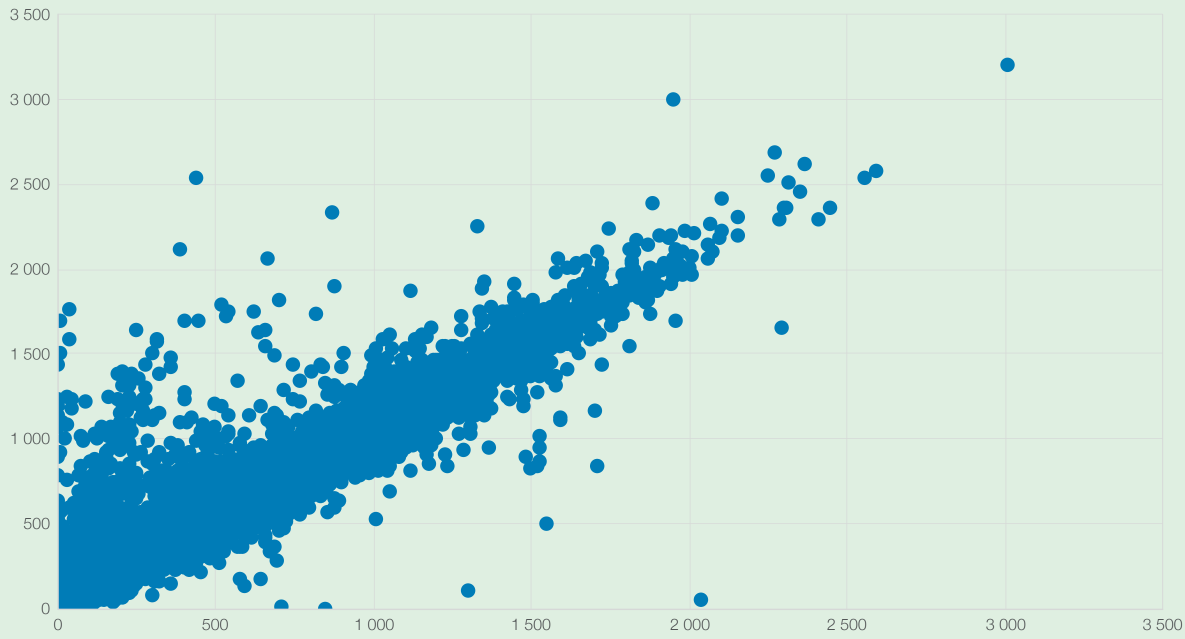
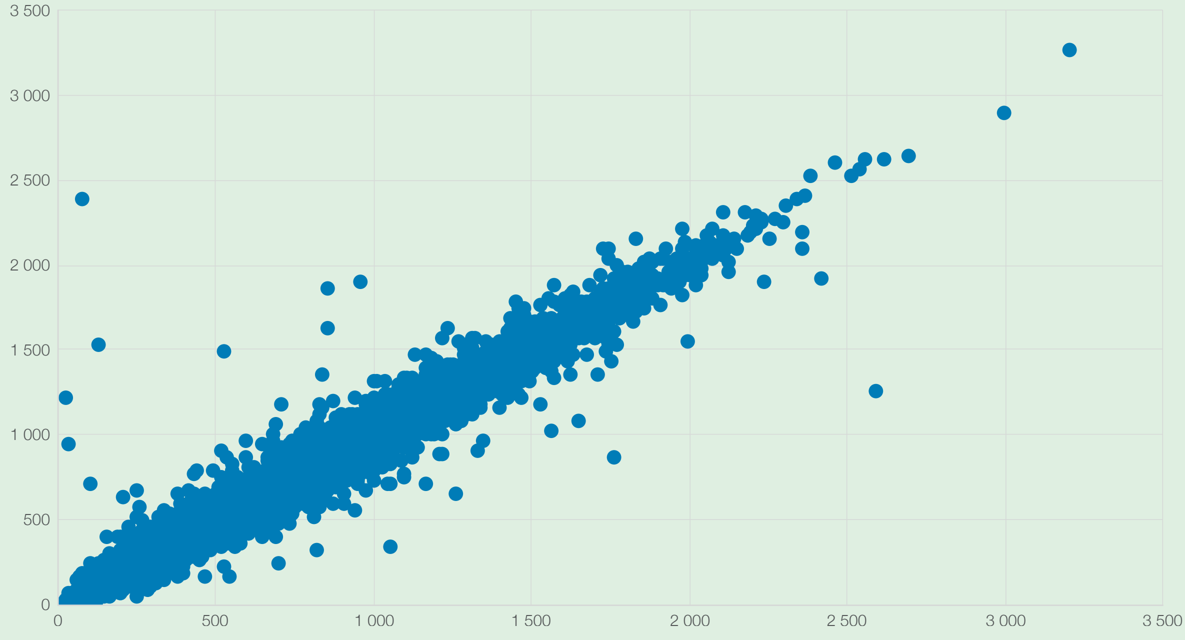


Figure 2: Comparison of 2018 versus 2019, Lurits enrolment data by school



Another challenge for analysing learner enrolment patterns in the Lurits data is that learners seem to have been allocated a new anonymous unique identification variable when they changed schools. This is likely a result of how the SA-SAMS data is collected, which follows the principle that a learner is only uniquely identified within a school. Once learners change schools, they are assigned a new unique identifier in the SA-SAMS data. The Lurits data is supposed to remedy this problem by assigning each learner a unique identification variable that uniquely identifies them regardless of which school they are enrolled in. Unfortunately, the anonymised Lurits data does not seem to have remedied this problem – learners still appear to have been assigned new unique identifiers when they entered a new school. That is, learners who changed schools were recorded as dropouts and new “drop-ins” into the system.

This phenomenon is most pronounced between Grades 7 and 8 when learners leave primary school and enter secondary school. As a result, it is not possible to track learners from primary school into secondary school. In addition, it is difficult to accurately calculate the number of learners that repeated a grade, since those that repeated and consequently moved to a different school are captured as new entrants, not as repeaters. This means our estimates of repetition are likely to be underreported. Similarly, learners who leave a school are captured as dropouts, leading to our estimates of dropout rates likely being over-reported. By the same token, learners who change schools are captured as new entrants into the system, resulting in "drop-in" rates also being overestimated. The extent of these issues shown in Figure 3, which compares dropout rates per grade in 2018 with drop-in rates in 2019. The figure shows that the percentage of learners who dropped out in 2018 is roughly the same as the percentage of learners who "dropped in" in the same year throughout the primary schooling phase, which means we can infer with relative certainty that many of those learners captured as dropouts in 2018 are the same learners captured as drop-ins in the following year. From Grade 7, however, the dropout rates in each grade are higher than the drop-in rates. Unfortunately, it is not clear what approach is best for dealing with this problem, and therefore a number of different approaches are used to estimate repetition and dropout rates based on the Lurits data in this report. More details on the different approaches are provided throughout the report.

Despite these limitations, the aim of this report is to triangulate across the different available datasets to answer questions such as:

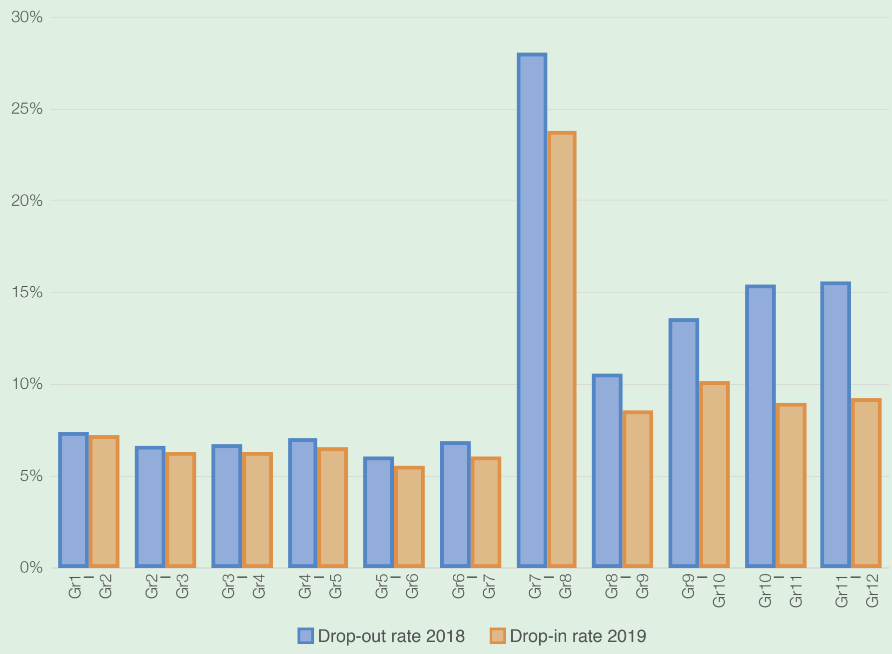
How many learners progressed through the school system without any repetition?

How many learners are still in the school system after repeating at least once?

How many learners dropped out of the school system?

What is the association between performance in the SBAs and grade progression, repetition, and dropout?

Figure 3: Comparison of learners who "dropped out" and "dropped in" by grade, Lurits 2018 and 2019



3. Enrolment patterns[4]

The available data suggests that South Africa has had relatively stable enrolment in recent years, with the exception of Limpopo, which has seen a slight decline in enrolment over the last 20 years, and the Western Cape, which has experienced enrolment growth in all grades over the period 2007-2019. In terms of enrolment by grade. Van der Berg, Van Wyk, et al. (2019), in their paper on the value of the SA-SAMS data, also produced for the MSDF, present evidence of clear enrolment “bulges” in Grades 4 and 10 – that is, these grades consistently have much higher enrolment than other grades. The available Lurits data provides more evidence of this, as seen in Figure 4. These enrolment bulges in certain grades are indicative of high repetition rates, resulting in more than one cohort stuck in these grades. It is clear from the figure that the enrolment bulge is particularly pronounced in Grade 10, reflecting high repetition rates in this phase. The fact that these bulges occur in Grades 4 and 10 in particular may be linked to the national repetition policy, according to which learners are not allowed to repeat more than once in each schooling phase[5]. Since Grade 4 is the beginning of the intermediate phase, and Grade 10 is the beginning of the further education and training (FET) phase, it could be that struggling learners who were not allowed to repeat before reaching these grades (if they had already repeated once in the previous phase) are then held back when they start a new phase. This high prevalence of grade repetition (a reflection of inefficiency in the system) is a defining feature of South Africa’s basic education system that has received too little attention in research and policy discussions alike. The next section considers the question of grade repetition in more detail.

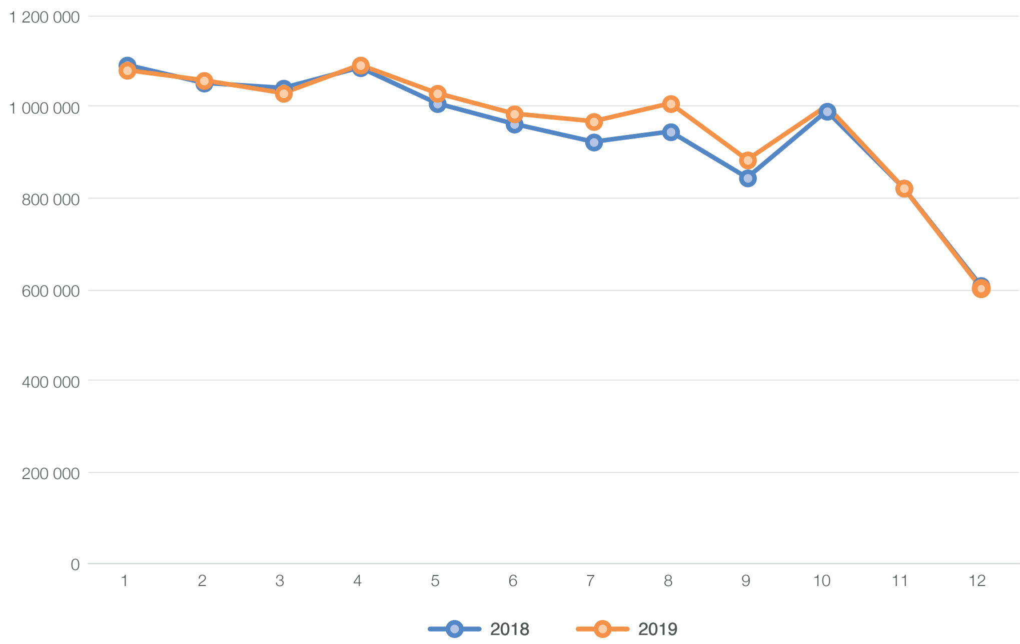
It is interesting to note from Figure 4 that enrolment sharply decreases in Grades 11 and 12, suggesting that many of the learners who repeat Grade 10 (and therefore contribute to the enrolment bulge in this grade) drop out of the system between Grades 11 and 12. However, due to the problems related to unique learner identifiers outlined in Information Box A above, it is not possible to know whether the learners that appear as dropouts in these grades dropped out of the system entirely, or whether they simply changed schools and were assigned new unique identifiers when they did so. The next sections on repetition and dropout rates deal with this issue in more detail.

Enrolment bulges in Grades 4 and 10 **indicate high repetition rates,** with more than one cohort stuck in these grades

4 The analyses of enrolment, repetition and dropout in this report is limited to public schools.

5 There are four schooling phases in South Africa: foundation phase (Grades R-3), intermediate phase (Grades 4-6), senior phase (Grades 7-9) and further education and training (FET) phase (Grades 10-12).

Figure 4: Cross-sectional analysis of Lurits enrolment data, 2018 & 2019



High repetition rates mean that **a large proportion of learners are over-age** for their grade

4. Repetition patterns

The available Lurits data shows that repetition is extremely common in South Africa. Though government has put in place a repetition policy that states that no child should be held back more than once in any education phase (i.e. foundation (Grades R-3), intermediate (Grades 4-6), senior (Grades 7-9), and further education and training (FET) phase (Grades 10-12)), this does not necessarily keep repetition rates low, nor is this policy always fully complied with by many schools, including primary schools. While holding learners back could provide an opportunity for remedial support, this is often not realistic in developing countries like South Africa where fiscal constraints are binding. On the other hand, promoting learners to higher grades when they have not mastered the required content increases heterogeneity at higher grades, making it very difficult for teachers to teach at the right level[6].

6 See Van der Berg, Wills, *et al.* (2019: 7-18) for a discussion of the arguments for and against grade repetition as remedial action for struggling learners.

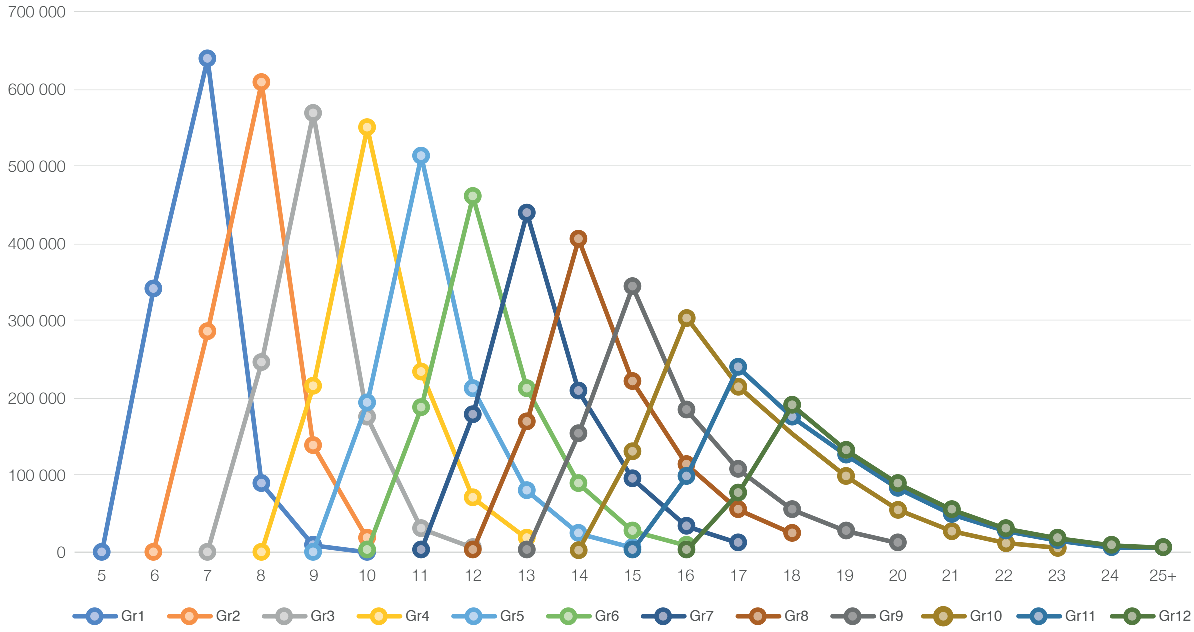
The result of high rates of repetition is that a large proportion of learners are over-age for their grade, an often-overlooked feature of the South African schooling system. Figure 5 plots the age distributions of all learners in the national Lurits database by grade in 2019. More uniformly distributed curves indicate that a larger proportion of learners are in the correct age for the grade. The figure shows that the age distribution of each grade becomes increasingly skewed to the right in each grade after Grade 1, indicating that increasingly larger proportions of learners are over-age in each grade, while the peak of the curves (usually representing those appropriately aged) declines across grades. This is because the proportion of over-age learners is cumulative - learners remain over-age when they have repeated once, and the percentage of over-age learners would only drop if large numbers of such learners drop out.

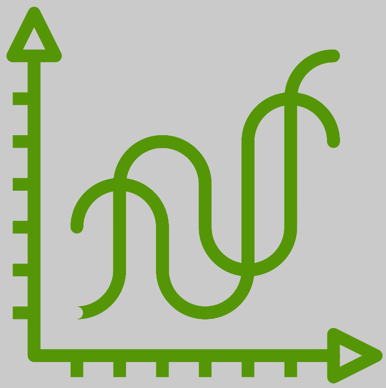
Figure 6 shows the proportion of over-age learners by grade. The figure shows that a large proportion of learners (roughly a third) are already at least one year over-age by Grade 4. This constitutes a sharp increase compared to the 9% of learners that are over-age in Grade 1. Said differently, while one in 10 Grade 1 learners are over-age, this proportion increases steadily over the foundation phase so that one in three learners are over-age by Grade 4. Since being over-age is a proxy for grade repetition, these results indicate that grade repetition is extremely common in South Africa, even in the primary school grades.

The cumulative effect of repetition through the schooling system can also be seen in Figure 6, which shows the proportion of learners that are over-age by grade. The figure shows that the proportion of over-age learners decreases slightly between Grades 11 and 12, suggesting that over-age learners are more likely to drop out between these grades than they are in other grades. The association between being over-aged and dropout is discussed explicitly in Section 5. A striking finding from Figure 6 is that the majority of learners (between 55% and 59%) in Grades 10, 11, and 12 are over-age for their grade. That is, in a typical Grade 10, 11, or 12 class in South Africa, it is more common for learners to be over-age than to be the correct age-for-grade.

Most learners
in Grades 10, 11 and 12
are over-age for their grade

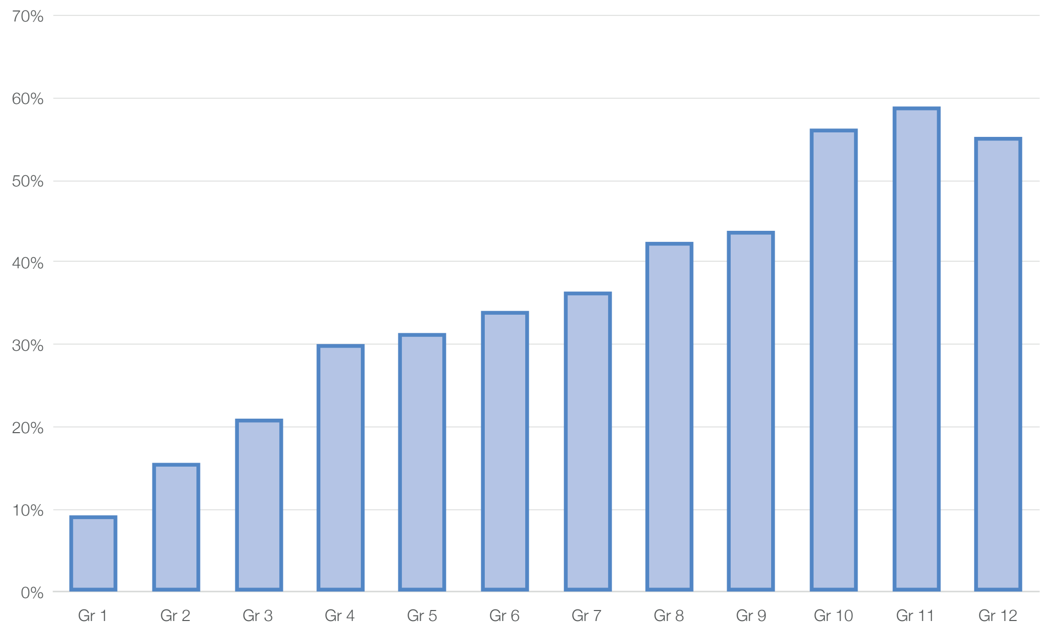
Figure 5: Enrolment by age and grade for public schools, Lurits 2019





~~More than~~ 20%
of learners in
Grades 10, 11 and 12
are three or more years
over-age

Figure 6: Proportion of over-age learners by grade, Lurits 2019

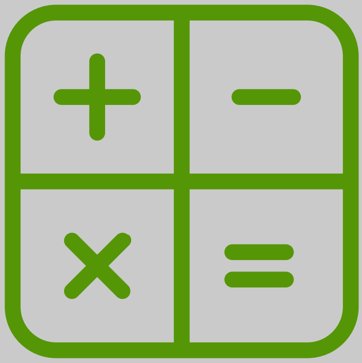


Information Box B: Why are over-age learners a problem?

The vast majority of children in South Africa (at least 91% according to 2019 Lurits data) enrol in Grade 1 in the calendar year that they turn seven. As a result, being over-age is a proxy for earlier grade repetition. As such, the proportion of over-age learners is an indication of how widespread grade repetition is in the system. Widespread grade repetition adds major costs to public spending on education: Van der Berg, Wills, et al. (2019) estimate that the cost of having repeaters in the system could be as high as between R20 billion and R29 billion. These authors argue that if repetition rates in only Grade 10 (the grade with the highest repetition rates) were to halve, as a conservative estimate R2 billion could be saved and relocated to other spending items in the education budget.

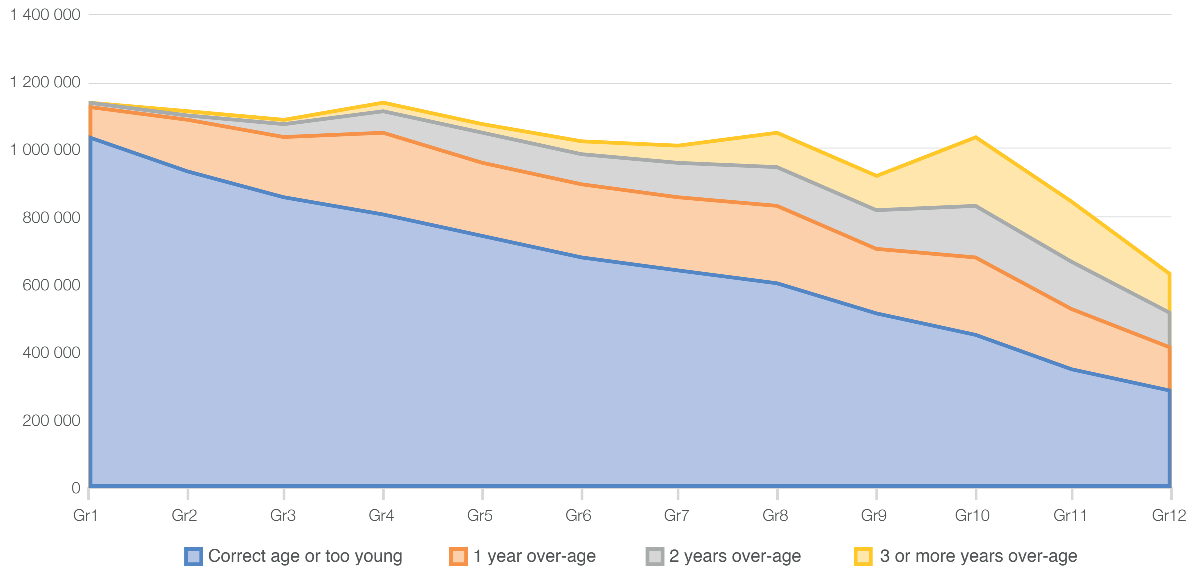
The same information is presented in Figure 7 and Figure 8, which show the proportions of learners who are the correct age-for-grade (the blue areas of the stacked bar graphs), as well as those who are one, two, and three or more years over-age (the orange, grey, and yellow areas of the graphs, respectively). These figures indicate that there are sharp increases in the proportions of over-age learners in Grade 10 and to a lesser extent Grade 4, reflecting the enrolment bulges in these grades (i.e. learners being more likely to be held back in these grades than other grades) discussed in the previous section.

Children in **poorer schools are less likely to master basic literacy and numeracy skills,** leading to learning deficits and repetition



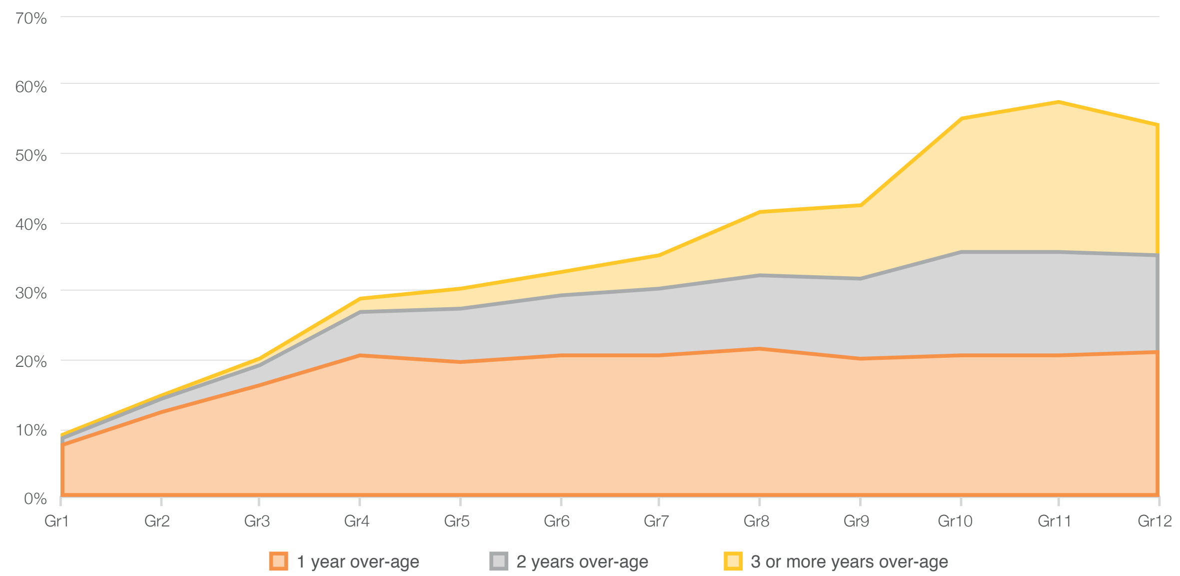
The increase of the proportion of learners who are two and especially three or more years over-age in Grade 10 could be indicative of many learners repeat~~ing~~ Grade 10 more than once. This, too, is investigated in more detail in the next section. The large proportion of learners in the FET phase who are three or more years over-age is particularly noteworthy: Figure 8 shows that more than 20% of learners in Grades 10, 11, and 12 are three or more years over-age. Given that most learners enrol in Grade 1 at the appropriate age, the large proportions of learners who are three or more years over-age indicate that there are many learners in the system who repeat three or more times during their schooling careers. This points to high levels of inefficiency within the system – what should take learners one year in terms of acquiring the skills and knowledge required to progress to the next grade takes many South African learners two, three, or even four years, in some cases. This is under the assumption that learners who repeat will eventually manage to meet the requirements to pass that grade. This is, unfortunately, not the case for a large number of learners: Many learners who are held back (especially in the FET phase) never manage to acquire the knowledge and skills required to be promoted to the next grade, and many go on to drop out of the schooling system entirely. The question of high dropout rates is returned to in more detail in Section 5. Before proceeding to that discussion, we consider how the problem of over-age learners differs by school wealth.

Figure 7: Number of on-track and over-age learners by grade, Lurits 2019



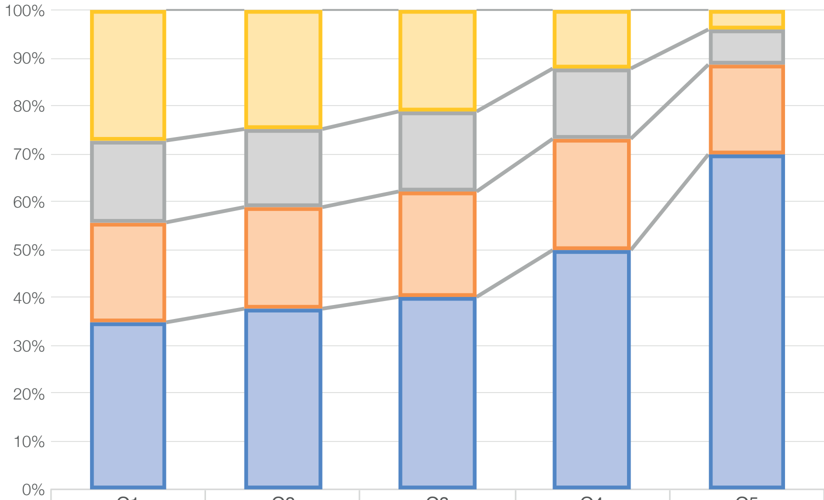
~~Roughly one~~ in four learners who enrol in Grade 10 will repeat this grade

Figure 8: Proportions of over-age learners by grade, Lurits 2019



Grade repetition differs by school wealth, with higher repetition rates in poorer schools. Given unequal access to quality schooling, children in poorer schools are less likely to master basic skills of literacy and numeracy, giving rise to learning deficits which make it challenging to keep pace with the curriculum, thus leading to higher repetition rates in these schools (Van der Berg, Wills, et al., 2019). The Lurits data shows how grade repetition is more prevalent among poorer schools, as can be seen in Figure 9, which shows the proportions of over-age learners in Grade 10 by school quintile. The figure shows that the patterns of over-age learners observed in Figure 6, Figure 7, and Figure 8 above are largely concentrated in Quintile 1 to 4 schools, with the majority of Grade 10 learners (between 60% and 65% in Quintiles 1-3) in these schools being over-age. By contrast, only 30% of Grade 10 learners in Quintile 5 schools were over-age in 2019 – still a high proportion if one considers that most of these children are drawn from more affluent home backgrounds, but far less than in the non-fee-paying schools.

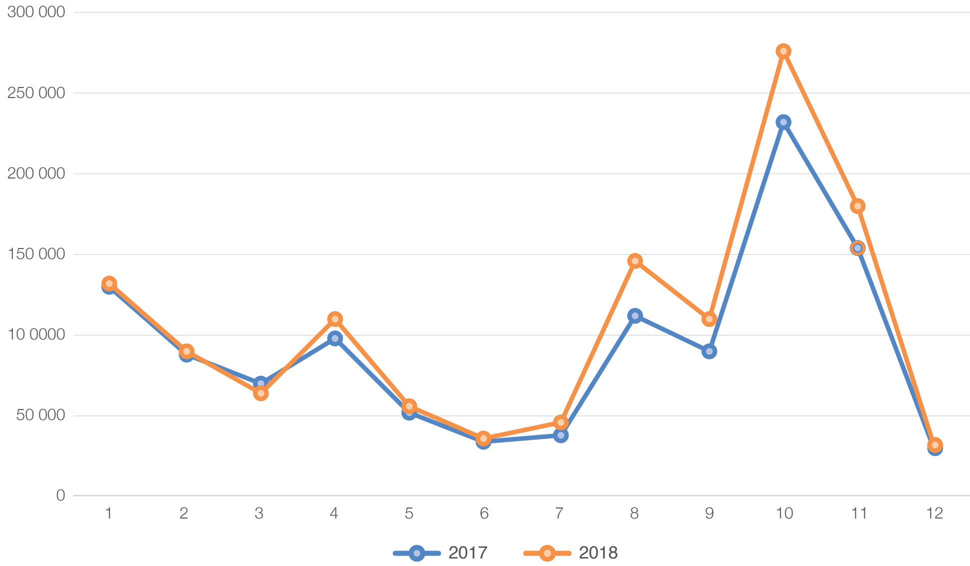
Figure 9: Proportions of over-age learners in Grade 10 by school quintile, Lurits 2019



	Q1	Q2	Q3	Q4	Q5
3+ years	59 228	57 260	59 977	16 726	5 228
2 years	37 164	38 427	46 796	19 837	10 116
1 Year	45 532	49 079	62 816	31 293	25 185
Appropriate	75 614	87 586	113 769	67 807	93 239

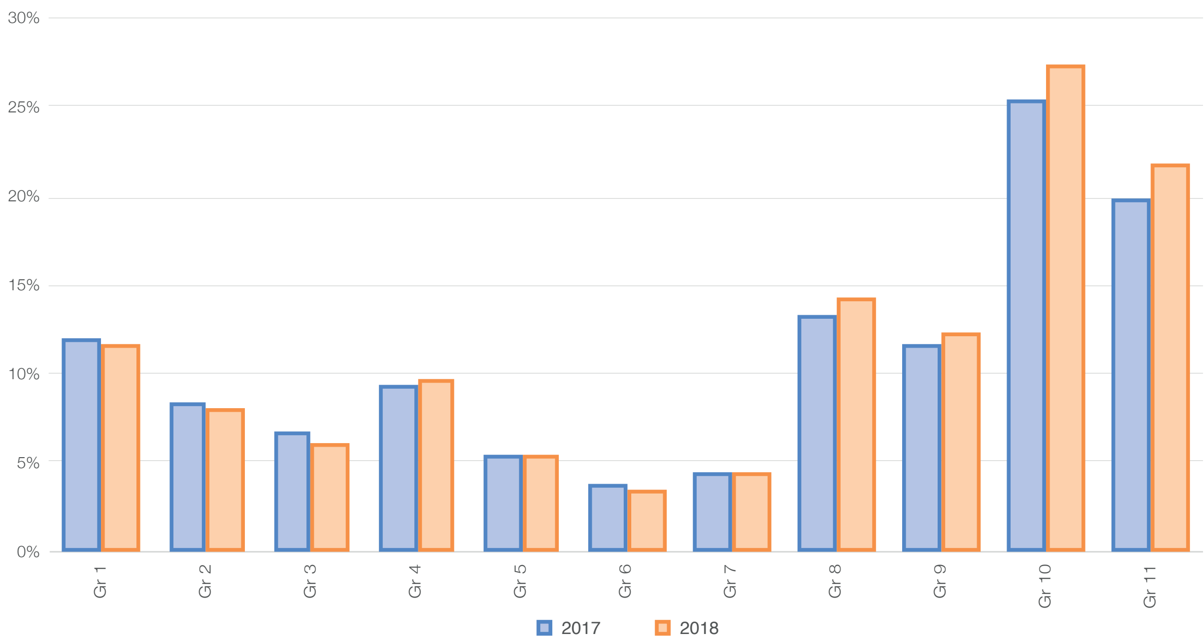
Figure 10 shows actual number of repeaters by grade, and shows the repetition patterns underlying the enrolment bulges and proportions of over-age learners per grade discussed above. Figure 11 shows the same information, but repetition is expressed in terms of rates (rather than the number of learners repeating). A number of noteworthy findings are evident from the figures. Firstly, they show clearly that repetition is much more common in the secondary schooling phase than in the primary schooling phase. The high repetition rate in Grade 10, especially, confirms that Grade 10 is a big hurdle on the road to matric – roughly one in four learners who enrol in Grade 10 will repeat this grade, according to the 2017 and 2018 Lurits data. The figure further shows that repetition rates decrease sharply in Grades 11 and 12, mirroring the enrolment patterns discussed in Section 3 above. The decrease in repetition rates between Grade 10 and Grade 12 are therefore further indicative of Grade 10 being a "gatekeeping" grade, where the greatest number of learners are held back, and the majority of those repeaters drop out of the system before reaching Grade 12. Once again, however, it is not possible to determine from the anonymised Lurits data whether the learners who repeat Grade 10 are the same learners who drop out before Grade 12.

Figure 10: Number of repeaters by grade, 2017 & 2018, Lurits



Dropout rates can be calculated from Lurits data using the UNESCO method

Figure 11: Repetition rates by grade, 2017 & 2018, Lurits



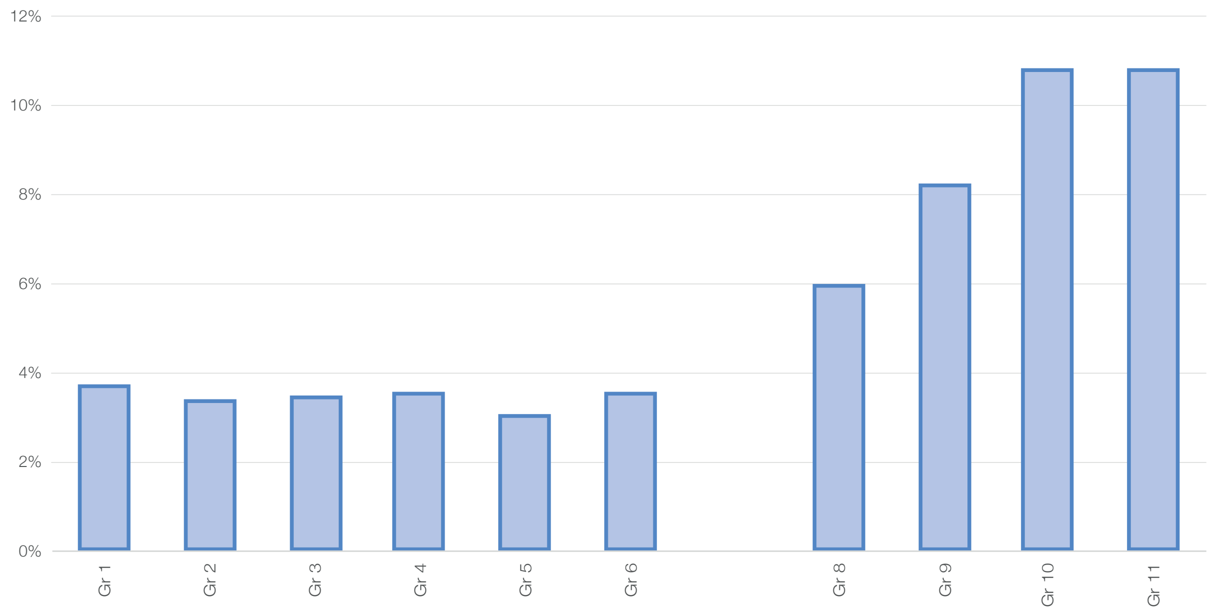
5. Dropout rates

Dropout rates can be calculated from Lurits data using the so-called UNESCO method. Essentially, this calculates dropouts using information from two years. To determine how many students have been promoted from one grade to the next, this figure can be calculated as enrolment in the next grade in the next year, minus those who are repeating that grade. If enrolment in the initial grade in the initial year was X, and Y persons were promoted and Z repeat the initial grade, dropout from the initial grade is then simply calculated as X – Y – Z.

Calculating dropout in this way is of course subject to some error. Those considered dropouts from the system could have joined another education system, e.g. students considered 'dropouts' from the public-school system could simply have moved to private schools, or they could have died or moved to another country. Similarly, movements from another country or province could also affect the accuracy of the dropout estimates, or students who have dropped out in one year may re-enter the system in another year.

Figure 12 shows dropout rates by grade for the country, based on Lurits data. As no dropouts are shown for Grade 7 or Grade 12, it not possible to calculate dropout rates in these grades using the UNESCO method, which requires enrolment numbers in the subsequent grade[7]. Dropout rates are generally low through primary school, though still higher than is desirable. The real dropout problem is now concentrated in Grades 10 and 11, where 11% of learners in each of those grades dropped out in 2018. The cumulative effect of these high dropout rates (especially in the secondary schooling phase) is that a large proportion of learners who start school in South Africa will drop out before reaching matric. This is a major source of inefficiency in the system, especially when one considers that an NSC (matric) certificate is the only widely-recognised school-level qualification in South Africa. This means that learners who drop out before reaching Grade 12 or who reach Grade 12 but do not pass the NSC examinations have nothing to show (at least in terms of qualifications) for the time they spent in school – which is usually more than 10 years (Gustafsson, 2011).

Figure 12: Dropout rates by grade, Lurits 2018



7 Dropout is calculated at the school level, and primary schools end at Grade 7. As such, it is not possible to use school-level enrolment data to calculate dropout rates in Grade 7.

These results illustrate the power of simple cross-sectional enrolment data (which is collected mainly for administrative purposes) for shedding light on the functioning of a country's schooling system. For example, the preceding three sections have relied solely on enrolment data to produce important insights into the efficiency of the South African schooling system. These analyses of enrolment data show that the South African schooling system does not manage to equip the majority of learners with the capabilities and skills necessary to progress successfully through the system without being held back at some point in their schooling careers. Grade repetition is extremely common, even in the primary school grades, and becomes even more pronounced in the secondary schooling phase, peaking in Grade 10. There is evidence of a strong bottleneck effect in the FET phase, whereby learners seem to get "stuck" in Grade 10 for two or more years, and only a fraction of learners manage to make it to Grade 12. The fact that the majority of learners will repeat at least one grade (and a significant proportion will repeat more than once) implies major inefficiencies in public spending on education. In addition to widespread repetition, high dropout rates are also a major source of inefficiency in the system. High dropout rates mean that a large proportion of learners who start school in South Africa will never reach matric. This situation is especially troubling when one considers that the NSC or matric examination is the only widely recognised school-level qualification in South Africa.

Cross-sectional enrolment data is limited, however, in that it does not allow one to trace the trajectories of individual learners through the schooling system. As such, it is very difficult to investigate inefficiencies in how learners flow through the schooling system in more detail using only cross-sectional enrolment data. For example, while the Lurits data clearly shows significant enrolment bulges in certain grades, without reliable unique identifiers for all learners, it is not possible to determine whether the learners who repeat Grade 10 have repeated previous grades as well, or whether those learners who repeat Grade 10 are the same learners who later drop out in Grades 11 and 12. In short, without reliable unique learner identifiers in national-level administrative education data, it is not possible to trace the trajectories of individual learners through the system. Doing so is essential for predicting the likely trajectories of learners currently enrolled in school, which would aid in identifying for example which learners are at risk of dropping out. Cross-sectional data is also limited in that it does not provide any indication of changes in the system (for example in enrolment) over time. Understanding trends in enrolment over time is essential for evidence-based education planning.

A large proportion of learners drop out before matric,
a major source of inefficiency in the system

6. The value of longitudinal data: Insights from the Western Cape

This section draws considerably on the results presented in an earlier paper by Van Wyk (2021) on learner flows in the Western Cape using longitudinal data for the period 2007 to 2019. The aim is to illustrate how high-quality longitudinal data can be used to (i) investigate learner flows through the schooling system, and (ii) understand trends in enrolment over time. The longitudinal analysis is unfortunately limited to the Western Cape since that province has its own administrative education data system, Cemis, which contains reliable learner-level data for at least the last 13 years. As explained in Information Box A, patterns of learner flows in the Western Cape are not always representative of the rest of the country. As such, the intention of this section is primarily to show the kind of analysis that is possible – and the kind of conclusions that can be drawn – with the availability of reliable learner-level data collected consistently over time. If the other provinces continue to improve in their submission of their data to the SA-SAMS system, such longitudinal analysis would be possible at the national level in the near future.

Longitudinal learner-level Cemis data over 13 years **allows investigation of changes in progression over time** in the Western Cape school system

High-quality administrative data can be used to inform education policy and practice if the data is collected consistently for long periods of time. High-quality longitudinal administrative data allows one to compare enrolment trends over time, which is instructive for effective education planning. Learner-level data that spans long enough time periods can also be used to investigate how cohorts of learners move through the system. This makes it possible to answer key questions such as "What is the profile of learners who dropped out of the system?" and "What factors are associated with successful grade progression?" This section aims to illustrate how longitudinal learner-level data can be used for both cohort analysis and for understanding trends in enrolment over time. This is done by presenting analysis of unit-level learner records in the Western Cape (Cemis data). Since the results of the analysis of Cemis data are presented in an earlier report (Van Wyk, 2021), this section only discusses the main results of that analysis to illustrate the value of longitudinal learner-level data.

6.1. Enrolment patterns in the Western Cape

By comparing enrolment data in the Western Cape over a 13-year period, Van Wyk (2021) is able to show that the province had consistently had an enrolment bulge in Grade 4, as is the case nationally. Interestingly, however, he presents evidence that there had been a change in progression patterns in the secondary schooling phase over time. Specifically, he finds that enrolment in Grade 8 had increased disproportionately over this period, while enrolment in Grade 10 had *decreased* over the same period. The reasons behind these unexpected changes in enrolment over time are unclear, though they could point to changes in grade retention practices in the province during this period. This result illustrates how longitudinal learner-level data can be used to investigate changes in enrolment patterns over time, which are important to take into account in education planning at the provincial as well as national level.

6.2. Analysis of the Western Cape 2007 Grade 1 cohort

By tracking a group (cohort) of learners that enrolled in Grade 1 in a given year as they move through the schooling system, Van Wyk (2021) is able to present a more detailed account of grade progression, repetition, and dropout patterns than is possible with cross-sectional data. He presents a number of noteworthy results in this regard:

- Less than half (45%) of learners that started Grade 1 in 2007 had successfully progressed to Grade 8 within the normal eight years, that is, without ever repeating.
- By the time they should have been in Grade 12, 39% of the original 2007 Grade 1 cohort had dropped out of the system entirely.[8]
- Progression rates for this cohort were extremely low in the FET phase, even among learners who had managed to reach Grade 10 without ever repeating – only 63% of these learners managed to progress to Grade 12 without repeating or dropping out.
- The result of these high repetition and dropout rates is that less than a third (28%) of the 2007 Grade 1 cohort had progressed successfully to Grade 12 without any repetition or dropout.

Progression **rates in the Western Cape are much higher** for recent cohorts

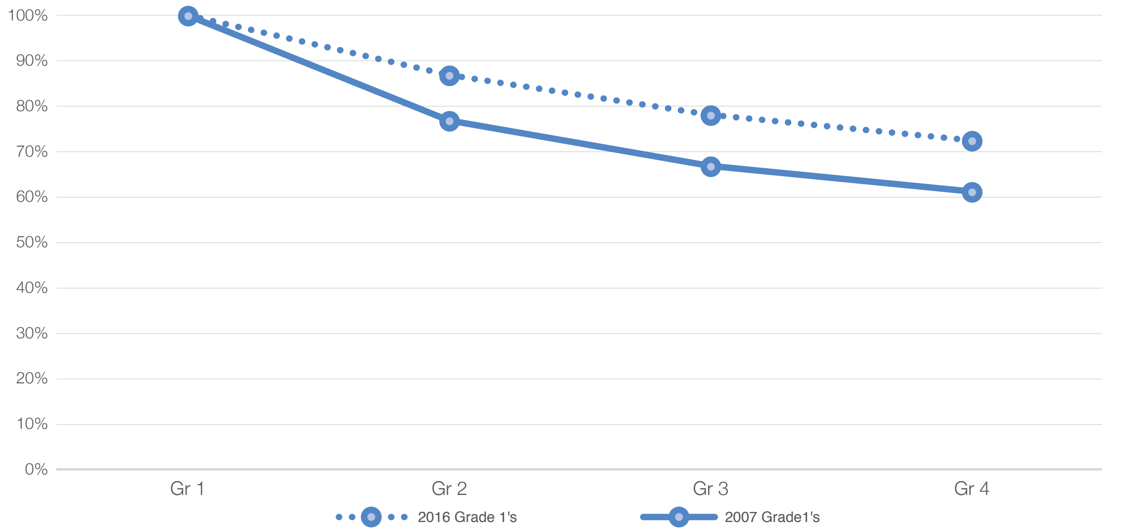
These results illustrate how longitudinal data can provide a much more detailed picture of learner flows through the system than cross-sectional data. Having a more detailed picture makes it easier to detect changes in the efficiency of the system over time. Using the longitudinal learner-level Cemis data, Van Wyk (2021) is able to present evidence of enhanced performance and improved efficiency in the Western Cape school system over time. These results are discussed in the next sub-section.

8 Since Cemis data is collected at the provincial level, it is not possible to trace students recorded in Cemis if they move to a different province. Calculating dropout rates at the provincial level is therefore likely to result in over-estimates of dropout rates.

6.3. Evidence of enhanced performance and improved efficiency in the Western Cape over time

Given the evidence of changes in grade retention practices over the period 2007 to 2019 in the Western Cape discussed above, Van Wyk (2021) compares progression patterns of the 2007 Grade 1 cohort with those of learners who started Grade 1 in 2016. This comparison is shown in Figure 13, which plots the number of learners who progressed to Grades 2, 3, and 4, as proportions of the original Grade 1 cohorts. It is clear from the figure that progression rates in the early grades were higher among the 2016 cohort.

Figure 13: Progression of the Western Cape 2007 and 2016 Grade 1 cohorts to Grade 4 in three years, Cemis

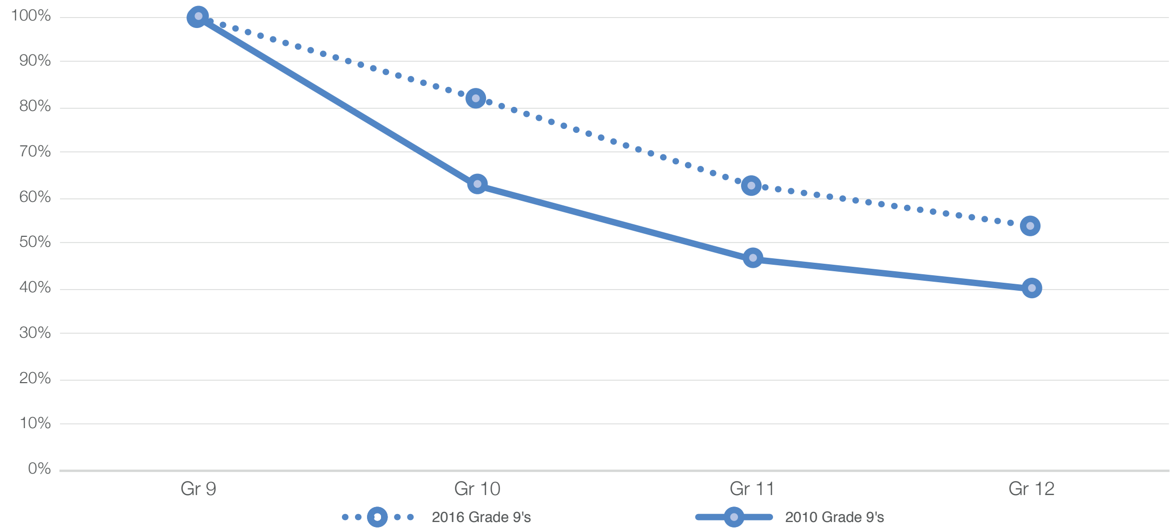


Source: Van Wyk (2021)

Van Wyk (2021) further presents evidence of improved efficiency in learner progression over time in the Western Cape, as seen in Figure 14, which compares the progression patterns of the cohort of learners that were in Grade 9 in 2010 with the cohort that were in Grade 9 in 2016, and shows that progression rates were much higher for the later cohort.

Accelerated learner **flows without a drop in education standards imply improved system efficiency** in the Western Cape

Figure 14: Progression of the Western Cape 2010 and 2016 Grade 9 cohorts to Grade 12 in three years, Cemis.



Source: Van Wyk (2021)

It is important to note that there is no evidence that this improvement in the efficiency with which learners progressed through the schooling system in the Western Cape was accompanied by any deterioration in learner performance – in fact, there is some evidence that the province's systemic test results (in grades 3, 6 and 9) improved over the period 2011 to 2019 (Western Cape Education Department, 2020). At the secondary schooling level, the province's performance in the Trends In Mathematics and Science Study (TIMSS) remained the same over the period 2015 to 2019 (Reddy, Juan, Winnaar, Arends, Harvey, Hannan & Zulu), as did the province's matric results (Department of Basic Education, 2013, 2019). This evidence suggests that the accelerated learner flows were achieved without a reduction in education standards in the province, thus implying improved system efficiency.

This section has illustrated how longitudinal learner-level administrative data can be used for both cohort analysis and for understanding trends in enrolment and performance over time. The next section illustrates a further advantage of longitudinal learner-level administrative data, namely that learners' marks in SBA's can be linked to later academic performance, grade promotion, repetition, and dropout.

Longitudinal data can be used to
investigate ~~both~~ cohort analysis ~~and~~ trends in enrolment and performance

7. The association between SBA marks and learner flows in Limpopo

The preceding sections on learner flows through the schooling system have revealed that only a small proportion of learners who start school in South Africa will progress to Grade 12 without ever repeating or dropping out. The analysis discussed in this section is aimed at identifying potential reasons behind these high repetition and dropout rates. To do so, we make use of regressions to predict the probability of smooth progression between certain grades – that is, without repetition or dropout. Unfortunately, learner-level DDD data for two or more years was only available for a select sample of learners in Limpopo. We confine our regression analysis to the Limpopo sample for the period 2018 to 2020. We focus on learner marks in Mathematics (or Mathematical Literacy) and English First Additional Language (EFAL) as predictors in our models of the probability of progression without repetition or dropout. The availability of SBA results in Limpopo's DDD data provides a unique opportunity to investigate the relationship between SBA and grade progression, something that has received little attention in analysis of grade progression in South Africa to date – although this is largely due to the unavailability of longitudinal SBA data until now. This section's aims are twofold: (1) To illustrate how longitudinal administrative data that includes learner marks can be used to investigate how achievement in earlier grades is related to later repetition and dropout; and (2) To investigate what other learner characteristics are associated with how learners progress through school, after taking into account performance in SBA's.

For a given SBA performance, over-age learners repeat less, suggesting that the **national progression policy is applied in Limpopo**

SBA marks are intended to provide learners and teachers with an indication of how successfully learners are mastering the curriculum on a continuous basis (every term). This information is intended to aid learning in that it provides a signal to both learners and teachers of how well they are keeping up with the curriculum. While teachers are supposed to provide opportunities for both informal (for example, class discussions) and formal assessment (oral presentations, tests and examinations) throughout the school year, only the results of formal assessments are recorded as school-based assessment in the SA-SAMS database. Since SBA marks are supposed to provide information of how well learners are keeping up with the curriculum, SBA marks should be used to inform teacher and principal decisions to promote a learner to the next grade. Until now, however, it was not possible to determine whether SBA marks do in fact inform grade promotion decisions in this way. Our regression analysis therefore makes an important contribution to understanding the relationship between SBA marks and learner flows through the system. In addition, the availability of SBA marks makes it possible to control for learner performance so that other learner characteristics (apart from SBA achievement) that influence progression, repetition, and dropout, can be identified. The result of our regression models in terms of these two contributions to our understanding of learner flows in Limpopo are summarised below[9].

9 Regression results can be found in Tables A1 to A3 in the Appendix.

Higher SBA marks in mathematics and EFAL are significantly associated with greater on-track grade progression between grades (e.g. Grades 5 and 7). This result constitutes important new evidence of the association between SBA marks and later academic outcomes, such as on-track grade progression two years later. The specific finding that SBA marks in mathematics and EFAL in one grade are good predictors of on-track grade progression over the next two grades is an important finding, since it suggests learners who perform poorly in these subjects are at risk of later grade repetition (and even dropout). While evidence from more provinces is required to know to what extent this is the case nationally, this evidence from Limpopo makes an important first contribution to our understanding of how SBA marks can be used to identify learners who may be at risk of later repetition and dropout. Again, this illustrates the value of administrative data for informing education policy and practice.

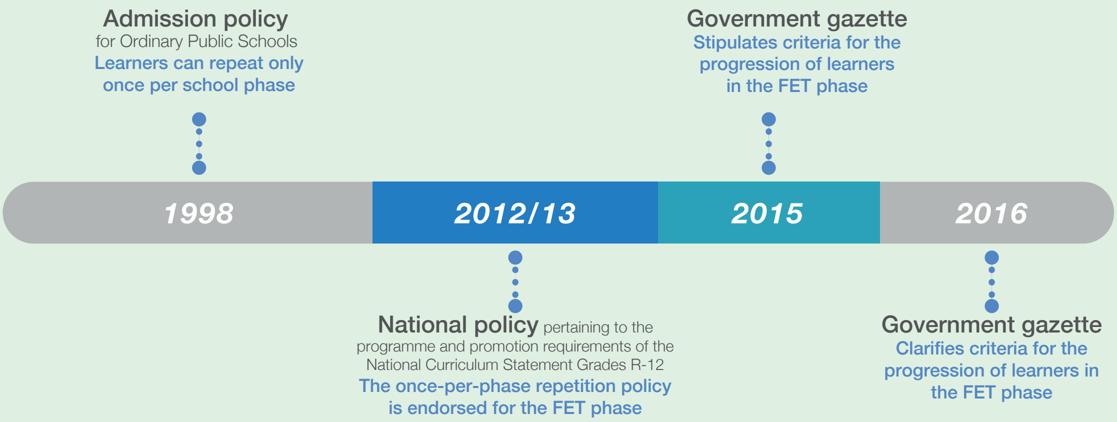
For a given mark in mathematics and EFAL, girls are more likely than boys to progress on-track until Grade 10, where they are less likely than boys to progress to Grade 12 without repeating or dropping out. This result is noteworthy since it points to gender-specific reasons for repetition and dropout that seem to change in the FET phase. Unfortunately, the available administrative data does not contain information that would allow us to investigate what these gender-specific reasons for repetition and dropout could be. Existing evidence from Gustafsson (2011) suggests that part of the reason for lower successful progression rates through the FET phase among girls could be pregnancy: Evidence from the National Income Dynamics Study (NIDS) shows that 13% of 16-year-old girls (who should be in Grade 10 or 11 if they are on-track in terms of age-for-grade) in South Africa indicated that they had given birth in the past or were pregnant at the time of taking the survey, a figure which increases dramatically to 47% by age 20. Although Gustafsson (2011) also presents evidence that pregnancy is by no means always associated with dropping out, the fact that pregnancy affects girls combined with the fact that the probability of falling pregnant increases at the ages when girls are in the FET phase, makes it likely that pregnancy might be one of the reasons for the gendered progression patterns in the FET phase observed for Limpopo.

Higher SBA marks
are associated with greater
on-track progression
between grades

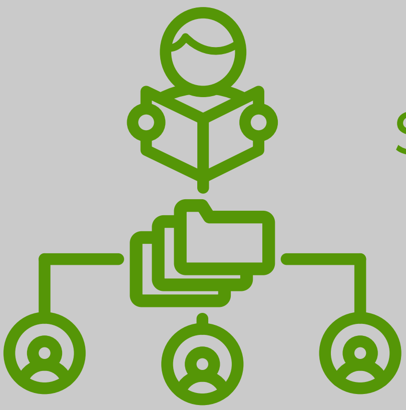
For a given SBA mark in mathematics and EFAL in a particular grade, over-age learners are less likely to repeat that grade than learners who are on-track in terms of age-for-grade. This result is found for all grades in Limpopo, and suggests that teachers in Limpopo take learners' age into account when making decisions about grade promotion. That is, when faced with two students who performed the same in mathematics and EFAL, teachers appear to be more likely to promote the over-age learner. This may be evidence that schools in Limpopo do apply the national progression policy that learners may only repeat once per school phase. While this policy has largely been applied in the General

Education and Training Phase (Grade R to Grade 9) since being gazetted in 1998, it was only endorsed in the FET phase in 2013 (Kika and Kotze, 2018). The fact that over-age learners in all grades (including Grades 10, 11, and 12) are less likely to repeat (for given mathematics and EFAL marks) suggests that the national progression policy is applied in all school phases in Limpopo, including the FET phase. This may be due to the endorsement of the progression policy in the FET phase in government gazettes issued since 2015, as shown in Information Box C. The fact that teachers are more likely to promote over-age learners (given mathematics and EFAL marks) in Limpopo is an important result, especially considering qualitative evidence presented in a recent paper by Mogale and Modipane (2021) that teachers in high schools in Limpopo were not adequately supported by their school management teams in the implementation of the once-per-phase repetition policy in the FET phase. Our regression results suggest that despite this, teachers in Limpopo do take learners' age into account in their grade promotion decisions.

Information Box C: Timeline of the development of legislation relating to progression policy in South Africa



These results show how the progressively improving availability of SBA data over time makes it possible to investigate important trends in grade promotion practices at the provincial level. Such analysis will soon be possible for more provinces if they continue to consistently submit their SBA data over time.



Progressively improving availability of **SBA data makes it possible to investigate important trends in grade promotion practices** at provincial level

8. Summary of main findings

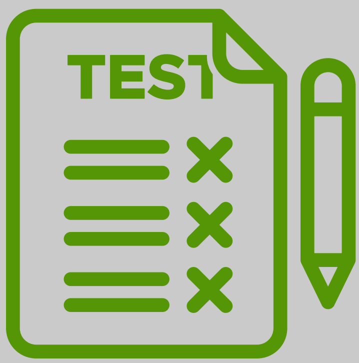
By triangulating across different education datasets in South Africa, this report has shown what kind of analysis is possible with high-quality learner-level administrative education data. This analysis has allowed us to highlight important trends in how learners progress through the schooling system. The results point to major inefficiencies in the system, with both repetition being extremely common in all grades, and many learners dropping out of the system before reaching matric. Learners tend to stay in the system until the FET phase (Grade 10), even if they are held back. From Grade 11, however, enrolment drops off significantly, and dropout is extremely common. Our analysis also shows that grade repetition is more common at the start of a new schooling phase, especially the start of the intermediate (Grade 4) and FET (Grade 10) phases. This evidence suggests that the national progression policy that learners can only repeat once per schooling phase is widely applied throughout the country. Progression patterns in the secondary schooling phase in the Western Cape is somewhat of an outlier in this regard, as repetition in that province is concentrated in Grade 8. Lastly, given new longitudinal SBA data at the provincial level, we are able to present evidence of the association between SBA marks and later learner outcomes in terms of on-track grade progression, repetition and dropout.

Taken together, these results all point to the conclusion that learners are not successfully acquiring the knowledge and skills required in the curriculum within the time that they are supposed to, if at all. There is clearly much scope for improvement in the efficiency with which learners progress through the schooling system. This does not necessarily imply, however, that learners should be progressed through the system even if they do not meet the requirements for being promoted to the next grade. Rather, it is important that the factors underlying the slow progression through the system are addressed.

9. Conclusion

This report was compiled to draw from the strength of available datasets to present a picture of learner performance and flows through the South African education system. By triangulating across different available datasets, we are able to present evidence of major inefficiencies in the South African schooling system, with the majority of learners not acquiring the knowledge and skills required to be progress through the system on time, and many learners dropping out without obtaining a matric qualification. It is hoped that the results presented in this report make clear the crucial importance of high-quality administrative data for informing education policy and practice.

The majority of learners **do not acquire the knowledge and skills required to progress** through the system on time



References

Department of Basic Education (2013) *National Senior Certificate Examination 2013: School Performance Report.* Pretoria.

Department of Basic Education (2019) 'Ministerial Release of the 2019 NSC Results National Senior Certificate Programme', pp. 1–44. Available at: http://www.ecexams.co.za/2019_Nov_Exam_Results/2019 NSC 2019 Highlights Report.pdf.

Figlio, D. N., Karbownik, K. and Salvanes, K. G. (2015) 'Education research and administrative data', *NBER Working Papers,* 53(21593)~~, pp. 1689–1699~~.

Gustafsson, M. (2011) 'The when and how of leaving school: The policy implications of new evidence on secondary schooling in South Africa', *Stellenbosch Economic Working Paper 09/2011.* Stellenbosch University, Department of Economics.

Howie, S.J., Combrinck, C., Tshele, M., Roux, K., McLeod Palane, N. & Mokoena, G.M. (2016) *PIRLS Literacy 2016: South African Highlights Report.* Pretoria.

Kika, J. & Kotze, J. (2018) ~~*Unpacking Grade Repetition Patterns in Light of the Progression Policy in the Further Education and Training Phase.*~~

Mogale, M. L. and Modipane, M. C. (2021) 'The implementation of the progression policy in secondary schools of the Limpopo province in South Africa'; *South African Journal of Education,* 41(1), pp. 1–10. doi: 10.15700/saje.v41n1a1853.

Reddy, V., Winnaar, L., Juan, A., Arends, F., Harvey, J., Hannan, S., Namome, C., Sekhejane, P. & Zulu, N. (2019) TIMSS 2019: *Highlights of South African Grade 9 Results in Mathematics and Science.* Cape Town: HSRC.

Reddy, V., Juan, A., Winnaar, L., Arends, F., Harvey, J., Hannan, S. & Zulu, N. (2019) 'TIMSS 2019: Highlights of the Western Cape Province Grade 9 Results in Mathematics and Science'. Cape Town: HSRC.

Western Cape Education Department (2020) *Annual report 2019/2020, AIMS Mathematics.* Cape Town. doi: 10.3934/math.2020i.

Wills, G., Shepherd, D. & Kotze, J. (2016) 'Interrogating a Paradox of Performance in the WCED: A Provincial and Regional Comparison of Student Learning'. Stellenbosch Economic Working Paper 14/2016. Stellenbosch University, Department of Economics.

Van der Berg, S., Wills, G., Selkirk, R., Adams, C. & Van Wyk, C. (2019) 'The cost of repetition in South Africa'. Stellenbosch Economic Working Paper 14/2016. Stellenbosch University, Department of Economics

Van der Berg, S., Van Wyk, C., Selkirk, R., Rich, K. & Deghaye, N. (2019) 'The promise of SA-SAMS & DDD data for tracking progression, repetition, and drop-out'. *Stellenbosch Economic Working Paper 17/2019.* Stellenbosch University, Department of Economics.

Van Wyk, C. (2021) 'Learner flow through patterns in the Western Cape using CEMIS datasets from 2007 to 2014: A longitudinal cohort analysis'. *Stellenbosch Economic Working Paper 01/2021*. Stellenbosch University, Department of Economics

Appendix

Table A1: Regression results of the probability of successful grade progression between 2018 and 2020, Limpopo DDD data

	(1) Gr5-7	(2) Gr8-10	(3) Gr9-11	(4) G10-12
Maths mark2018	0.005***	0.007***	0.008***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Eng mark2018	0.006***	0.010***	0.012***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Chose maths 2018				0.118***
				(0.006)
Chose maths 2019			-0.107***	
			(0.007)	
Female	0.018***	0.047***	0.010***	-0.005*
	(0.002)	(0.003)	(0.004)	(0.003)
Q2	0.000	-0.014	-0.036***	-0.007
	(0.005)	(0.009)	(0.011)	(0.008)
Q3	-0.001	-0.008	-0.040***	-0.021**
	(0.007)	(0.011)	(0.013)	(0.010)
Q4&5	-0.045***	-0.023	0.040*	0.001
	(0.011)	(0.020)	(0.023)	(0.021)
Missing Q	-0.095***	-0.061***	0.082***	0.015
	(0.016)	(0.019)	(0.018)	(0.016)
1 year over-age (2017)	0.015***	-0.070***	-0.122***	-0.012**
	(0.004)	(0.005)	(0.005)	(0.005)
2 years over-age (2017)	0.028***	-0.099***	-0.169***	-0.030***
	(0.007)	(0.007)	(0.006)	(0.006)
3+ years over-age (2017)	-0.017	-0.124***	-0.190***	-0.043***
	(0.012)	(0.009)	(0.007)	(0.006)
Constant	0.202***	-0.179***	-0.341***	-0.412***
	(0.012)	(0.014)	(0.020)	(0.012)
Observations	**133 617**	**126 935**	**102 382**	**142 739**
R-squared	**0.148**	**0.277**	**0.338**	**0.327**

Robust standard errors in parentheses. Reference group is Male Q1 learners that are the correct age for grade in 2017, that chose mathematical literacy in Grade 10, where applicable.
*** p<0.01, ** p<0.05, * p<0.1

Table A2: Regression results of the probability of repeating specific grades, Limpopo DDD data 2018

	(1) Grade 1	(2) Grade 4	(3) Grade 8	(4) Grade 10	(5) Grade 10	(6) Grade 11	(7) Grade 11
				Maths	Math Lit	Maths	Math Lit
Maths(Lit) mark2018	-0.010***	-0.006***	-0.007***	-0.009***	-0.015***	-0.009***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Eng mark2018		-0.007***	-0.010***	-0.013***	-0.015***	-0.010***	-0.013***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.001	-0.009***	-0.029***	-0.018***	-0.048***	-0.009**	-0.032***
	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)
Q2	0.003	-0.002	0.016**	0.059***	0.046***	-0.008	-0.004
	(0.004)	(0.004)	(0.008)	(0.013)	(0.013)	(0.012)	(0.013)
Q3	0.014***	0.007	0.015	0.089***	0.046***	0.029**	0.023
	(0.005)	(0.006)	(0.010)	(0.016)	(0.016)	(0.014)	(0.016)
Q4&5	-0.001	0.035***	0.052***	0.112***	0.109***	0.033	0.140***
	(0.008)	(0.009)	(0.013)	(0.021)	(0.029)	(0.026)	(0.034)
Missing Q	0.000	0.031***	0.049***	0.050***	0.053**	0.042**	0.055**
	(0.009)	(0.010)	(0.014)	(0.017)	(0.021)	(0.016)	(0.026)
1 year over-age (2018)	-0.083***	-0.035***	-0.006	-0.087***	-0.048***	-0.077***	-0.068***
	(0.004)	(0.003)	(0.004)	(0.007)	(0.006)	(0.006)	(0.006)
2 years over-age (2018)	-0.078***	-0.106***	-0.005	-0.115***	-0.093***	-0.096***	-0.101***
	(0.013)	(0.006)	(0.007)	(0.008)	(0.008)	(0.009)	(0.007)
3+ years over-age (2018)	-0.035	-0.135***	-0.040***	-0.138***	-0.165***	-0.104***	-0.146***
	(0.025)	(0.010)	(0.009)	(0.011)	(0.009)	(0.010)	(0.008)
Constant	0.776***	0.859***	0.986***	1.192***	1.529***	1.040***	1.484***
	(0.010)	(0.011)	(0.017)	(0.020)	(0.018)	(0.022)	(0.023)
Observations	**141 620**	**140 206**	**121 791**	**69 290**	**60 033**	**54 208**	**61 405**
R-squared	**0.351**	**0.299**	**0.269**	**0.274**	**0.334**	**0.222**	**0.283**

Robust standard errors in parentheses. Reference group is Male Q1 learners that are the correct age for grade in 2018.
*** p<0.01, ** p<0.05, * p<0.1

Table A3: Regression results of the probability of dropping out in Grades 10 and 11, Limpopo DDD data 2018

	Gr10	Gr10	Gr11	Gr11
	Mathematics	Math Lit	Mathematics	Math Lit
Maths(Lit) mark2018	-0.001***	-0.002***	-0.001***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Eng mark2018	-0.002***	-0.004***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.012***	0.020***	0.012***	0.009***
	(0.002)	(0.003)	(0.002)	(0.003)
Q2	-0.001	0.005	-0.007*	-0.003
	(0.004)	(0.006)	(0.004)	(0.005)
Q3	0.006	0.012*	0.002	0.014*
	(0.004)	(0.007)	(0.005)	(0.008)
Q4&5	0.044*	0.082***	0.026**	0.084***
	(0.024)	(0.013)	(0.011)	(0.014)
Missing Q	0.043***	0.064***	0.044***	0.096***
	(0.007)	(0.014)	(0.009)	(0.018)
1 year over-age (2018)	0.005*	-0.000	0.001	-0.009***
	(0.002)	(0.003)	(0.002)	(0.003)
2 years over-age (2018)	0.017***	0.007**	0.014***	-0.006*
	(0.003)	(0.003)	(0.004)	(0.003)
3+ years over-age (2018)	0.129***	0.125***	0.107***	0.082***
	(0.006)	(0.005)	(0.005)	(0.005)
Constant	0.149***	0.261***	0.136***	0.334***
	(0.008)	(0.010)	(0.010)	(0.016)
Observations	**74 629**	**68 110**	**57 290**	**68 291**
R-squared	**0.055**	**0.097**	**0.061**	**0.105**

Robust standard errors in parentheses. Reference group is Male Q1 learners that are the correct age for grade in 2018.
*** p<0.01, ** p<0.05, * p<0.1

RESEP
Research on Socio-Economic Policy

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918 · 2018