The promise of SA-SAMS & DDD data for tracking progression, repetition and drop-out

SERVAAS VAN DER BERG
CHRIS VAN WYK
REBECCA SELKIRK
KATE RICH
NICOLA DEGHAYE

Stellenbosch Economic Working Papers: WP17/2019

www.ekon.sun.ac.za/wpapers/2019/wp172019

December 2019

KEYWORDS: South Africa, education, educational outcomes, longitudinal data, school-based assessment, school dropout, repetition
JEL: I21, I24, I25, O12, C81

ReSEP (Research on Socio-Economic Policy)
http://resep.sun.ac.za

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH

SOUTH AFRICA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

www.ekon.sun.ac.za/wpapers

The promise of SA-SAMS & DDD data for tracking progression, repetition and drop-out

Servaas van der Berg (Resep, Department of Economics, Stellenbosch University)[1]
Chris van Wyk (Resep, Department of Economics, Stellenbosch University)
Rebecca Selkirk (Resep, Department of Economics, Stellenbosch University)
Kate Rich (Resep, Department of Economics, Stellenbosch University)
Nicola Deghaye (Resep, Department of Economics, Stellenbosch University)

Abstract

This paper analyses the SA-SAMS school administration data that the Michael and Susan Dell Foundation in partnership with the Department of Basic Education collects quarterly from schools in order to assess its usefulness for better understanding the school system. The disaggregated SA-SAMS data housed in the Data Driven Districts operational data store is typically provided in the form of data dashboards for analytical purposes to the education authorities. Although only non-random samples of the data are available in longitudinal form, the analysis shows that even these samples can already divulge important relationships and features of the education system. These include the relationship between performance in earlier grades and performance in matric, the relationship between performance, repetition and subsequent dropout, the choice between Mathematics and Mathematical Literacy, and the utility of using school-based assessments in investigating later educational outcomes. The SA-SAMS data also contains much better information on the number of disabled learners in schools than previous Annual Survey of Schools (ASS or EMIS) data. Repeating and expanding such analysis in the future with lengthened longitudinal data and larger samples as data collection improves should be very fruitful for an improved understanding of the school system.

Acknowledgment: The authors wish to thank the Michael and Susan Dell Foundation for financial support and for making this data available for analysis, as part of an attempt to assess the usefulness of this data for research on the South African school system. This paper draws from a presentation to Resep's Quantitative Education Research Conference, held at Stellenbosch on 12 and 13 September 2019.

[1] *Corresponding author. Email: svdb@sun.ac.za. Postal address: Department of Economics, Stellenbosch University, Private Bag X1, Stellenbosch 7602, South Africa.*

1. INTRODUCTION

1.1 Background to this study

For the past five years, the New Leaders Foundation in collaboration with the Michael and Susan Dell Foundation has been working with the Department of Basic Education in seven provinces to collect data from the SA-SAMS administrative data system from all schools on a quarterly basis. This data is then used to create a dashboard of data on 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 of 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. This paper, as well as one by Van Biljon and Burger (2019), is the result. In this analysis, the focus falls mainly on those provinces and indeed schools for which data over a long enough period is available. The intention with this is not to present analysis for all of the school system, as the data available over the full period 2015-2018 is only a sample. However, these samples in different provinces can give us some impression of the nature of the analysis that is possible with such data. In many cases, it may also indicate what may be more general patterns in the school system.

2. THE DATA UNDERLYING THE DDD DASHBOARDS

2.1 How we can use the data

The SA-SAMS data is submitted by schools and housed into the DDD operational data store that stores disaggregated data. This underlying data store also provides longitudinal data that allows for tracking individual learners through the education system from year to year. This is useful for answering cohort-specific 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?

Aggregate data such as the Annual Survey of Schools (ASS, also often referred to as EMIS data) allows analysis of a specific cohort or pseudo-cohort, but does not give any information regarding which learners are repeating or dropping out of the school system. Unit-level data such as SA-SAMS and DDD is therefore useful in that it can provide insights into the profiles of learners who dropped out, repeated, or progressed without any repetition.

In order for unit-level data to be useful, it is important to attach a unique learner identifier to each learner, and for this unique identifier to follow a learner across grades and schools. A major limitation in the current study was that the learner identifier did not consistently follow a learner across schools, and as a result, a number of observations were lost from the sample, or erroneously recorded as dropouts. This problem was particularly acute between Grade 7 and 8, as this is a time where the majority of South African learners change schools in moving from primary to high school. For this reason, the transition from Grade 7 to Grade 8 could not be included in the longitudinal analysis.

Using only cross-sectional enrolment numbers in the same manner as in the EMIS data, it is possible to construct a pseudo-cohort of learners and track this pseudo-cohort across years. This pseudo-cohort does not include the exact same learners, owing to repetition, but it gives a general idea of movement through the school system. An example of this is given in Table 1, which shows enrolment in a Gauteng sample of schools from SA-SAMS, over a four-year period. A pseudo-cohort can be tracked by following the diagonal, as in the highlighted fields: the 2015 Grade 9 cohort should have progressed to Grade 10 in 2016, Grade 11 in 2017, and Grade 12 in 2018. Changes in the enrolment numbers in each of these years can give an indication of repetition and dropout. The large increase in enrolment numbers between Grade 9 and 10 indicates that there is a high rate of repetition in Grade 10, resulting in more than one cohort "parking" in this grade. The subsequent large drops in Grade 11 and 12 enrolment indicate a combination of high repetition in earlier grades, and dropout.

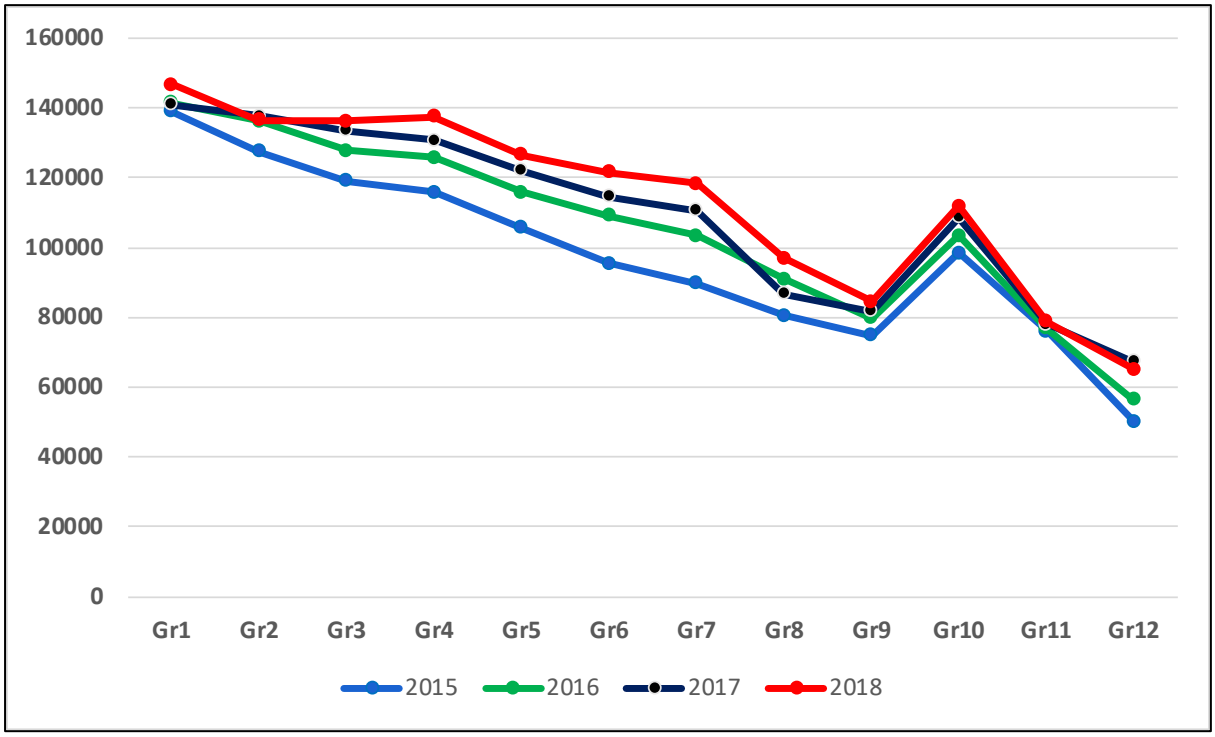
Table 1: Enrolment in a Gauteng sample of schools based on SA-SAMS data, 2015-2018

	Gr1	Gr2	Gr3	Gr4	Gr5	Gr6	Gr7	Gr8	Gr9	Gr10	Gr11	Gr12
2015	138 891	127 496	118 977	115 773	105 655	95 438	89 757	80 416	74 832	98 413	76 011	50 087
2016	141 64	136 297	127 701	125 629	115 702	109 076	103 333	90 961	79 770	103 471	76 726	56 508
2017	142 027	137 467	133 467	130 667	121 898	114 478	110 591	86 816	81 789	108 644	78 024	67 401
2018	146 652	136 169	136 169	137 371	126 541	121 489	118 305	96 865	84 505	111 826	78 864	65 157

Source: Own calculations using SA-SAMS data

By simply considering enrolment numbers across the Gauteng sample in a manner similar to using cross-sectional data, a clear pattern emerges that is relatively stable across time, although the schools in the sample appear to experience moderate enrolment growth over time. As illustrated by Figure 1, the South African school system does a relatively good job of retaining learners in the school system throughout primary school, but there is high dropout and repetition in secondary school.

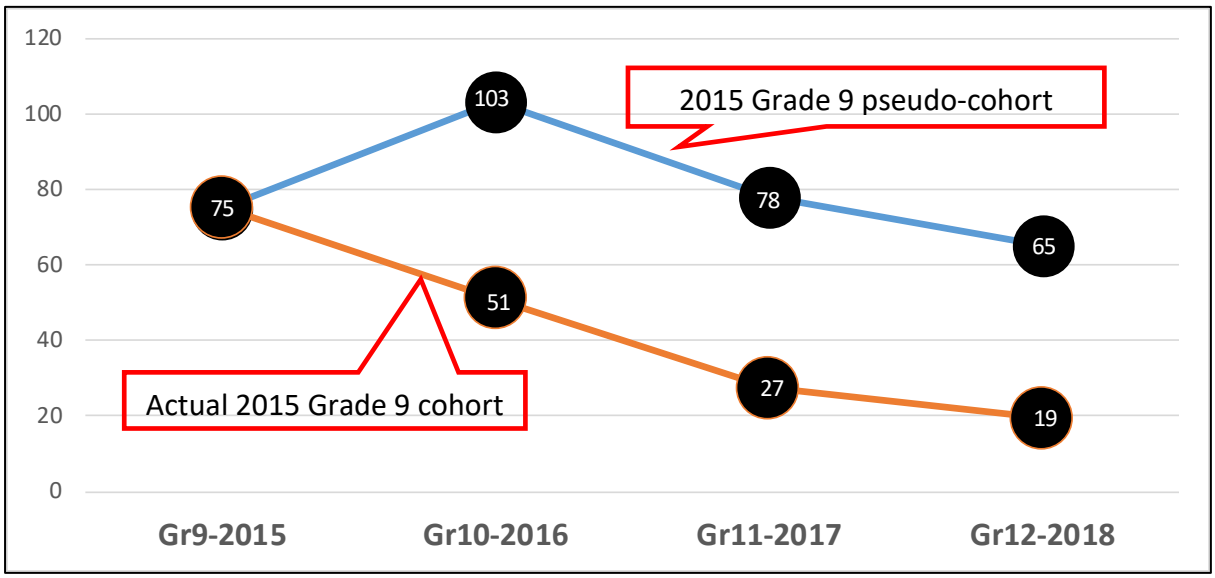
Figure 1: Cross-sectional analysis of SA-SAMS data (2015-2018): Gauteng sample



Source: Own calculations using SA-SAMS data

With unit-level data, it is possible to track an actual cohort of learners through the school system, and to compare this to the pseudo-cohort constructed from repeated cross-sections of the data. This gives a better idea of grade progression, as illustrated in Figure 2. The pseudo-cohort shows that in Grade 9 in 2015 there were approximately 75 000 learners, and this number dropped to approximately 65 000 learners who were in Grade 12 in 2018, the year in which Grade 9 2015 learners should have reached Grade 12 if they had not repeated or dropped out. Contrasting this with the actual 2015 cohort of Grade 9 learners, i.e. retaining only those who had been in Grade 9 in 2015 and had passed each subsequent year without repeating, it can be seen that of the initial 75 000 learners, only 19 000 reached Grade 12 without dropping out or repeating. The remaining approximately 46 000 learners of the 65 000 2018 Grade 12 enrolment are therefore repeaters from earlier cohorts in Grade 10, 11, or 12 in one of the years between 2016 and 2018.

Figure 2: Pseudo vs actual 'on track' grade progression for the 2015 Grade 9 cohort: Gauteng



Source: Own calculations using SA-SAMS data

From Table 2 it is clear that learner unit-level data offers a far more accurate and complete picture of on-track learners than aggregate or pseudo-cohort data. The table shows the progression of the 2015 Grade 9 cohort through the school system between 2015 and 2018. 'On track' learners are those on the highlighted diagonal, while those above the diagonal represent repeaters from that cohort who are still in the school system. Numbers below the diagonal are data errors, such as learners recorded as being in Grade 9 in 2015 and, for example, Grade 11 or 12 in 2016. It is clear here that the national progression policy prohibiting repetition more than once in a phase is not strictly adhered to, with 199 learners still in Grade 9 in 2018, for at least four consecutive year. The high dropout in the three years following Grade 9 is also apparent, with nearly 25 000 of the initial 75 000 learners dropping out of the sample between 2015 and 2018. (As this is a sample of schools, it is possible that some of such 'dropout' was not really dropout from the school system, but children moving to other provinces, or to schools not in the sample, including private schools.)

Table 2: Progression of learners that enrolled in Grade 9 in 2015 for the period 2015-2018: Gauteng

	2015	2016	2017	2018
Grade 9	75 259	10 536	1 158	199
Grade 10	0	51 001	27 421	10 021
Grade 11	0	885	27 414	21 149
Grade 12	0	19	625	19 453
Still in school	75 259	62 441	56 618	50 822
Repeaters	0	10 536	28579	31 369
Dropout	0	12 818	5 823	5 796
Cumulative dropout	0	12 818	18 641	24 437

Source: Own calculations using SA-SAMS data

2.2 How representative are the samples?

For cohort analysis it is important to track the same schools over time, and therefore the sample was reduced to those schools that submitted data in cases where it was useful to track learners over four years.

When viewing the results, it is important to note that these results are not representative of the broader population because of sample-selection issues. Specifically, far fewer schools submitted SA-SAMS data in 2015 than in later years. Table 3 illustrates this. It compares the number of Gauteng schools that submitted data each year, compared to the number of schools in EMIS in 2018. Quintile 5 schools were particularly underrepresented: The sample that could be used to track learners across all the years from 2015 to 2018 was only 47% of all Gauteng schools, but even much lower at 26% in Quintile 5 Gauteng schools. Where data from fewer years is used, those schools represented in all the most recent relevant years were included.

Table 3: Number of Gauteng schools by year and quintile: Actual (EMIS 2018) and those in the DDD sample (2015-2018)

Year	Q1	Q2	Q3	Q4	Q5	Quintile unknown	Total
2015	205	198	342	351	174	111	1 381
2016	260	249	401	444	407	293	2 054
2017	262	243	405	462	563	446	2 381
2018	269	256	411	474	630	489	2 529
Same schools (2015-2018)	**202**	**195**	**340**	**348**	**168**	**99**	**1 352**
Same school sample as % of all schools	73%	76%	82%	73%	26%	12%	47%
All schools in EMIS:2018	275	257	416	477	641	814	2 880

Source: Own calculations using SA-SAMS and EMIS data

3. PROGRESSION PATTERNS

Using the unit-level DDD data, it is possible to investigate progression patterns and to profile dropouts and the correlates of dropout and repetition.

3.1 Progression without dropout or repetition

This section uses 2016 to 2018 data for the sample of schools that appears in each of these three years in order to do this. The regressions "predict" the probability of successful progression over a two-year period with zero repetition or dropout, using simple OLS Linear Probability Models. Table 4 considers progression from Grade 5 to Grade 7. In all three provinces, higher Grade 5 Mathematics and also English First Additional Language (EFAL) marks are significantly associated with higher probability of reaching Grade 7 without repetition or dropout. A one percentage point increase in these marks leads to an approximately 0.3 to 0.5 percent increase in the probability of reaching Grade 7 without repetition or dropout. After controlling for these marks, Grade 5 marks in Natural Science and Social Science have smaller effects and are non-significant in Gauteng. Learners who are absent more frequently in Grade 5 are more likely to repeat or to drop out of the sample, while school quintile does not appear to play a large or consistent role after controlling for other factors. In the Eastern Cape and Limpopo, girls are more likely to progress without repeating or dropping out, and across all provinces, learners who are already three or more years over-age in Grade 5 are significantly less likely to progress without further repetition or dropout by Grade 7. The positive coefficients in some cases for over-age may relate to the fact that the official policy is that learners should not repeat more than once in a school phase, so this may lead to some over-age children being less likely to repeat in some schools.

Table 4: Probability of progression without repetition from Grade 5 to Grade 7 for three provincial samples, 2016 Grade 5 cohort

	Outcome: P(No dropout/repetition Gr5-7)		
	GP	EC	LIM
Gr5 Maths	0.0029***	0.0048***	0.0040***
	(0.0005)	(0.0003)	(0.0002)
Gr5 EFAL	0.0039***	0.0046***	0.0046***
	(0.0005)	(0.0004)	(0.0003)
Gr 5 Nat Science	0.0008	0.0014***	0.0014***
	(0.0005)	(0.0004)	(0.0002)
Gr 5 Social Science	0.0003	0.0010***	0.0011***
	(0.0006)	(0.0003)	(0.0002)
Days absent (2016)	-0.0028***	-0.0051***	-0.0050***
	(0.0006)	(0.0008)	(0.0009)
Reference: Q1-3			
Quintile 4	0.0192*	0.0030	0.0335
	(0.0110)	(0.0253)	(0.0339)
Quintile 5	0.0203	-0.0492*	0.0317
	(0.0219)	(0.0294)	(0.0202)
Female	-0.0036	0.0162***	0.0130***
	(0.0036)	(0.0038)	(0.0026)
Reference: Correct age			
1 year over-age	-0.0143**	0.0159***	0.0102**
	(0.0058)	(0.0047)	(0.0040)
2 years over-age	-0.0384***	0.0017	0.0047
	(0.0102)	(0.0069)	(0.0072)
3+ years over-age	-0.1103***	-0.0790***	-0.0314***
	(0.0191)	(0.0092)	(0.0107)
Constant	0.3463***	0.1945***	0.1987***
	(0.0260)	(0.0212)	(0.0192)
Observations	63,285	63,047	95,498
R-squared	0.0943	0.1382	0.1580

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

Table 5 examines progression from Grade 9 to 11. With the exception of Social Science in the Eastern Cape, Grade 9 performance in Mathematics, EFAL, Life Orientation and Social Science are all positively associated with successful progression. Being absent more often from school in Grade 9 is negatively associated with progression, though this is not significant in Gauteng. Being in a quintile 5 school in the Eastern Cape and Limpopo strongly and significantly increases the probability of reaching Grade 11 from Grade 9 without repeating or dropping out. Girls are significantly more likely to repeat or drop out in these grades in both Gauteng and the Eastern Cape. In secondary school, over-age learners are far more likely to drop out or repeat between Grade 9 and 11 than is the case for the end of primary school (as will be shown), and this probability increases the more years a learner is over-age.

Table 5: Probability of progression without repetition from Grade 9 to Grade 11 for three provincial samples, 2016 Grade 9 cohort

	Outcome: P(No dropout/repetition Gr9-11)		
	GP	EC	LIM
Gr9 Maths	0.0022***	0.0033***	0.0044***
	(0.0007)	(0.0005)	(0.0004)
Gr9 EFAL	0.0048***	0.0065***	0.0055***
	(0.0008)	(0.0005)	(0.0004)
Gr9 LO	0.0056***	0.0060***	0.0048***
	(0.0008)	(0.0005)	(0.0005)
Gr 9 Social Science	0.0045***	0.0001	0.0061***
	(0.0008)	(0.0005)	(0.0004)
Days absent (2016)	-0.0010	-0.0031***	-0.0021***
	(0.0008)	(0.0007)	(0.0007)
Reference: Q1-3			
Quintile 4	-0.0073	-0.0151	0.0283
	(0.0306)	(0.0508)	(0.0262)
Quintile 5	-0.0273	0.1594***	0.1594***
	(0.0502)	(0.0165)	(0.0227)
Female	-0.0146**	-0.0217***	-0.0042
	(0.0061)	(0.0046)	(0.0037)
Reference: Correct age			
1 year over-age	-0.0522***	-0.0879***	-0.0913***
	(0.0072)	(0.0061)	(0.0052)
2 years over-age	-0.0687***	-0.1593***	-0.1276***
	(0.0100)	(0.0079)	(0.0064)
3+ years over-age	-0.0573***	-0.2170***	-0.1303***
	(0.0122)	(0.0089)	(0.0075)
Constant	-0.4898***	-0.1646***	-0.4579***
	(0.0585)	(0.0334)	(0.0417)
Observations	45,843	50,373	91,652
R-squared	0.2168	0.2052	0.3044

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

Table 6 examines progression from Grade 10 to 12. A "Mathematics Equivalent" score is calculated by multiplying the mark for Mathematical Literacy, which is less demanding than Mathematics, by 0.75 and retaining the Mathematics mark unchanged.[2]

Compared to earlier grades, a learner's Mathematics, EFAL and Life Orientation marks in Grade 10 appear to be more strongly associated with progression without repetition or dropout in these grades. Higher absenteeism in Grade 10 is associated with a lower probability

[2] *Up to 2007, Mathematics could be taken at Higher or Standard Grade. At that time the former was given a weight of 400 marks in the determination of the aggregate mark, and the latter only 300 marks. The "equivalent" mark calculated here follows the same pattern, though (Shepherd and Van der Berg 2019) applied a more sophisticated methodology of (Simkins 2013) based on estimating "correspondence scores" for these two subjects in 2010 derived from the performance in these two subjects of students at similar points in the aggregate points distribution in matric.*

of successful progression, while the effect of school quintiles is not consistently clear. Girls are less like to progress without repetition or dropout, and the effects of being over-age in Grade 10 are unclear, except in the Eastern Cape where over-age learners are consistently less likely to progress without repetition or dropout, regardless of how many years they are over-age at the start of Grade 10.

Table 6: Probability of progression without repetition from Grade 10 to Grade 12 for three provincial samples, 2016 Grade 10 cohort

	Outcome: P(No Rep/Drop Gr10-12)		
	GP	EC	LIM
Gr10 Maths Equivalent	0.0101***	0.0083***	0.0093***
	(0.0006)	(0.0006)	(0.0004)
Gr10 EFAL	0.0086***	0.0106***	0.0107***
	(0.0005)	(0.0005)	(0.0003)
Gr10 LO	0.0062***	0.0052***	0.0047***
	(0.0005)	(0.0006)	(0.0004)
Days absent (2016)	-0.0021***	-0.0019**	-0.0024***
	(0.0005)	(0.0007)	(0.0004)
Reference: Q1-3			
Quintile 4	-0.0030	-0.1350**	0.0025
	(0.0170)	(0.0544)	(0.0506)
Quintile 5	-0.0665**	0.0860*	0.0710**
	(0.0283)	(0.0465)	(0.0316)
Female	-0.0190***	-0.0252***	-0.0196***
	(0.0051)	(0.0051)	(0.0029)
Reference: Correct age			
1 year over-age	0.0110	-0.0347***	0.0043
t	(0.0071)	(0.0061)	(0.0053)
2 years over-age	0.0099	-0.0453***	0.0084
	(0.0087)	(0.0084)	(0.0066)
3+ years over-age	0.0246**	-0.0623***	-0.0048
	(0.0098)	(0.0100)	(0.0070)
Constant	-0.4483***	-0.4481***	-0.5013***
	(0.0295)	(0.0397)	(0.0334)
Observations	57,870	65,202	136,252
R-squared	0.2812	0.2678	0.3062

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

3.2 The effect of Grade 10 repetition on dropout between Grade 10 and Grade 11

The regressions in Table 7 include only learners who were in Grade 10 in 2017, in order to determine whether those who are repeating Grade 10 in 2017 are more likely to drop out between Grade 10 and 11 than those who are in Grade 10 for the first time in 2017. The outcome variable is the probability of dropping out without progressing or repeating.

In all provinces, those repeating Grade 10 in 2017 appear more likely to drop out by 2018, although, surprisingly, this is not statistically significant in Gauteng. Higher absenteeism appears to be associated with higher probability of dropout, with Term 3 absenteeism appearing particularly problematic. There is some endogeneity resulting from the way the data is reported, though: If someone left school before the end of the year, no absenteeism may be recorded for them in the terms where they were no longer in school. School quintile is only significantly associated with dropout in the Eastern Cape, where learners in richer schools are less likely to drop out even after controlling for other factors. Consistent with the progression findings, over-age learners are more likely to drop out, with the probability of dropping out increasing the more years a learner is over-age. Also consistent with the findings on progression, girls are more likely to drop out between Grade 10 and Grade 11.

Table 7: Probability of dropping out before Grade 11 for learners in Grade 10 in 2017 in three provincial samples

	Outcome: P(dropout after Grade 10)		
	GP	EC	LIM
Gr10 repeater	0.0176	0.0354***	0.0708***
	(0.0120)	(0.0071)	(0.0030)
Days absent (Term 1)	0.0013	0.0039***	0.0061***
	(0.0029)	(0.0012)	(0.0020)
Days absent (Term 2)	0.0051	0.0038***	0.0006
	(0.0054)	(0.0012)	(0.0008)
Days absent (Term 3)	0.0092***	0.0116***	-0.0067***
	(0.0019)	(0.0009)	(0.0007)
Days absent (Term 4)	-0.0052	0.0077***	-0.0151***
	(0.0033)	(0.0013)	(0.0019)
Reference: Q1-3			
Quintile 4	0.0213	-0.0400**	0.0124
	(0.0476)	(0.0195)	(0.0159)
Quintile 5	-0.0269	-0.0457**	-0.0036
	(0.0511)	(0.0186)	(0.0278)
Female	0.0170**	0.0181***	0.0155***
	(0.0066)	(0.0032)	(0.0019)
Reference: Correct age			
1 year over-age	0.0472***	0.0197***	0.0045**
	(0.0071)	(0.0039)	(0.0023)
2 years over-age	0.1091***	0.0692***	0.0115***
	(0.0108)	(0.0055)	(0.0027)
3+ years over-age	0.2498***	0.1902***	0.1603***
	(0.0155)	(0.0075)	(0.0041)
Constant	0.3449***	0.0285**	0.0772***
	(0.0849)	(0.0130)	(0.0130)
Observations	62,028	54,427	139,811
R-squared	0.1514	0.1093	0.0928

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

3.3 Over-age and enrolment patterns

The data is able to shed some light on the patterns of enrolment, repetition and dropout in the SA school system.

Table 8 illustrates some of the issues in working with administrative data. Several learners are recorded as being far too young for their grade – for example, learners younger than 10 supposedly in Grade 9. This should also lead one to doubt whether the over-age figures are entirely accurate, but the story is likely to remain the same – the number of over-age learners is very large, even if it is slightly overestimated due to administrative error.

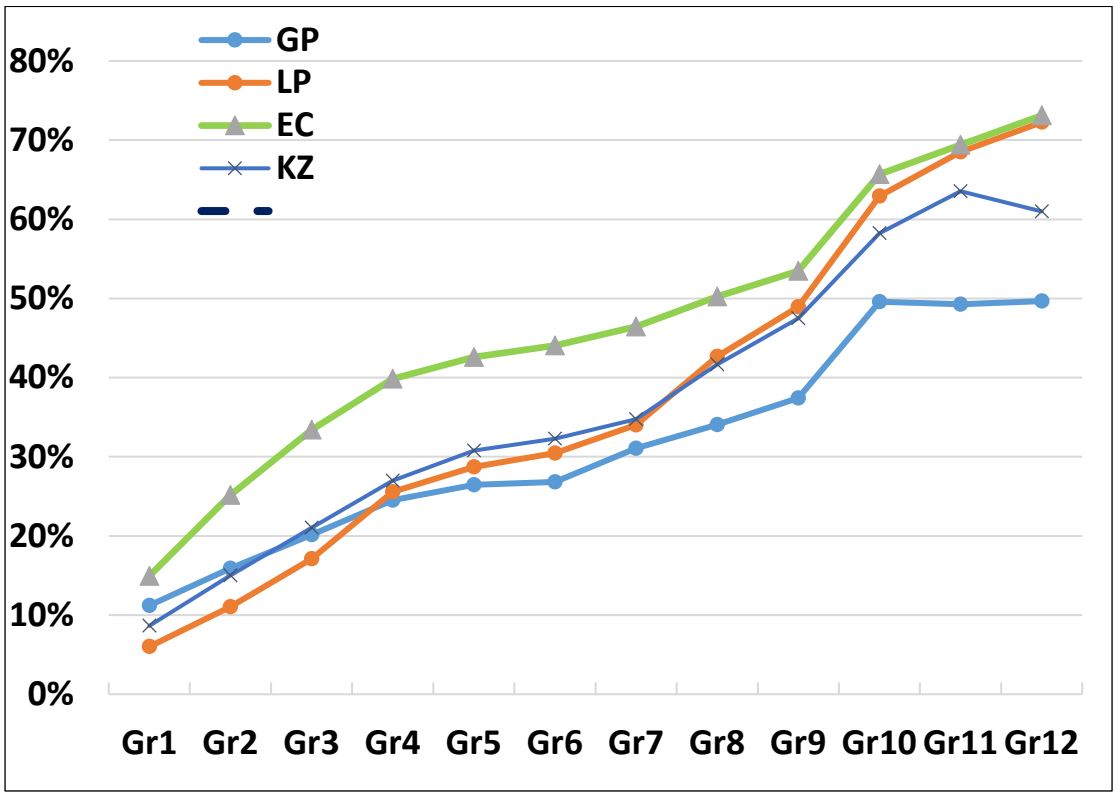
The share of over-age learners rises steadily with each grade, as can also be seen in Figure 3. This reflects the fact that the percentage is cumulative – learners remain over-age once they have repeated once, and the percentage would only drop if learners drop out. The percentage of learners that is over-age keeps on rising until Grade 10 for Gauteng and Grade 11 for Kwazulu-Natal, suggesting some dropout in higher grades, but keeps rising continuously for the other provinces. More than 70% of Eastern Cape learners are over-age by Grade 12, compared to 50% in Gauteng. In neighbouring Botswana, in contrast, the percentage of learners who are over-age actually declines after Grade 10, due to the fact that that country's high stakes Grade 10 examination forces many weaker-performing learners to drop out of school.

Table 8: Age by grade in Eastern Cape sample, 2018

Age	Gr1	Gr2	Gr3	Gr4	Gr5	Gr6	Gr7	Gr8	Gr9	Gr10	Gr11	Gr12
<4	38	5	12	8	3	1	2	2	1			
5	834	26	15	6	3		7	2	1	2		
6	58 950	520	22	22	11	7	3	3	3			
7	97 641	41 296	511	33	6	4	4	3	1			
8	22 388	86 238	32 450	458	20	14	4	8	3	1		1
9	3 163	32 137	79 253	29 270	517	30	21	6	2	2	2	1
10	854	7 443	39 429	75 856	27 473	598	31	8	5	3	3	1
11	421	1 688	10 960	42 650	66 368	25 762	657	16	4	4	1	1
12	261	670	3 175	17 448	39 869	61 874	24 888	690	25	8	4	
13	166	376	1 010	6 069	18 368	38 249	56 629	22 029	741	17	8	2
14	130	245	503	2 257	7 107	18 387	36 436	49 605	19 402	677	9	5
15	77	147	223	823	2 661	7 495	18 652	31 871	40 963	15 159	482	18
16	54	77	205	380	1 077	3 119	8 649	19 388	29 000	36 191	10 271	472
17	48	60	197	179	468	1 274	4 192	10 959	18 566	31 191	26 933	7 704
18	24	48	170	107	262	619	1 987	6 022	11 503	25 456	24 816	23 491
19	30	46	144	56	139	218	840	2 861	6 385	19 012	21 528	24 425
20	14	35	107	20	55	90	323	1 094	2 891	11 933	16 289	21 431
21	13	27	95	12	34	45	139	509	1 201	6 418	10 855	16 560
22	7	24	59	3	13	22	62	207	439	3 145	6 379	11 503
23	5	16	45	6	13	10	20	75	184	1 298	3 057	6 199
24	8	9	34	1	4	6	8	46	85	565	1 432	3 195
25	1	3	6	1	2	3	3	23	28	217	617	1 388
25+	10	3	10	4	5	-	8	24	56	412	646	1 514
Total	**185 137**	**171 139**	**168 635**	**175 669**	**164 478**	**157 827**	**153 565**	**145 451**	**131 489**	**151 711**	**123 332**	**117 911**
Over-age	27 674	43 054	56 372	70 016	70 077	69 537	71 319	73 079	70 338	99 647	85 619	86 215
% over-age	**15%**	**25%**	**33%**	**40%**	**43%**	**44%**	**46%**	**50%**	**53%**	**66%**	**69%**	**73%**

Correct age in grade 1 taken to be 6 or 7 years, over-age 8 years or more. Ages for higher grades follow the same pattern.
Source: Own calculations using SA-SAMS data

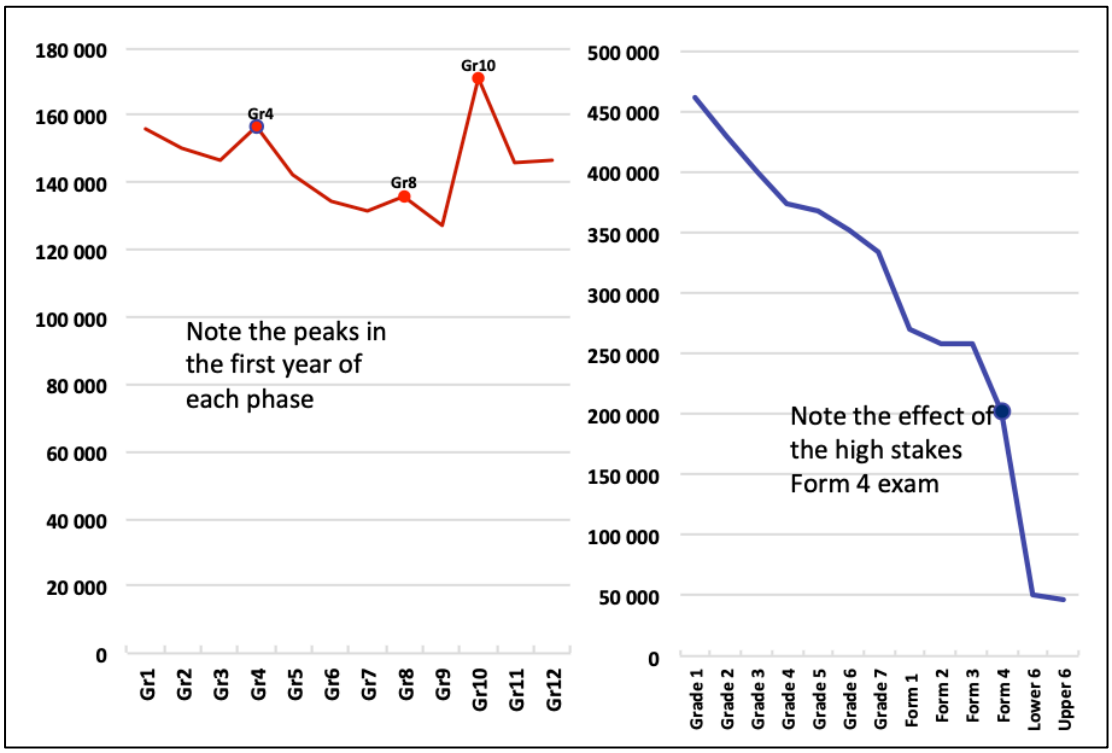
Figure 3: Over-age by grade and province 2018 in three provincial samples, 2018, with a comparison to Botswana in 2014



Source: Own calculations using SA-SAMS data

There is a noticeable jump in over-age proportions in Grade 10. This result from the repetition policy, which limits repetition to once per school phase, after which learners are supposed to progress automatically. The effect of the repetition policy is even more pronounced in the clear peaks in Grade 4, Grade 8 and Grade 10 in Figure 4 – the first year of each phase. In contrast, as shown in the right-hand panel, Zimbabwe has a very different enrolment pattern, with a steady decline in enrolment in each grade throughout their school system, and a particularly large drop in enrolment after the high stakes examination in Form 4.

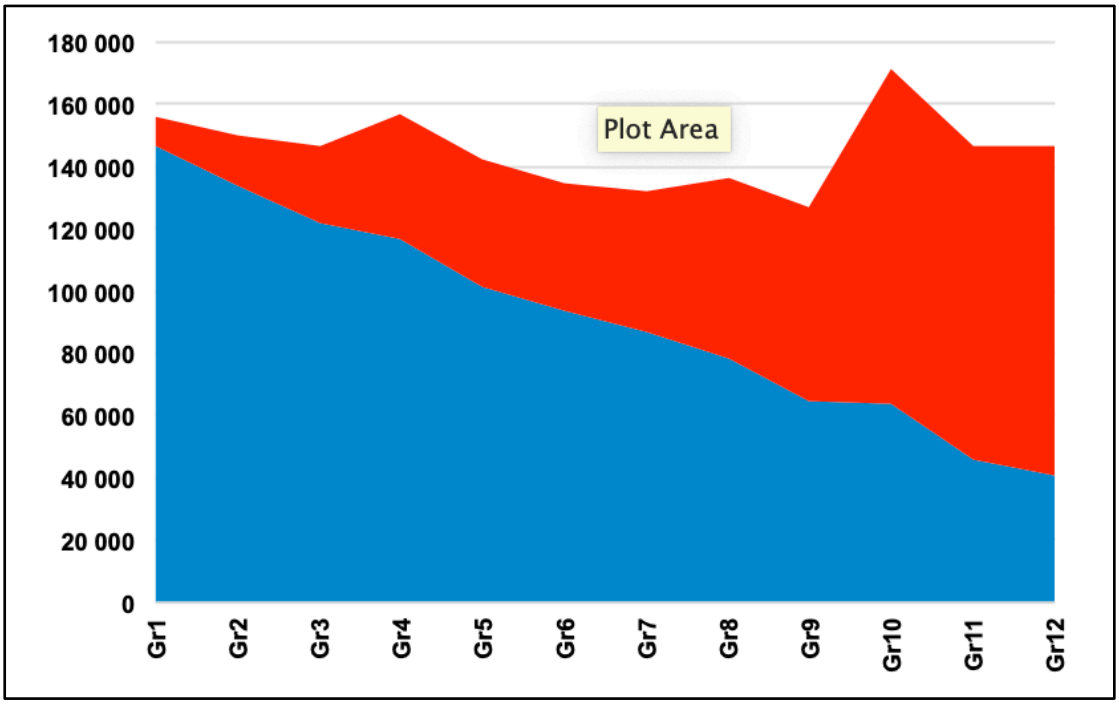
Figure 4: Comparing enrolment patterns in the Limpopo sample with Zimbabwe, 2018



Source: Own calculations using SA-SAMS data

The peaks in Grades 4, 8 and 10 are also evident in the figure below for the Eastern Cape.

Figure 5: Total enrolment, showing over-age, for the Eastern Cape sample, 2018

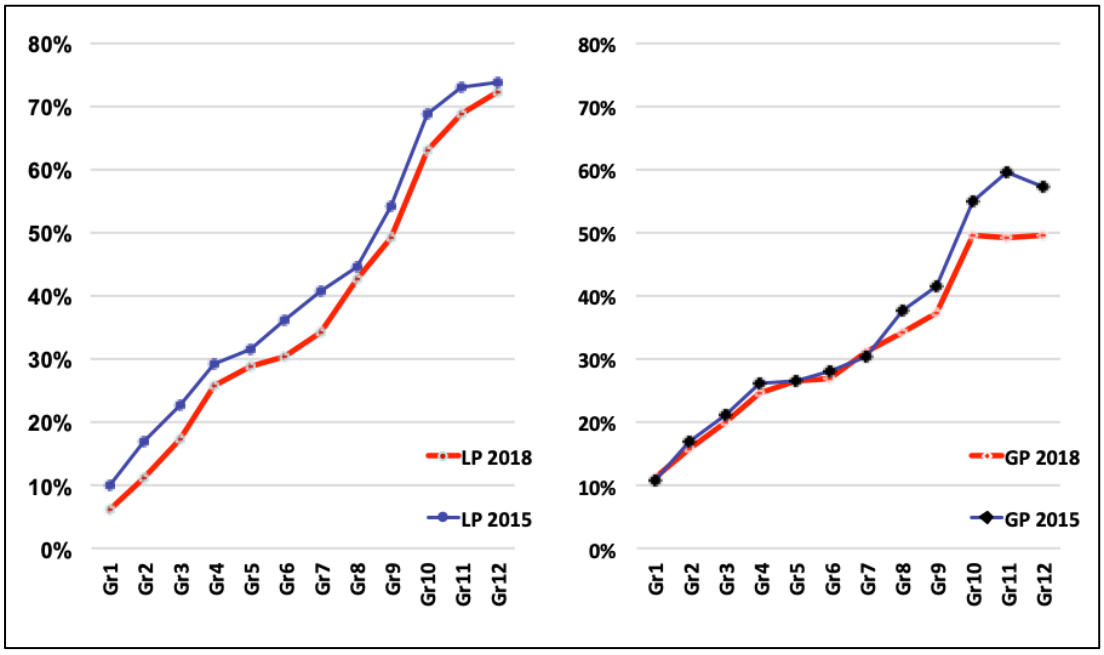


Source: Own calculations using SA-SAMS data

Figure 6 shows the enrolment patterns in Gauteng and Limpopo. It seems that in recent years repetition has been declining slightly across all grades in Limpopo. In Gauteng the proportion of learners who are over-age has fallen more drastically in the later grades. As seen in Source: Own calculations using SA-SAMS data

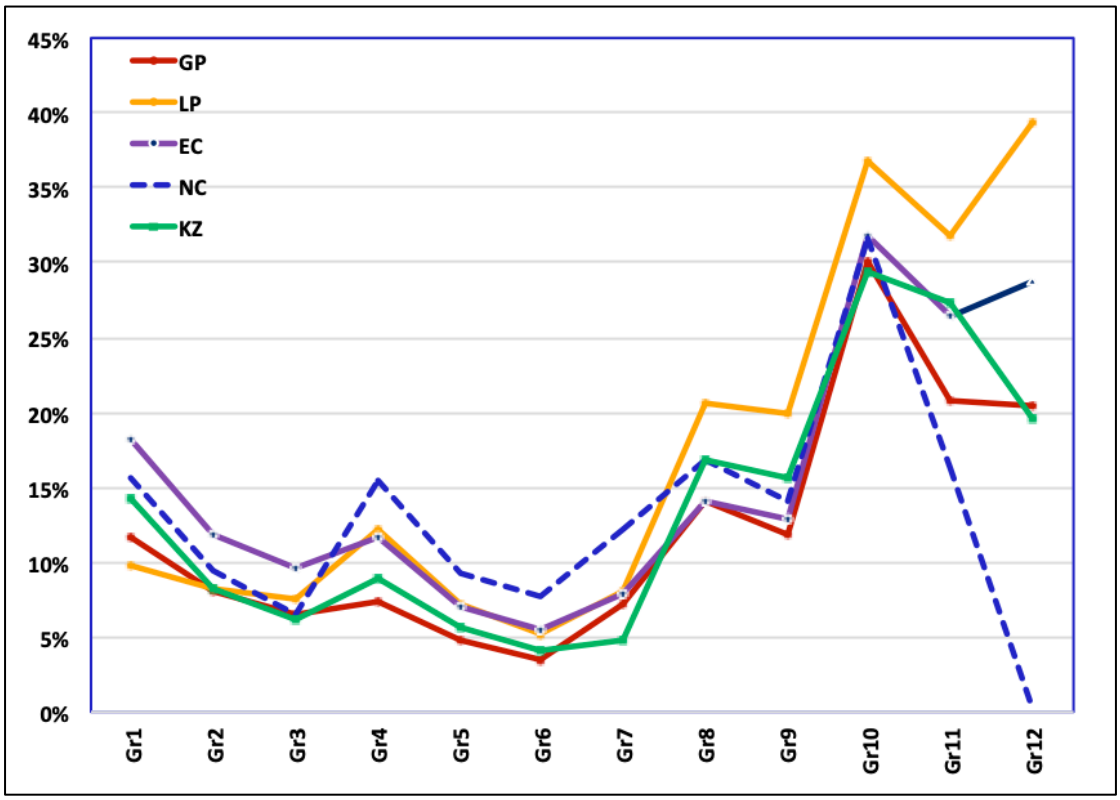
Figure 7, Limpopo has similar repetition rates to the other provinces in the lower grades, but its repetition rate rises above the others in the later grades. The automatic promotion policy is again evident – once learners have repeated Grade 10, according to the promotion policy they should automatically be promoted from Grade 11 to 12 even if they have not passed.

Figure 6: Over-age in the Gauteng and Limpopo samples, 2015 and 2018



Source: Own calculations using SA-SAMS data

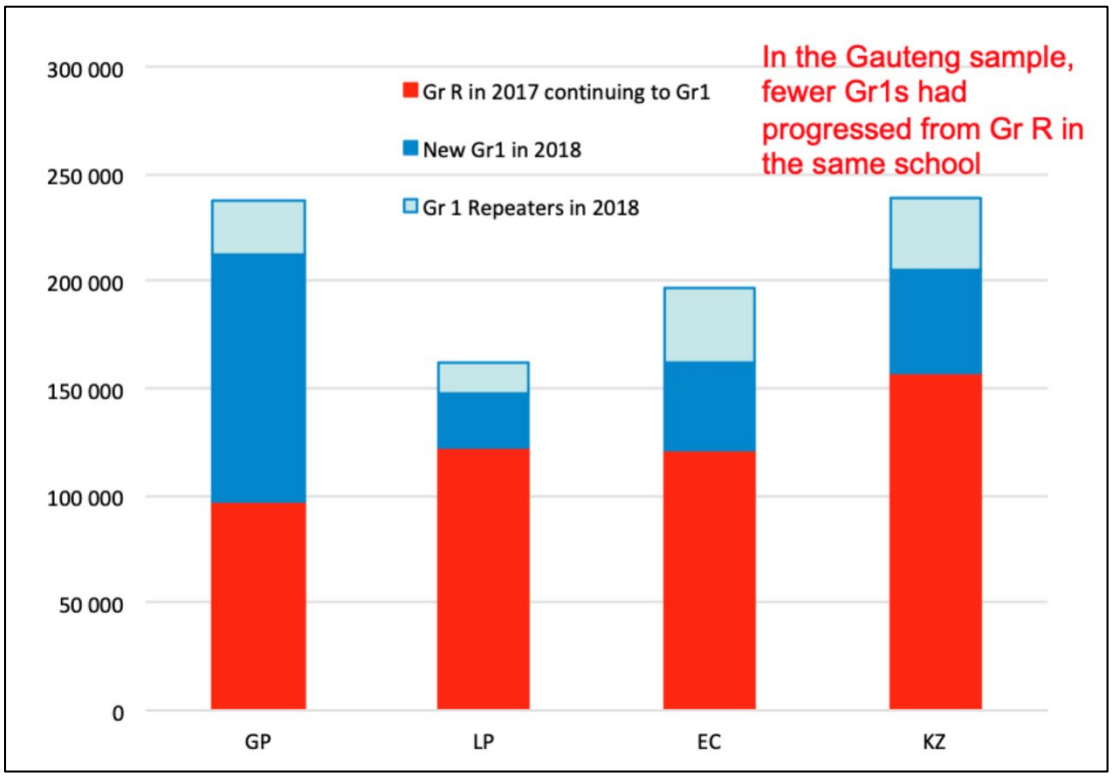
Figure 7: Repetition rates by grade in five provincial samples, 2017



Source: Own calculations using SA-SAMS data

The patterns of enrolment in Grade 1 again differ across the provinces. Gauteng appears to have a particularly high proportion of new Grade 1s who had not previously been captured in the SA-SAMS (Figure 8). This may be because many learners attend community-based Grade R, so they are not captured in the school-based SA-SAMS data before Grade 1. This has implications for funding and subsidization of these learners – Grade R is subsidised if a learner attends a school-based Grade R programme. This also provides access to the school feeding programme, funded by the state, which does not apply to community-based Early Childhood Development Centres.[3]

Figure 8: Composition of Grade 1 enrolment in four provincial samples, 2018



Source: Own calculations using SA-SAMS data

4. CAN SCHOOL-BASED ASSESSMENT PREDICT PERFORMANCE AND PROGRESSION?

4.1 Does school-based assessment predict future performance?

The regressions in this sub-section provide an indication of whether current performance can predict future performance across a two-year period. This is useful as it offers insight into the usefulness and consistency of school-based assessments. For all grades except grade 12, performance is taken to be the subject mark in Term 4. For matric, the Term 3 results are used, as the DDD data does not consistently include matric examination (NSC) marks.

Table 9 shows how Grade 5 performance can be used to predict Grade 7 Mathematics performance. The strong positive association between Grade 5 and Grade 7 Maths performance points to encouraging consistency in school-based assessments between grades. Higher Grade 5 performance

[3] *Registered community-based ECD centres do receive a government subsidy, but this only covers part of the cost of providing community-based ECD and is not explicitly linked to provision of food.*

in EFAL, Natural Science and Social Science are also all strongly positively associated with Grade 7 Maths performance, though understandably the coefficients for Grade 5 Maths performance is much higher, showing that this acts as the best predictor. Absenteeism in Grade 7 is negatively associated with Maths performance, while school quintile is not significantly associated with performance once after controlling for other factors, with the exception of quintile 4 learners in Limpopo outperforming learners in quintile 1 to 3 schools. In all provinces, girls outperform boys after controlling for their Grade 5 performance, and learners do worse the more they are over-age at the start of Grade 5.

Table 9: Regressions of 2018 Grade 7 Mathematics performance based on 2016 Grade 5 performance in three provincial samples

	Outcome: Gr7 Maths mark		
	GP	EC	LIM
Gr5 Maths	0.3617***	0.2725***	0.3143***
	(0.0242)	(0.0155)	(0.0158)
Gr5 EFAL	0.1015***	0.1261***	0.0846***
	(0.0260)	(0.0186)	(0.0161)
Gr 5 Natural Science	0.1465***	0.0872***	0.1342***
	(0.0240)	(0.0153)	(0.0164)
Gr 5 Social Science	0.1204***	0.0800***	0.1131***
	(0.0226)	(0.0131)	(0.0144)
Absenteeism (2018):			
Days absent (Term 1)	-0.3373***	-0.1240**	-0.2339***
	(0.0722)	(0.0521)	(0.0737)
Days absent (Term 2)	-0.1292**	-0.2447***	-0.1866***
	(0.0610)	(0.0561)	(0.0539)
Days absent (Term 3)	-0.2494***	-0.1466***	-0.2116***
	(0.0456)	(0.0505)	(0.0431)
Days absent (Term 4)	0.0410	-0.3916***	-0.2876***
	(0.0737)	(0.0620)	(0.0682)
Reference: Q1-3			
Quintile 4	0.0238	1.9585	6.3526***
	(0.7446)	(2.9704)	(1.2392)
Quintile 5	0.2007	1.1212	3.4847
	(1.3934)	(2.2899)	(2.6115)
Female	1.3936***	1.0841***	1.2592***
	(0.1514)	(0.1213)	(0.1132)
Reference: Correct age			
1 year over-age	-1.1777***	-1.5147***	-1.5805***
	(0.1792)	(0.1391)	(0.1536)
2 years over-age	-1.6752***	-1.8683***	-2.2910***
	(0.3207)	(0.2130)	(0.2367)
3+ years over-age	-2.5495***	-3.1419***	-3.4175***
	(0.5872)	(0.2940)	(0.3337)
Constant	14.7602***	18.1316***	15.4571***
	(1.5592)	(0.9782)	(1.6238)
Observations	49,853	49,356	76,794
R-squared	0.4599	0.3406	0.3955

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

At the end of Grade 9, learners have to choose between taking Mathematics and Mathematical Literacy in Grade 10. As a result of this, when considering the Grade 9 predictors of Grade 11 performance, the aforementioned Mathematics Equivalent score is used as the outcome variable (Table 10). It should be remembered, though, that there is further sample selection bias in these results, as these can only be shown for learners who have successfully progressed from Grade 9 in 2016 to Grade 11 in 2018, grade levels in which dropout is endemic.

Compared to primary school, the association between Mathematics performance between grades is smaller, but there is nevertheless a significant and positive association. Grade 9 EFAL, Life Orientation and Social Science performance are also all positively associated with higher Grade 11 Mathematics performance after controlling for the Grade 9 marks, though once again the coefficients are smaller than for Mathematics performance in Grade 9. All significant absenteeism effects indicate that higher Grade 11 absenteeism is associated with poorer Mathematics performance, particularly in the third term of Grade 9, while learners in quintile 4 or 5 schools in the Eastern Cape and Limpopo outperform learners in quintile 1 to 3 schools. (The weaker performance of quintile 5 schools after controlling for other factors in Gauteng is surprising.) Girls perform worse than boys in Mathematics in secondary schools once other factors are controlled for, which could be reinterpreted as that girls in this age group find it more difficult to maintain their Maths performance (though the low weight of 0.75 given to the Maths Literacy mark in calculating the Maths Equivalent mark may play a role). Older learners perform significantly worse in Mathematics, but there is not a large difference according to how many years a learner is over-age in Grade 9, for those who do successfully progress to Grade 11.

Table 10: Regressions of 2018 Grade 11 Mathematics Equivalent performance based on 2016 Grade 9 performance in three provincial samples

	Outcome: Grade 11 Maths Equivalent mark		
	GP	EC	LIM
Gr9 Maths	0.2556***	0.3400***	0.2808***
	(0.0235)	(0.0279)	(0.0198)
Gr9 EFAL	0.1862***	0.1545***	0.1774***
	(0.0235)	(0.0216)	(0.0202)
Gr9 LO	0.0821***	0.1664***	0.1461***
	(0.0255)	(0.0227)	(0.0204)
Gr 9 Social Science	0.2121***	0.0447**	0.0912***
	(0.0232)	(0.0213)	(0.0207)
Absenteeism (2018):			
Days absent (Term 1)	0.0009	-0.2237***	-0.2381**
	(0.0810)	(0.0665)	(0.1008)
Days absent (Term 2)	-0.1934***	-0.2473***	-0.1137
	(0.0717)	(0.0665)	(0.1051)
Days absent (Term 3)	-0.2457***	-0.2949***	-0.2636***
	(0.0479)	(0.0440)	(0.0572)
Days absent (Term 4)	-0.0813	-0.1481**	0.2414
	(0.0671)	(0.0689)	(0.1478)
Reference: Q1-3			
Quintile 4	0.7682	3.0611	5.1170***
	(0.7062)	(2.2587)	(1.1310)
Quintile 5	-4.5532***	4.2621***	7.6959***
	(1.3246)	(0.5703)	(1.0801)
Female	-3.9751***	-2.9011***	-4.4912***
	(0.2783)	(0.2056)	(0.2185)
Reference: Correct age			
1 year over-age	-1.3950***	-2.2685***	-1.5240***
	(0.2868)	(0.2371)	(0.2216)
2 years over-age	-1.4878***	-3.1147***	-1.3967***
	(0.4215)	(0.3162)	(0.3100)
3+ years over-age	-0.8630	-4.4028***	-2.1193***
	(0.6563)	(0.3523)	(0.3897)
Constant	-5.1182***	-7.3342***	-6.1700***
	(1.6820)	(1.9352)	(2.1482)
Observations	14,415	18,494	30,689
R-squared	0.3329	0.2355	0.2554

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

Table 10 uses Grade 10 performance to predict Grade 12 Mathematics performance, as measured by the Mathematics Equivalent score. As final Grade 12 results were not available in the DDD data, Term 3 Grade 12 results are used. Grade 10 Maths performance is more strongly associated with Grade 12 performance than Grade 9 performance is associated with Grade 11 performance. Both EFAL and, to a lesser extent, LO, in Grade 10 are associated with better Grade 12 Maths performance. Absenteeism

in Grade 12 has relatively large negative effects on Mathematics performance, but Grade 11 absenteeism has a discernible and significant effect on Grade 12 performance only in the Eastern Cape. The effects of school quintile are not strong or consistent after applying the other controls. Girls perform 4-5 percentage points worse than boys, after controlling for Grade 10 performance, while over-age learners perform worse than correctly aged learners even after controlling for previous performance, as the previous regressions also showed.

Table 11: Regressions of 2018 Grade 12[4] Mathematics equivalent performance based on 2016 Grade 10 performance in three provincial samples

	Outcome: Gr12 Maths mark		
	GP	EC	LIM
Gr10 Maths Equivalent	0.5629***	0.5533***	0.5322***
	(0.0258)	(0.0307)	(0.0200)
Gr10 EFAL	0.2023***	0.2643***	0.2322***
	(0.0223)	(0.0271)	(0.0187)
Gr10 LO	0.1437***	0.1793***	0.1192***
	(0.0204)	(0.0299)	(0.0215)
Absenteeism (2018):			
Days absent (Term 1)	-0.3432***	-0.1854**	-0.2290**
	(0.0693)	(0.0871)	(0.0889)
Days absent (Term 2)	-0.1916*	-0.1331*	-0.2488***
	(0.0997)	(0.0768)	(0.0844)
Days absent (Term 3)	-0.1851**	-0.4563***	-0.4037***
	(0.0874)	(0.0633)	(0.0704)
Days absent (2017)	0.0415	-0.0760*	0.0344
	(0.0299)	(0.0431)	(0.0281)
Reference: Q1-3			
Quintile 4	0.6102	1.4529	-2.0419**
	(0.6577)	(1.1071)	(0.9558)
Quintile 5	-2.0415	1.7451*	9.8402***
	(1.3485)	(0.9915)	(1.6717)
Female	-4.2716***	-4.9271***	-5.4927***
	(0.2130)	(0.2517)	(0.1705)
Reference: Correct age			
1 year over-age	-2.9553***	-3.9683***	-3.3926***
	(0.2229)	(0.3026)	(0.2168)
2 years over-age	-4.1099***	-5.8955***	-4.2597***
	(0.3043)	(0.3945)	(0.2725)
3+ years over-age	-5.2731***	-7.4599***	-5.3274***
	(0.3715)	(0.4834)	(0.3209)
Constant	1.4660	-12.6777***	-0.2849
	(1.5353)	(1.8469)	(1.5798)
Observations	20,539	21,833	41,063
R-squared	0.4446	0.4313	0.4361

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

[4] *Note: Term 3 results, not the National Senior Certificate examinations, as these were unavailable in the data set.*

4.2 Patterns of subject performance

The earlier analysis has shown that subject performance in higher grades as well as progression to higher grades can to some extent be 'predicted' by earlier performance in school-based assessments. The next few tables show how subject marks in some major subjects are correlated for two provincial samples in the Grade 9 school-based assessment in Term 4 as well as matric marks in Term 3. The correlations are for the full sample of learners who were in Grade 9 in 2015 and Grade 12 in 2018 (i.e. not the same sample as in the regressions below).

Table 12 and shows that the Science and Maths marks in both the Gauteng and Limpopo samples are quite highly correlated. In Gauteng the correlations between Maths and EFAL are lower than that between Science and EFAL. The correlations for the third term of matric (Table 13) are extremely high between Maths and Physical Science in both provincial samples, but the correlations between Maths Literacy and Physical Science is much lower.

Table 12: Correlations between Maths, EFAL and Natural Science marks in Grade 9 in the Gauteng and the Limpopo samples, 2015

	Maths	EFAL	Science
		Gauteng	
Maths	1		
English FAL	0.488	1	
Natural Sciences	0.568	0.622	1
		Limpopo	
Maths	1		
English FAL	0.570	1	
Sciences	0.647	0.575	1

Source: Own calculations using SA-SAMS data

Table 13: Correlations between Maths, Maths Literacy, EFAL and Physical Science marks in Grade 12 in the Gauteng and the Limpopo samples, 2018

	Maths	Maths literacy	EFAL	Physical Science
			Gauteng	
Maths	1			
Maths literacy	n/a	1		
English FAL	0.565	0.555	1	
Physical Science	0.861	0.537	0.612	1
			Limpopo	
Maths	1			
Maths literacy	n/a	1		
English FAL	0.518	0.547	1	
Physical Science	0.810	0.500	0.542	1

Source: Own calculations using SA-SAMS data

5. SUBJECT CHOICE BETWEEN MATHEMATICS AND MATHEMATICAL LITERACY

5.1 Initial Mathematics subject choice

The choice made by Grade 9 learners between taking Mathematics or Mathematical Literacy from Grade 10 (one of these two subjects has to be offered) has lasting consequence for a learner's tertiary education and career possibilities, as relatively good performance in Mathematics is a prerequisite for programmes in medicine, engineering, the sciences and even commerce. Thus learners who have the ability to pass Mathematics but opt instead for Maths Literacy may unnecessarily limit their choice of careers at an early stage. Conversely, learners who are ill-equipped to pass Mathematics but choose to take it anyway face a higher risk of failing the subject and possibly also failing Grade 11 or matric.

Ideally, Grade 9 school-based assessment should provide a signal of learners' prospects of passing Mathematics in Grades 10-12, and influence their decision to take Maths or Maths Literacy. However, previous research has found school-based continuous assessment to be a poor predictor of performance in the matric exam, particularly in weaker schools (Van der Berg and Shepherd, 2015).

Quit a large number of learners take Mathematics in Grades 10-12 despite doing poorly in Maths in Grade 9, as shown in Table 14 below for Gauteng. This table only refers to that part of the provincial sample captured in 2015 in Grade 9 and in 2018 in Grade 12, i.e. it excludes dropouts and repeaters.

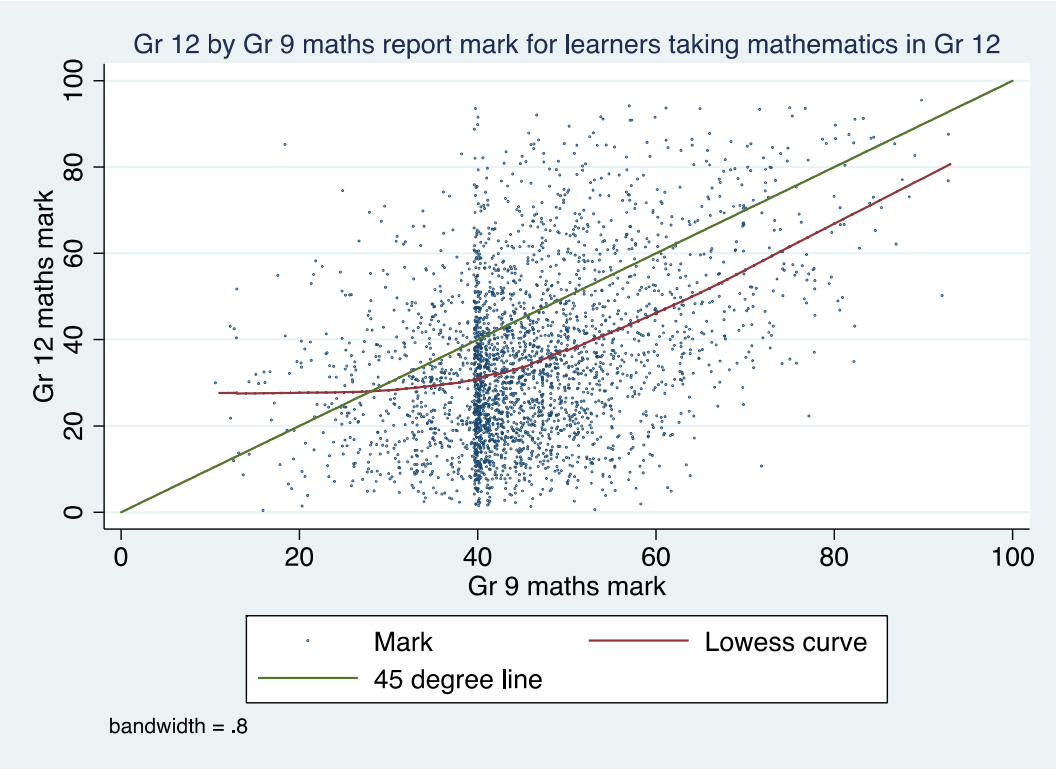
Table 14: Number of learners achieving a pass mark in Maths in Gr12 in 2018 by grade achieved in Maths in Gr9 in the Gauteng sample

	Gr12 Maths mark, 2018 (Term 3)		
Grade 9 2015 Maths mark	**<30**	**>=30**	**Total**
10-19	14	15	29
	48%	52%	100%
20-29	71	61	132
	54%	46%	100%
30-39	207	180	387
	53%	47%	100%
40-49	720	778	1 498
	48%	52%	100%
50-59	175	357	532
	33%	67%	100%
60-69	31	174	205
	15%	85%	100%
70-79	4	91	95
	4%	96%	100%
80-89	0	30	30
	0%	100%	100%
90-99	0	4	4
	0%	100%	100%
Total	1 222	1 690	2 912
	42%	58%	100%

Source: Own calculations using SA-SAMS data

Those who do poorly in Grade 9 tend to do poorly in Grade 12. The figure below for Gauteng shows that Grade 12 Maths marks (for learners who took Maths) tend to be below those achieved in Grade 9 Maths (most points lie below the diagonal). Furthermore, most learners who took Maths despite getting below 40% at the Grade 9 level, as well as many of those who got above 40% but did not achieve a high mark, also tend to achieve below 40% in Grade 12. However, National Senior Certificate marks were not available, so these Grade 12 marks are school-based assessment results from term 3, and do not necessarily reflect the learner's final performance in the external matric exams. The figure also reveals substantial clustering of marks around 40% in Grade 9 Maths, suggesting that these learners achieved a "true" mark lower than this and were pushed up to achieve a pass mark in Mathematics. (Regressions models, not shown, that included a dummy for a Grade 9 mark of exactly 40% indicate this was roughly equivalent to a mark of around 31%.)

Figure 9: Gr12 Maths mark by Gr9 Maths mark for learners taking Maths in Gr12, Gauteng sample



Source: Own calculations using SA-SAMS data

The regressions below explore the predictors of taking Mathematics and Physical Science in Grade 10 and continuing to take these subjects in Grade 12 for Gauteng and Limpopo. Regressions include controls for district and home language, though these are not reported.

The sample is limited to learners who were in Grade 9 in 2015 and successfully progressed to Grade 12 in 2018 without repetition. It is further limited to learners with data available for each year and who did not change schools, in schools with more than 15 learners taking Maths in their grade. These limitations were necessary to explore the link between Grade 9 performance and Grade 10 and Grade 12 subject choice, given that we only had four years of data available, but the result is a relatively small subsample. Given the high rates of repetition, those who progressed from Grade 9 to Grade 12 without repetition form a very select group that is by no means representative of the total sample of learners in the school system. It also excludes learners whose schools did not submit data in any of the years, which again may not be a random subsample of schools.

As expected, Grade 9 Maths, English First Additional Language (EFAL) and Natural Sciences marks (all from Term 4) predict choosing Maths in Grade 10 as well as continuing with it to matric rather than changing to Maths literacy. In Gauteng, a 10 percentage point increase in Grade 9 Maths mark is associated with an 8 percentage point increase in the probability of taking Maths in Grade 12, holding other variables constant. However, it is also interesting to note that the much of the variation in Mathematics and Science subject choice is left unexplained even by Grade 9 marks.

Interestingly, Grade 9 Maths marks are not generally more strongly predictive of the choice to take Mathematics rather than Mathematics Literacy than are EFAL or Natural Sciences marks. Grade 9 marks are more strongly predictive of taking Maths and indeed also Science in Grade 12 than taking these subjects in Grade 10, which suggests that weaker students may take Maths or Science initially but switch subjects by the time they reach Grade 12.

For a given mark, in Gauteng girls are nearly 8 percentage points less likely to take Maths in Grade 12, though the association is insignificant in Grade 10. The association is even stronger for Science. However, this should be interpreted in light of the fact that boys are more likely than girls to drop out of school, so there may be some selection effect here – academically weaker girls may remain in school, while weaker boys are more likely to drop out. Over-age learners are less likely to take Maths or Science than their age-appropriate counterparts, even when holding constant their Grade 9 marks.

Table 15: Predictors of Maths and Science subject choice, Gauteng

	(1)	(2)	(3)	(4)	(5)
	Maths Gr 12	Maths Gr 10	Science Gr 12	Science Gr 10	Maths & Science Gr 12
Female	-0.077***	-0.002	-0.102***	-0.044***	-0.101***
	(0.014)	(0.009)	(0.015)	(0.010)	(0.015)
1 year over-age	-0.068***	-0.100***	-0.048***	-0.070***	-0.043***
	(0.014)	(0.011)	(0.015)	(0.010)	(0.014)
2 years over-age	-0.105***	-0.119***	-0.085***	-0.087***	-0.064***
	(0.022)	(0.014)	(0.024)	(0.011)	(0.023)
3 or more years over-age	-0.075**	-0.116***	-0.085**	-0.073***	-0.062*
	(0.037)	(0.019)	(0.035)	(0.016)	(0.037)
Quintile 4	0.053*	0.025	0.056**	0.026	0.067**
	(0.027)	(0.024)	(0.025)	(0.021)	(0.029)
Quintile 5	-0.074*	-0.009	-0.083**	-0.032	-0.076**
	(0.044)	(0.033)	(0.035)	(0.031)	(0.037)
% taking Maths/Science	0.008***	0.008***	0.007***	0.007***	0.006***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Gr 9 Maths mark	0.008***	0.003***	0.006***	0.003***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gr 9 English FAL mark	0.007***	0.006***	0.007***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gr 9 Natural Sciences mark	0.013***	0.008***	0.012***	0.009***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.990***	-0.536***	-0.927***	-0.545***	-0.875***
	(0.059)	(0.047)	(0.056)	(0.039)	(0.058)
Observations	5,290	16,975	5,084	16,957	5,290
R-squared	0.348	0.297	0.285	0.214	0.276

Source: Own calculations using SA-SAMS data

Holding constant other variables, including Grade 9 marks, learners in schools where a higher proportion of learners in their grade take Maths or Science respectively are more likely also to take Maths or Science. This may suggest that some schools are more likely to encourage learners to do Maths regardless of what mark they achieved. Alternatively, this may reflect peer effects.

Compared to quintiles 1-3, learners in quintile 5 schools are less likely to take Maths, holding all other variables (including Grade 9 marks) constant, though the coefficient is only marginally significant, and not significant for Science. This result is based on a small quintile 5 sample size, but may reflect quintile 5 schools attempting to maintain high pass rates by discouraging learners who may be likely to fail from taking Maths.

Table 16: Predictors of Maths and Science subject choice, Limpopo

	(1)	(2)	(3)	(4)	(5)
	Maths Gr 12	Maths Gr 10	Science Gr 12	Science Gr 10	Maths & Science Gr 12
Female	-0.053***	-0.006	-0.075***	-0.036***	-0.084***
	(0.016)	(0.009)	(0.021)	(0.009)	(0.020)
1 year over-age	-0.117***	-0.111***	-0.118***	-0.125***	-0.093***
	(0.022)	(0.014)	(0.026)	(0.015)	(0.024)
2 years over-age	-0.199***	-0.182***	-0.201***	-0.197***	-0.194***
	(0.033)	(0.017)	(0.043)	(0.018)	(0.036)
3 or more years over-age	-0.235***	-0.232***	-0.167**	-0.242***	-0.209***
	(0.049)	(0.020)	(0.084)	(0.019)	(0.049)
Quintile 4	-0.140*	-0.096*	-0.054	-0.130**	-0.057
	(0.075)	(0.050)	(0.052)	(0.056)	(0.062)
Quintile 5	-0.829***	0.013		-0.165	-0.758***
	(0.068)	(0.091)		(0.113)	(0.077)
% taking Maths/Science	0.008***	0.008***	0.007***	0.008***	0.005***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Gr 9 Maths mark	0.006***	0.003***	0.007***	0.003***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gr 9 English FAL mark	0.003**	0.004***	0.002	0.004***	0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gr 9 Natural Sciences mark	0.004***	0.002***	0.004***	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.336***	-0.151**	-0.183	-0.209***	-0.387***
	(0.120)	(0.068)	(0.133)	(0.076)	(0.129)
Observations	3,501	14,398	2,979	13,969	3,501
R-squared	0.362	0.371	0.221	0.274	0.234

Source: Own calculations using SA-SAMS data

5.2 Switching from Maths to Maths Literacy

The regressions in Table 17 include only learners who were in Grade 10 and taking Mathematics and not Mathematical Literacy, in 2016 and progressed to Grade 12 in 2018. The outcome variable is the probability of having changed to Mathematical Literacy by Grade 12, with no repetition or dropout.

As would be expected, in all three provincial samples, higher Mathematics performance is associated with a lower probability of changing to Mathematical Literacy before reaching Grade 12. Absenteeism

does not appear to play a significant separate role. School quintile is only significant in Limpopo, where quintile 5 learners are more likely to switch to Mathematical Literacy. This supports anecdotal evidence that quintile 5 schools often encourage weaker learners to change to Mathematical Literacy in order to improve a school's matric results. With the exception of the Eastern Cape, female learners are more likely to switch to Mathematical Literacy, as are over-age learners.

Table 17: Probability of having switched to Maths Literacy in Gr12 after taking Mathematics in Gr10 in three provincial samples, 2016-2018

	P(Maths Lit Gr12| Maths Gr10)		
	GP	EC	LIM
Gr10 Maths mark	-0.0107***	-0.0019***	-0.0061***
	(0.0009)	(0.0004)	(0.0006)
Gr10 EFAL	-0.0028**	-0.0000	-0.0025***
	(0.0011)	(0.0004)	(0.0006)
Gr10 LO	-0.0019*	-0.0004	-0.0013
	(0.0010)	(0.0005)	(0.0009)
Absenteeism (2018):			
Days absent (Term 1)	0.0052	0.0033	-0.0022
	(0.0050)	(0.0028)	(0.0027)
Days absent (Term 2)	0.0001	0.0039	0.0041
	(0.0042)	(0.0028)	(0.0048)
Days absent (Term 3)	-0.0002	0.0011	0.0021
	(0.0042)	(0.0019)	(0.0027)
Reference: Q1-3			
Quintile 4	-0.0103	0.0382	-0.0126
	(0.0329)	(0.0559)	(0.0375)
Quintile 5	0.0161	-0.0101	0.2482***
	(0.0687)	(0.0152)	(0.0391)
Female	0.0544***	-0.0072	0.0454***
	(0.0113)	(0.0046)	(0.0067)
Reference: Correct age			
1 year over-age	0.0732***	0.0125**	0.0423***
	(0.0144)	(0.0052)	(0.0077)
2 years over-age	0.1373***	0.0270***	0.0917***
	(0.0201)	(0.0077)	(0.0147)
3+ years over-age	0.2099***	0.0328***	0.1620***
	(0.0297)	(0.0106)	(0.0194)
Constant	0.7867***	0.2483***	0.5361***
	(0.1002)	(0.0548)	(0.0673)
Observations	9,704	13,276	26,674
R-squared	0.2426	0.0858	0.1378

Robust standard errors in parentheses.
Race, home language, and school district are controlled for.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

6. DATA ON LEARNERS WITH DISABILITIES

The DDD data presents a new data source based on individual level data at the school level. It thus offers a potentially important insight into how well disability data is reported in the individual data

compared to Annual Survey of Schools (ASS or EMIS) data. In this case, the 2014 ASS data is compared with 2018 DDD/SA-SAMS[5] data for the Eastern Cape[6].

Table 18: Disability enrolment in ordinary public schools in the Annual School Survey (2013, 2014) & DDD/SA-SAMS (2018), Eastern Cape

	ASS 2013	ASS 2014	SA-SAMS 2018
Total no. of learners with disabilities	27 823	18 625	36 670
Total learner enrolment	1 883 883	1 892 941	1 987 763
Total no. of schools	5 501	5 501	5 222
Mean no. of learners with disabilities per 1,000 learners (all schools)	13 (0.74)	9 (0.67)	14 (0.60)
Number of schools reporting any disability enrolment	1 341	743	2 263
Mean no. of learners with disabilities per 1,000 learners (in schools which report learners with disabilities)	53 (2.77)	63.6 (4.46)	33 (1.27)
No. (%) of schools not reporting enrolment of any learners with disabilities	4 160 (75%)	4 758 (87%)	2 959 (57%)

Standard errors are shown in brackets unless otherwise stated.
Disability enrolment is measured as a school reporting at least 1 learner with a disability in the year in question.
Source: Own calculations using SA-SAMS and EMIS data

The data in Table 18 shows a substantial increase in reported enrolment of learners with disabilities in the Eastern Cape in 2018 (increases of between 32% and 97%, depending on whether the comparison is made with 2013 or 2014 data). Total learner enrolment increased by only 5-6% in the period. The change in numbers of schools reporting is also substantial; between 69% and 205% more schools reported enrolling at least one learner with a disability in 2018 than in 2013 and 2014, respectively. Thus it is clear that learner disability status was much more widely reported in 2018 (under SA-SAMS) than in 2013-2014 (under ASS). However, the observed increases in reporting cannot be attributed to the introduction of SA-SAMS alone, as the Screening, Identification Assessment and Support (SIAS) policy was implemented concurrently.[7]

Data quality also appears to have improved from 2013/14 to 2018. Inconsistency of reporting of disability in the ASS between 2013 and 2014is apparent in Table 18. These inconsistencies were pervasive in most provinces in the ASS 2011-2014. SA-SAMS has the potential to produce more consistent data over time as a learner's disability status is entered only once and aggregation of total enrolment is automated. Previously, schools were required to aggregate learners with disabilities by gender, grade and population group. Previous analysis of ASS data suggests that aggregation errors, at school-level, were widespread. Fewer schools reported unrealistically high rates of disability

[5] *DDS/SA-SAMS data as at November 2018*

[6] *Analysis was limited to one province, due to the large data file size of the national dataset. The Eastern Cape was chosen as the quality of data in SA-SAMS in that province is generally good and the province is sufficiently large to allow robust analysis of disability sub-groups*

[7] *This policy introduced new screening and identification procedures for learner disabilities and included training of school personnel on inclusive practices from 2015 onwards.*

enrolment[8] in 2018. Higher quality data allows more reliable conclusions to be drawn about the type of schools that were able to identify and report learners with disabilities. This is demonstrated by a comparison of results from linear probability regression models run on the Eastern Cape data from SA-SAMS in 2018 and from the ASS (2012-2014), as shown below.

Table 19: Estimating the probability of an Eastern Cape school reporting at least one learner with a disability: Comparing SA-SAMS and ASS

	SA-SAMS 2018	ASS 2012	ASS 2013	ASS 2014
Reference: Urban school				
Rural school	-0.066***	-0.037***	-0.028**	0.004
	(0.015)	(0.012)	(0.013)	(0.010)
Reference: Quintile 4 & 5 schools				
Quintile1	-0.420***	-0.470***	-0.486***	-0.465***
	(0.037)	(0.031)	(0.032)	(0.026)
Quintile2	-0.428***	-0.459***	-0.486***	-0.456***
	(0.038)	(0.031)	(0.032)	(0.026)
Quintile3	-0.322***	-0.425***	-0.430***	-0.429***
	(0.036)	(0.030)	(0.031)	(0.025)
Reference: Small school				
Large school (>600 learners)	0.230***	0.098***	0.108***	0.083***
	(0.018)	(0.016)	(0.016)	(0.013)
Reference: Ordinary school				
Full-service school	0.321***	0.380***	0.190*	-
	(0.085)	(0.109)	(0.115)	0.240***
Constant	0.799***	0.653***	0.693***	0.553***
	(0.035)	(0.029)	(0.030)	(0.024)
R-squared	0.111	0.073	0.071	0.081
Sample (number of schools)	5 147	5 428	5 428	5 428

Full service schools identified using 2017 data for SA-SAMS, 2011 data for ASS regressions
Standard errors in parentheses.
**** p<0.01, ** p<0.05, * p<0.1*
Source: Own calculations using SA-SAMS data

Analysis of SA-SAMS data shows a much larger effect of school size (and a somewhat larger urban effect) on the probability that a school reports at least one learner with a disability. The SA-SAMS data provides a much richer portrayal of the school wealth effect: Quintile 1 and 2 schools were 42 to 43% less likely, and Quintile 3 schools were 32% less likely, to report at least one learner with a disability than Quintile 4-5 schools, all else being equal. In the ASS data little difference between Quintile 1, 2 and 3 schools can be detected. Improved data quality is also reflected in a higher R^2 score in regressions on the SA-SAMS data and is probably driven by the large number of reporting schools.

More research is needed to identify the challenges Quintile 1 and 2 schools experience in identifying learners with disabilities. In particular, the impact of access to the medical diagnosis currently required to classify a learner as disabled in Quintile 1 and 2 schools should be investigated.

The learner-level disability data in SA-SAMS allows new research questions to be addressed. For example, grade-age profiles for learners with disabilities can now be produced. The analysis of drop-

[8] *Unrealistically high rates of disability were defined as more than 60% of learners, overall, or more than 80% of learners in a particular phase having disabilities.*

out and repetition, presented earlier in this report, can be done for learners with disabilities (disaggregated into learning disabilities and other disabilities), and can be contrasted to patterns among learners without disabilities. Estimation of mean age and grade at which learners are identified as having a disability, in different parts of the school system, should be possible. Learner-level data will allow analysis of the drivers of substantially and significantly higher rates of disability identification among boys observed previously in the ASS data (2011 to 2014) and in the Eastern Cape in 2018 (17/1,000 boys and 11/1,000 girls among learners have a disability). Further, SA-SAMS represents the first effort to integrate data collection across special and ordinary schools and enables comparison of learner characteristics and performance across the special, full-service and ordinary schools.

7. A BRIEF CONCLUSION

The analysis undertaken in this paper shows the value of learner-level data. This is the case whether the analysis is for a single year, as in the disability data discussed in the previous section, or for monitoring progression and performance across years. The growing availability of data through increased collection of the SA-SAMS data for use in the DDD holds the promise that it may in future become possible to have such data for all schools. Yet even the samples that had to be used to track the same schools over time makes it possible to investigate important processes in schools that were impossible previously. Although the results are not yet fully representative, they do provide first evidence on the value of school-based assessment and absenteeism as predictors of progression, later performance and subject choice, and the disability data, though still far from perfect, provides a much fuller picture of the prevalence of disability amongst South African learners.

8. BIBLIOGRAPHY

Shepherd, Debra, and Servaas Van der Berg. 2019. *Analysing matric data to identify promising schools for the 'Maths Challenge' of the Epoch and Optima Trusts.* Draft Report, Stellenbosch: Resep.

Simkins, Charles. 2013. *Performance in the South African educational system: What do we know?* Johannesburg: CDE.

Van Biljon, Chloé, and Cobus Burger. 2019. *The Period Effect: The Effect of Menstruation on Absenteeism of School Girls in Limpopo.* Working Paper, Resep, Stellenbosch: Stellenbosch Economic Working Paper.

Van der Berg, Servaas, and Debra Shepherd. 2015. "Continuous assessment and matriculation examination marks – an empirical examination." *South African Journal of Childhood Education* 78-94.