Estimation of promotion, repetition and dropout rates for learners in South African schools

Daniël Wilhelm Uys

Edward John Thomas Alant

A new procedure for estimating promotion, repetition and dropout rates for learners in South African schools is proposed. The procedure uses three different data sources: data from the South African General Household survey, data from the Education Management Information Systems, and data from yearly reports published by the Department of Basic Education. The data from the General Household survey are utilised to estimate repetition rates for learners in three different age groups. Keeping these repetition rates fixed, the data from the other two sources are used to estimate dropout and promotion rates, which are based on a birth-year-cohort approach for the different age groups. In particular, this procedure involves minimising the difference between actual flow-through rates and simulated flow-through rates for both the birth-year cohorts and age groups. The procedure gives different results when compared to published literature.

Keywords: South African schools, promotion rates, repetition rates, dropout rates, birth-year cohorts

Introduction

The quality of basic education is of great importance to the general and economic wellbeing of any nation. If the school system fails to produce well-educated learners, the effects on the socio-economic standard of the people and their workplace are numerous. Nearly 21 years since the first democratic election, the South African school system unfortunately still suffers from some difficulties, partly due to the country's long legacy of apartheid. Despite government's substantial efforts over the years to improve access and quality in education, learner performance is still low

---

Daniël Wilhelm Uys
Department of Statistics and Actuarial Science, University of Stellenbosch
E-mail: dwu@sun.ac.za
Tel: +27 (21) 808 3879.

Edward John Thomas Alant
POPII Research and Consulting
E-mail: edward@popii.info
Tel: +27 (21) 886 5967

(Chisholm, 2005). Shindler and Fleisch (2007) stated that access to education is lower than what most published sources suggest. According to Murtin (2013), the quality of basic education for a large proportion of Black learners is substandard. In addition, Sedibe (2011) also found that there is no equality in access to resources, such as libraries, laboratories and computers, between previously disadvantaged schools.

The South African schooling system starts at Grade 1 and consists of eleven further consecutive grades up to Grade 12. Some schools offer Grade 0 prior to Grade 1; for the purpose of this paper, this grade is not taken into account. The system is further divided into four schooling phases: Grades 1 to 3, 4 to 6, 7 to 9, and 10 to 12. Not all learners who enrolled in Grade 1 will complete Grade 12. Of those learners who complete Grade 12, not all of them will certainly do so in twelve years. The Department of Basic Education (DBE) reports on the number of learners in each grade.[1]. However, the number of learners reported includes possible repeaters, which are not specified, and excludes those learners who dropped out of the schooling system. Therefore, using DBE data, direct calculation of accurate promotion (or pass), repetition and dropout rates is a near impossible task. For this reason, researchers attempt to estimate these rates.

The estimation of promotion, repetition and dropout rates is of national concern, and comprehensive studies have been conducted to estimate these rates. In addition, researchers endeavoured to identify possible reasons why these rates are sometimes too high or too low. Fleisch and Shindler (2007) used the cohort Birth to Twenty dataset to address patterns and prevalence of initial school enrolment, late entry, promotion, and repetition in urban schools in South Africa. They also reported on the national concern about the perceived high dropout rates from secondary schools.

In an in-depth study, Burger, Van der Berg & Von Fintel (2013) proposed nearest neighbour estimates for promotion, repetition and dropout rates. They utilise these rates further by studying the consequences of two policies implemented by the DBE in the late 1990s. The first was that schools were no longer allowed to enrol learners who were more than two years older than the correct grade-ages, and the second that learners were not allowed to repeat more than once in each of the four schooling phases. In particular, they consider the effect of these two policies on the unemployment in the South African labour market.

Anderson et al. (2001) gave evidence that, although declining, there persists a racial gap in schooling in South Africa, mostly attributed to the high rate of grade repetition of Black learners, with only small differences in the enrolment rate across other racial groups.

Crouch (2005) stated that repetition rates are poorly reported and thus poorly estimated. Crouch (2005) further studies the problem of dropouts in-depth, claiming that 60% of the learners enrolled in Grade 1 do not reach Grade 12 or its equivalent in Further Education and Training colleges. Crouch (2005) also explained different

ways in which dropouts can be calculated, and concludes that the correct approach is to consider age-specific and grade-specific enrolment ratios from the South African General Household survey (GHS).[2] By using the proportion of learners enrolled, and applying this to the total population, the number of learners not in school can be determined. Reasons for dropping out were discussed, and further comparisons of dropout rates were made with those of other countries.

Modisaotsile (2012) reported on very high dropout rates and low literacy and numeracy levels, claiming that 50% of the learners enrolled in Grade 1 complete Grade 12. Modisaotsile (2012) also stated that sexual abuse, pregnancy and poverty are factors that increase the dropout rate in secondary schooling. In a study conducted in the KwaZulu-Natal province of South Africa, Grant and Hallman (2008) affirmed this by finding a strong correlation between school performance and pregnancy-related school dropout of female learners.

Obtaining credible estimates for promotion, repetition and dropout rates for learners in South African schools is worthwhile for various reasons. It helps identify problems regarding teachers and teaching methods that may exist in schools. These estimates can also be used by the DBE and tertiary institutions in planning their infrastructure. They can further be applied in actuarial calculations, such as loss-of-income claims. For example, the estimates are helpful to project the school career path of learners who sustained serious injuries arising from motor vehicle accidents or medical malpractice.

Data sources

Demographic data in South Africa, in common with most developing countries (and even a few developed countries), are not completely accurate. In this regard, Fleisch and Shindler (2007) stated that it is particularly significant and well documented that school and census data sets can be unreliable. In this paper, we start by estimating repetition rates for each grade. Keeping these rates fixed, we use enrolment numbers of learners to construct a flow-through model in order to ultimately obtain estimates for promotion and dropout rates for each grade. Three different data sources are utilised for the purpose of estimation.

The first data source is Statistics South Africa, which conducts censuses and various surveys in order to attain information about the South African population.[3] One such survey is the GHS, which collects data on a yearly basis, and aims to provide government and companies with information about South African households. Data of the GHS are accessible online and can easily be manipulated into a required and suitable format for further use. The survey's results are generalised to the South African population by using a method of weighting. Full details of the weighting procedure are available on the website. Data from the GHS are available from 2002 to 2011. Unfortunately, only the surveys done in 2009, 2010 and 2011 include questions on learner's birth years, their current grade and whether they repeated the grade or not. The data obtained from these questions in these three years are utilised for estimating repetition rates for Grades 1 to 12.

The second data source is the Education Management Information Systems (EMIS). The data are captured by the DBE on a yearly basis, usually during the months of March and April, and give the number of learners at every school in South Africa. The data are available, on request, from the DBE. Geographic, provincial and socio-economic information of the different schools are also given, as well as information regarding the age, current grade and gender of learners. The data also contain the number of repeaters; unfortunately, the age of the repeaters are not recorded. Data of the EMIS are available from 1997 to 2011.

Careful consideration of the EMIS data shows both discrepancies and vast differences in the quality of data between the nine provinces. Measurement errors, due to incorrect data capturing, and missing data in some of the provinces are obvious. Visual inspection of graphs constructed from the data for the various provinces shows that certain data years have significantly lower or higher numbers of learners than other years. This suggests that the data needs to be cleaned before further use. Consider, in this regard, Figure 1 where the number of learners for each of the data years is plotted against their age for the North-West province. An unrealistic high number of three-year olds is observed in 1997. It makes sense to remove such a data point. In addition, the substantially lower number of learners in 2001, from ages six to eighteen, is assumed to be incorrect and requires attention. Figure 2, also constructed from data obtained for the North West province, shows the number of learners for each data year plotted against the current grade of the learners. Clearly, again the number of learners in 2001 for Grades 1 to 12 is by far less than that for other years.
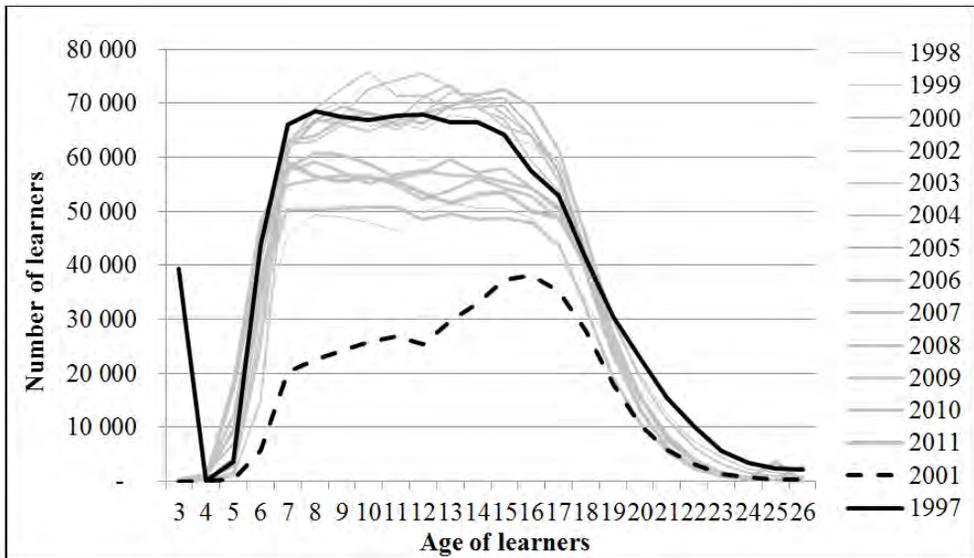

Figure 1: Number of learners per age per data year before adjustments (North-West province)
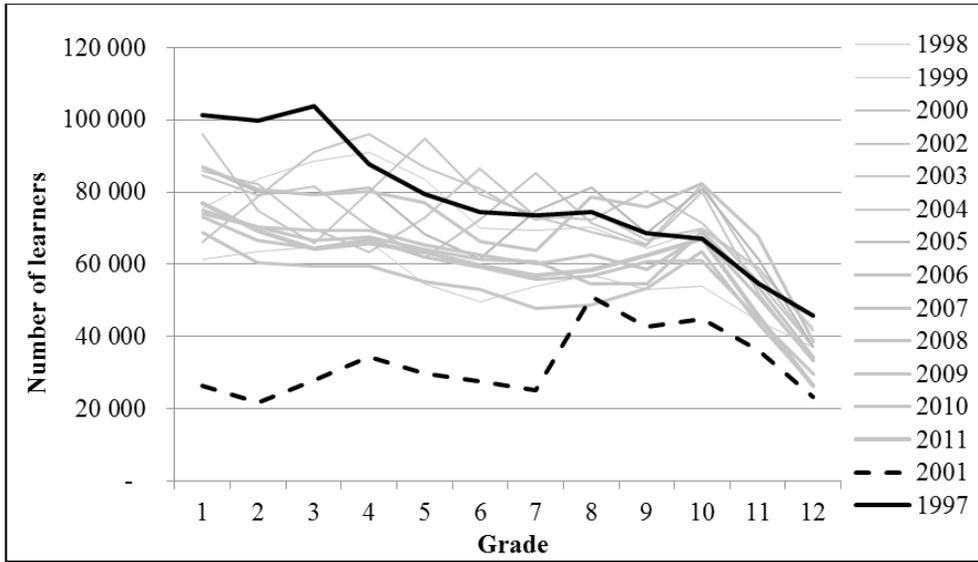
Figure 2: Number of learners per grade per data year before adjustments (North-West province)

Altering data should be done cautiously and consistently. For example, reconsider the graphs in Figures 1 and 2. Merely adjusting the number of learners at different ages in 2001 will affect the number of learners in different grades in 2001. For this reason, we decided to replace the number of learners per age in 2001 with a weighted average based on the number of learners in all the other years, in such a way that the number of learners per grade is adjusted correspondingly. The weights used in the calculation decline exponentially as we move further away from 2001. Therefore, the number of learners in 2000 and 2002 are awarded larger weights than the number of learners in data years that are further away from 2001. Figures 3 and 4 show the graphs for the data obtained for the North-West province once the adjustments have been made.
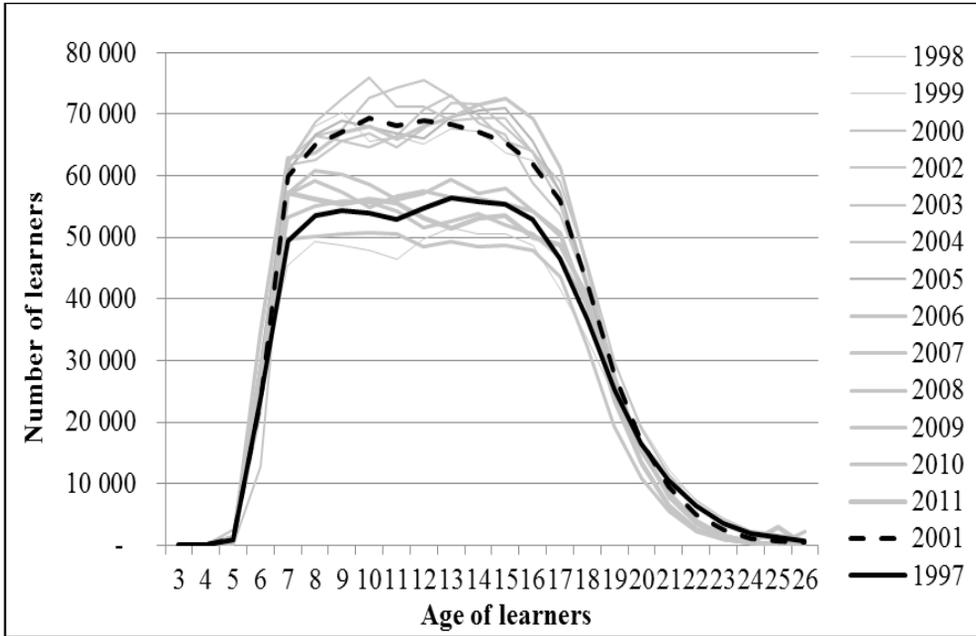
Figure 3: Number of learners per age per data year after adjustments (North-West province)

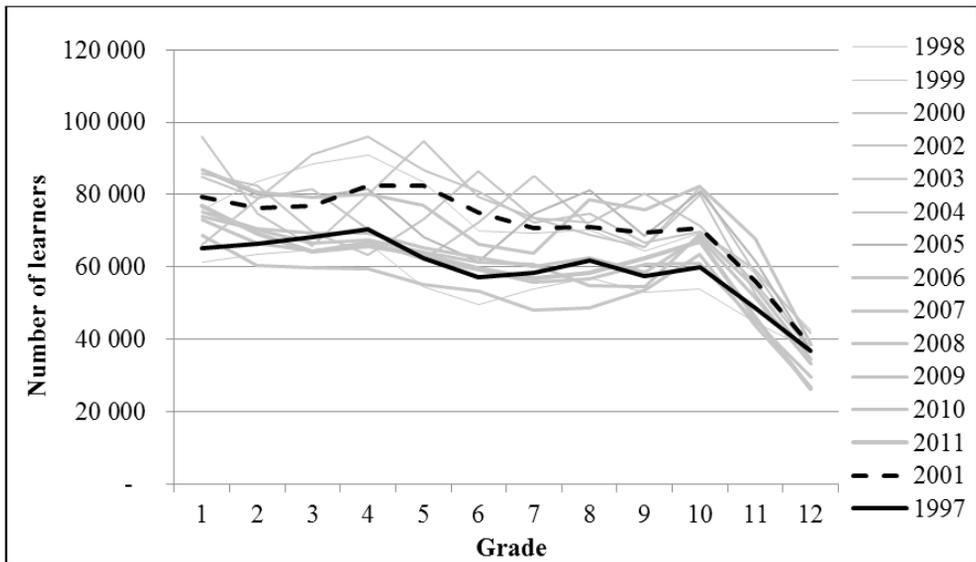*Figure 4: Number of learners per grade per data year after adjustments (North-West province)*

Two graphs for each of the nine provinces were constructed and visually inspected. The first is the number of learners plotted against age, and the second is the number of learners plotted against grade. For each province, these graphs were considered at the same time and, where necessary, adjustments were made. Where data

were missing or erroneous, data years were replaced with a weighted average based on the remaining data years. A consistent method of data adjustment was thus followed throughout. No adjustments were made to data from the Free State and Western Cape provinces, while other provinces such as Gauteng and Kwazulu-Natal required substantial changes. In total, sixteen adjustments were made. Due to space limitations, the graphs for data obtained for each province prior to and after adjustment are omitted. The adjusted EMIS data for each of the provinces are combined in a single data set, representing all South African schools. This is used in estimating dropout and promotion rates for Grades 1 to 11.

The third data source utilised in estimating the rates are yearly reports published by the DBE. The Education Statistics in South Africa reports from 1999 to 2005 and the Education Statistics in South Africa from 2006 to 2011 are available on the DBE's website.[4] The data in these reports are used specifically for estimating Grade 12 dropout and promotion rates.

The EMIS data are often used to compare the number of learners enrolled in a specific grade with the number of learners enrolled in the previous year in the previous grade. In this way, dropout rates can be determined. It should be noted that this method of calculation does not accurately reflect the true dropout rates, as it does not account for learner repetition and late entry into school. The method in this paper incorporates the estimation of repetition rates from data obtained from the GHS. It should be noted that, since the GHS surveys do not include exactly the same learners in their samples for consecutive years, the estimation of the repetition rates are not based on a birth-year-cohort approach. Nevertheless, these repetition rates are assumed to be reliable and are kept fixed when dropout and promotion rates for Grades 1 to11 are estimated from the cleaned EMIS data. The estimation of the dropout and promotion rates is based on a birth-year-cohort approach, where the learner's age at the first year of enrolment in primary school is considered. Finally, data from the reports of the DBE are used to estimate Grade 12 dropout and promotion rates. The method of estimation is discussed in detail in the next section.

Method

According to legislation determined by the South African Schools Act (SASA), it is compulsory for children to attend school in the year in which they turn seven.[5] However, many children enrol into Grade 1 in the year in which they turn six and some as late as the year in which they turn eight. Although there are learners in Grade 1 who are even younger and also older, we only consider, for the purpose of this study, learners who are in Grade 1 in the year in which they turn, six, seven or eight. Learners who are two or more years younger or older for a particular grade are not considered.

Since we consider turning seven in Grade 1 to be the normal age for learners to be in Grade 1, we denote this number of learners by n_{norm}. For Grade 1 learners, the abbreviated notation $_1 n_{norm-1}$ and $_1 n_{norm+1}$ are used for the number of learners who turn six and eight, respectively. Similarly $_2 n_{norm-1}$, n_{norm} and $_2 n_{norm+1}$ are the number of learners in Grade 2 who turn seven, eight and nine, respectively. In general, the notation $_g n_{norm-}$, $_g n_{norm}$ and $_g n_{norm+}$ refers to the number of learners in the three different age groups that is considered respectively in each of the grades, $g = 1, \ldots, 12$.

Data from the GHS survey for 2009, 2010 and 2011 provide the number of learners who repeat. Ultimately, repetition rates are calculated for each of the three age groups, mentioned in the previous paragraph, in each of the twelve grades. The repetition rate for grade for the age group $norm-$ for 2010 is calculated as follows:

$$\frac{\text{Number of learners who are one year older, repeating Grade } g \text{ in 2010}}{_g n_{norm-1} \text{ in 2009}}. \quad (1)$$

Similarly, for the same age group, the repetition rate for grade for 2011 is:

$$\frac{\text{Number of learners who are one year older, repeating Grade } g \text{ in 2011}}{_g n_{norm-1} \text{ in 2010}}. \quad (2)$$

One should note the way in which repetition rates in (1) and (2) are determined: the numerator and the denominator are based on different surveys. Although they are weighted up to reflect the total population, there may be incompatibilities in both the sampling of the different surveys and the populations used to weight up the surveys. We purposely determined the repetition rates using this cohort-based approach, therefore maintaining consistency with calculations in the remainder of the paper.

Calculating the mean of (1) and (2) for each of the twelve grades, provides us with average repetition rates for the age group $norm-$. Denote these repetition rates by

$$_g \overline{r}_{norm-1} \text{ for } g = 1, \ldots, 12.$$

Repetition rates for the age groups $norm$ and $norm+$ are calculated accordingly, and are denoted by $_g \overline{r}_{norm}$ and $_g \overline{r}_{norm+1}$ for $g = 1, \ldots, 12$, respectively. Table 1 shows

the numerical calculated repetition rates for each of the three different age groups and each of the twelve grades.

Table 1: Average repetition rates (in %) for the different age groups in each of the twelve grades

Age	Grade											
	1	2	3	4	5	6	7	8	9	10	11	12
6	6											
7	5	8										
8	7	7	7									
9		8	8	7								
10			11	7	5							
11				10	4	5						
12					11	6	4					
13						8	4	4				
14							7	5	7			
15								11	10	13		
16									16	20	12	
17										24	16	9
18											24	11
19												12

Since estimation of dropout and pass rates from the EMIS data is based on a birth-year-cohort approach, we add an additional subscript in the notation used for the number of learners in each grade and in each of the age groups. The notation

$_g n_{norm-1,c}$, $_g n_{norm\ c}$ and $_g n_{norm+1,c}$ will henceforth refer to the number of learners in the three different age groups in Grade $g = 1,\dots,12$, who were born in birth-year cohort .

The number of learners from birth-year cohort   who turned six in Grade 1, expressed relative to total number of learners from birth-year cohort   who turned six, seven and eight is given by

$$_1 prop_{norm-1,c} = \frac{_1 n_{norm-1,c}}{_1 n_{norm-1,c} + _1 n_{norm,c} + _1 n_{norm+1,c}} \tag{3}$$

where $_1 prop_{norm-1,c}$ refers to a proportion, and the pre- and post-subscripts are used accordingly with previous defined notation. It is evident that, in (3), $norm-$, $norm$ and $norm+$ represent the age groups six, seven and eight, respectively. The proportions

$$_1 prop_{norm,c} = \frac{_1 n_{norm,c}}{_1 n_{norm-1,c} + _1 n_{norm,c} + _1 n_{norm+1,c}} \quad \text{and} \quad _1 prop_{norm+1,c} = \frac{_1 n_{norm+1,c}}{_1 n_{norm-1,c} + _1 n_{norm,c} + _1 n_{norm+1,c}}$$

are determined in a similar manner for age groups *norm* and $norm+$, respectively. It should be noted that these proportions are only determined for learners who are in Grade 1. We calculated the mean over all cohorts , and represent the means as $_1 \overline{prop}_{norm-1}$, $_1 \overline{prop}_{norm}$ and $_1 \overline{prop}_{norm+1}$. We determined the proportions to be as follows:

$$_1 \overline{prop}_{norm-1} = 0.377; \quad _1 \overline{prop}_{norm} = 0.483; \quad _1 \overline{prop}_{norm+1} = 0.140 \tag{4}$$

We return to these three average proportions below.

A flow-through rate between two consecutive grades is defined as the number of learners in the higher grade expressed relative to the number of learners in the lower grade. These rates can be calculated per birth-year cohort, but differ from promotion (pass) rates in the sense that flow-through rates take repeaters into account, whereas repeaters are excluded from promotion rates. Therefore, we have to distinguish clearly between flow-through rates and promotion rates, and emphasise that these rates are not the same. The EMIS data are used to calculate flow-through rates between grades for each of the three age groups per birth-year cohort. Let the flow-through rate between Grade and for learners in the $norm-$ age group, that were born in cohort be given by

$$_g f_{norm-1,c} = \frac{_{g+1} n_{norm-1,c}}{_g n_{norm-1,c}} .$$

The flow-through rates for the age groups *norm* and $norm+$ are determined similarly as

$$_g f_{norm,c} = \frac{_{g+1} n_{norm,c}}{_g n_{norm,c}} \quad \text{and} \quad _g f_{norm+1,c} = \frac{_{g+1} n_{norm+1,c}}{_g n_{norm+1,c}} .$$

It should be noted that $_{11} f_{norm-1,c}$, $_{11} f_{norm,c}$ and $_{11} f_{norm+1,c}$ give flow-through rates from Grades 11 to 12 for the three different age groups from cohort , and that calculation of $_{12} f_{norm-1,c}$, $_{12} f_{norm,c}$ and $_{12} f_{norm+1,c}$ are not applicable in this instance. The flow-through rates are adjusted to allow for trends over time. This is done by

calculating a mean for each of the flow-through rates between Grades   and
, where the weights are proportional to the number of birth-cohort years available. This implies that later birth-cohort years receive a larger weighting relative to the earlier years. The average flow-through rates for all birth-year cohorts are denoted as

$$_g\overline{f}_{norm-1}, \; _g\overline{f}_{norm} \;\; \text{and} \;\; _g\overline{f}_{norm+1} \tag{5}$$

and the numerical values are given in Table 2.

Table 2: Average flow-through rates (in %) between grades for the different age groups

Age groups	Grades										
	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
norm - 1	71	88	90	91	92	90	98	89	91	81	76
norm	86	93	94	92	96	98	99	93	99	82	75
norm + 1	120	111	104	98	99	97	110	95	108	81	64

Consider again the average repetition rates given in Table 1. We now express the dropout rates, denoted by   , as proportions of these repetition rates. In accordance with the notation used earlier, denote these proportions as $_g\alpha_{norm-}$, $_g\alpha_{norm}$ and $_g\alpha_{norm+}$ for each of the age groups for $g = 1,\ldots,11$. It is necessary to place constraints on the   values to ensure convergence in the simulation. We, therefore, assume that each of these   values lies between 0,00001 and 3. The dropout rates for the three age groups for $g = 1,\ldots,11$ are then given as:

$$_gd_{norm-1} = \left(_g\alpha_{norm-1}\right)\left(_g\overline{r}_{norm-1}\right); \quad _gd_{norm} = \left(_g\alpha_{norm}\right)\left(_g\overline{r}_{norm}\right); \quad _gd_{norm+1} = \left(_g\alpha_{norm+1}\right)\left(_g\overline{r}_{norm+1}\right)$$

As starting values, we choose all the   values to be equal to 0.5. Since we now have repetition and dropout rates for each of the age groups for Grades 1 to 11, promotion rates, denoted by   , can easily be obtained as follows:

$$_gp_{norm-1} = 1 - {}_g\overline{r}_{norm-1} - {}_gd_{norm-1}; \quad _gp_{norm} = 1 - {}_g\overline{r}_{norm} - {}_gd_{norm}; \quad _gp_{norm+1} = 1 - {}_g\overline{r}_{norm+1} - {}_gd_{norm+1}$$

We use the average proportions calculated for the three different age groups for Grade 1 in (4), to divide 1000 (or any arbitrary quantity) learners into the three groups. Thus, 377 learners turn six, 483 turn seven, and 140 learners turn eight in Grade 1. Applying the repetition rates from Table 1, the current dropout rates,

which are now 50% of the repetition rates, and the attained promotion rates to subdivided 1000 Grade 1 learners result in the values given in Table 3. The values in Table 3 are used to obtain flow-through rates in a similar way as flow-through rates per cohort were obtained previously. For example, the flow-through rates from Grade 1 to Grade 2 for learners turning six, seven and eight in Grade 1 are

$$ {}_1 f^*_{norm-1} = \frac{343}{377} = 0,91, \quad {}_1 f^*_{norm} = \frac{471}{483} = 0,98 \quad \text{and} \quad {}_1 f^*_{norm+1} = \frac{157}{140} = 1,12 , $$

respectively. Flow-through rates between other grades are obtained similarly. In general, we use the notation

$$ {}_g f^*_{norm-1}, \quad {}_g f^*_{norm} \quad \text{and} \quad {}_g f^*_{norm+1} \tag{6} $$

to denote these flow-through rates between grades for the three different age groups, respectively. It should be noted again that (6) is defined for $g = 1, \ldots, 11$, but is not equal to twelve. We refer to these flow-through rates as the simulated flow-through rates to distinguish them from the actual average flow-through rates in (5) calculated in Table 2.

The main idea is to minimise the absolute difference between the actual flow-through rates in (5) and the simulated flow-through rates in (6). We decided to use the absolute difference, since this form of error minimisation is more resistant to outliers in the data (although we found that squared differences provided similar results). This is done for Grades 1 to 11, for the three different age groups. Therefore, this amounts to finding values for ${}_g \alpha_{norm-}$, ${}_g \alpha_{norm}$ and ${}_g \alpha_{norm+}$, under certain restrictions, such as to minimise the following quantity

$$ \sum_{g=1}^{11} \sum_{i=-1}^{1} \left| {}_g \overline{f}_{norm-i} - {}_g f^*_{norm-i} \right| \tag{7} $$

Table 3: The number of learners in each grade in the different age groups in the first simulation repetition

Age	Grade											
	1	2	3	4	5	6	7	8	9	10	11	12
6	377											
7	483	343										
8	140	471	303									
9		157	444	272								
10			171	410	245							
11				171	382	227						
12					161	368	211					
13						159	342	198				
14							155	328	186			
15								155	317	167		
16									163	289	134	
17										182	217	110
18											152	176
19												118

Simulation in @Risk software is used to obtain the minimum for the quantity in (7).[6] In the actual simulation, this minimum value equals 0.3912. The finally attained values that correspond with the minimum value are given in Table 4.

Table 4: The attained   values between grades for the different age groups in the final simulation repetition

Age group	Grades										
	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
norm - 1	3.00	0.53	0.48	0.30	0.72	1.01	0.00	1.26	0.29	0.46	1.00
norm	1.11	0.62	0.23	0.59	0.66	0.00	0.00	0.52	0.00	0.10	0.82
norm + 1	0.00	0.08	0.00	0.07	0.29	0.38	0.00	1.25	0.72	0.69	1.08

The   values in Table 4 give the estimated dropout rates, expressed as proportions of the repetition rates. Thus, dropout rates and, ultimately, promotion rates for the three different age groups in Grades 1 to 11 can now be calculated. By using the above procedure to obtain simulated flow-through rates that are nearly equal to the actual flow-through rates, the ratio of the number of learners in the different grades in the final simulation repetition is as good as possible an imitation of the ratio of number of learners in the different grades in the actual data.

Finally, data from the Education Statistics in South Africa reports[7] from 2000 to 2011 are used to estimate the dropout and promotion rates for Grade 12. These reports obtain results from the National Senior Certificate and show the number of

Grade 12 learners who dropped out and passed. Since the repetition rates for Grade 12 were also obtained from the GHS, calculation of the dropout and promotion rates for Grade 12 is trivial. The obtained rates are presented and discussed in the following section.

Estimated results

In this section, the method discussed earlier is applied and results are presented by means of tables and graphs. We start with the repetition rates given in Table 1 for the three different age groups for Grades 1 to 12. A graphical representation of the repetition rates is shown in Figure 5.
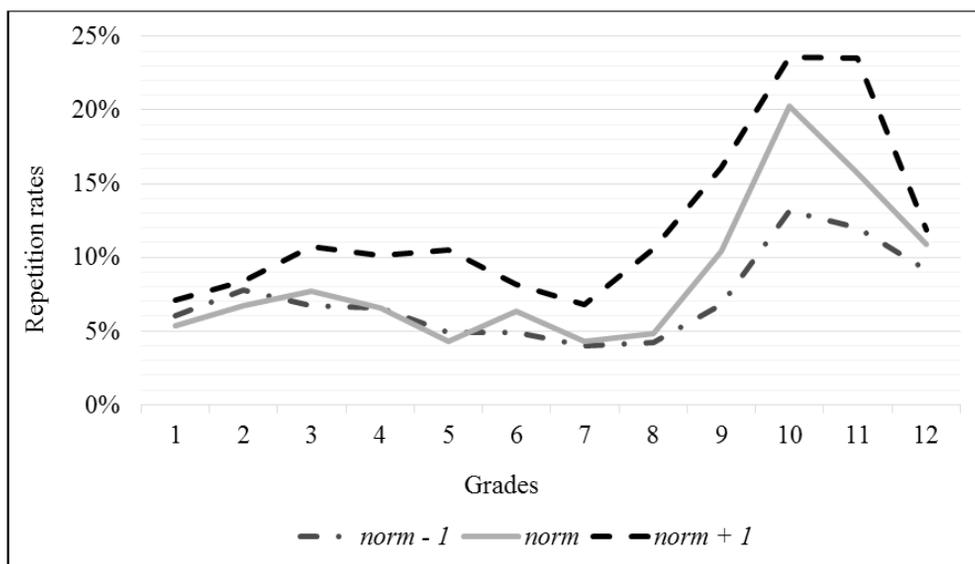

Figure 5: Estimated repetition rates for the different age groups

It is interesting to note that learners, who are one year older than the normal age for a certain grade, have higher repetition rates than learners who are in the other two age groups. A possible explanation for this is that learners in the $norm$ age group, who are more likely to repeat, will ultimately fall in the $norm+$ age groups. For example, a learner turning ten in Grade 4 may repeat Grade 4 once before moving on to Grade 5. Typically, such a learner falls in the $norm+$ age group in Grade 5 in the year in which the learner turns eleven. Learners in the $norm$ age groups, who are less likely to repeat, will remain in the $norm$ age group in later grades. Since we expect those learners who failed once to have a higher likelihood of failing again (Pugatch, 2012), a similar argument can be made for learners who are initially in the $norm-$ age group and who may fall in the $norm$ or $norm+$ age group in higher grades. Furthermore, when interpreting these rates, the DBE regulation with regard to learner progression must be borne in mind. Legislation states that learners are not permitted to repeat any grade more than once, and that learners are not allowed to

repeat more than once in any of the four school phases. We would, therefore, expect higher repetition rates in the first grade of each new school phase. It is, however, not clear how this legislation is enforced by the DBE, or how different schools adhere to it. It should also be noted that there exists interaction between age, repetition and dropout rates. For this reason, repetition rates should not be considered in isolation, but within the context of dropout and flow-through rates.

Table 5 and Figure 6 show the dropout rates for the different age groups. The Grade 1 dropout rate for the $norm-$ age group is relatively high compared to the other two age groups. In addition, the dropout rates for the $norm+$ age groups, for Grades 8 to 11, are higher relative to the same grades for the other two age groups. A reasonable explanation for this is that weaker learners are characterised by higher repetition rates, making these learners older than the normal age for a specific grade. Learners who already repeated in the past are more likely to repeat again. This ultimately results in higher dropouts in the $norm+$ age group, particularly in the higher grades.

Table 5: Estimated dropout rates (in %) for the different age groups

Age group	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
norm - 1	18	4	3	2	4	5	0	5	2	6	12	25
norm	6	4	2	4	3	0	0	3	0	2	13	23
norm + 1	0	1	0	1	3	3	0	13	12	16	25	22

One should recall that the method mentioned earlier uses the actual and simulated flow-through rates to determine dropout rates (as proportions of the repetition rates). Therefore, flow-through rates play an important role when determining dropout rates. The dropout rate in Grade 7 for all three age groups is zero. This is an unexpected result. A possible reason for this is the high flow-through rate between Grades 7 and 8. See Table 2 in this regard, where the mean of the flow-through rates between Grades 7 and 8 for the three different age groups is 102%. Thus, the number of learners in Grade 8 is considerably more than the number of learners in Grade 7, since under-reporting of the number of learners at the end of primary school (i.e., Grade 7), and the over-reporting of the number of learners at the beginning of secondary school (i.e., Grade 8). Since we chose the values between 0.00001 and 3, negative dropout rates cannot be obtained. It should be noted that very low dropout rates have been reported in published literature (Bot, 2011).
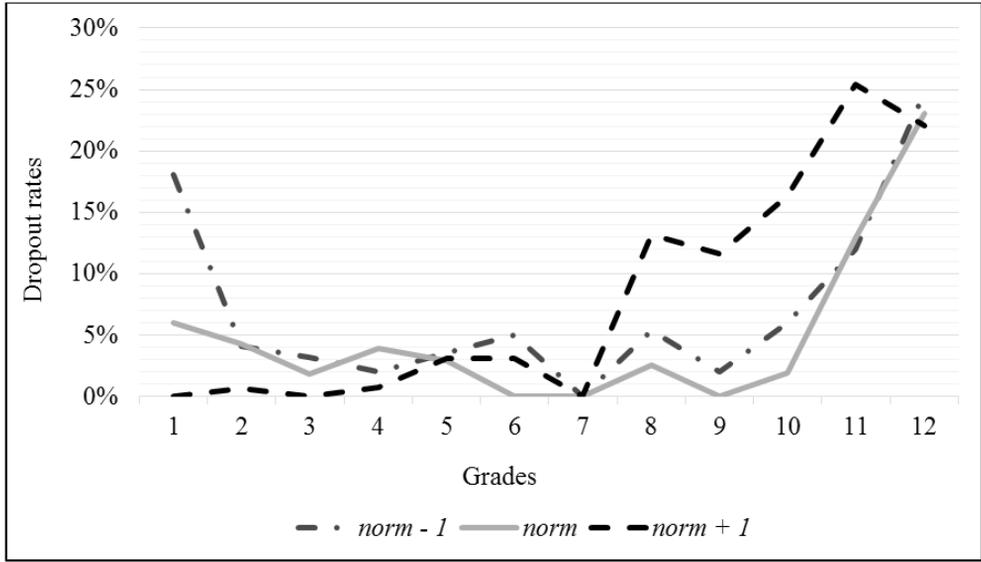
Figure 6: Estimated dropout rates for the different age groups

Table 6 and Figure 7 show the promotion rates for the three age groups. Again, the promotion rates for the $norm+$ age group are lower for Grades 8 to 11, if compared to the other two age groups. This is expected, since higher repetition and dropout rates were estimated above for this age group in Grades 8 to 11.

Table 6: Estimated promotion rates (in %) for the different age groups

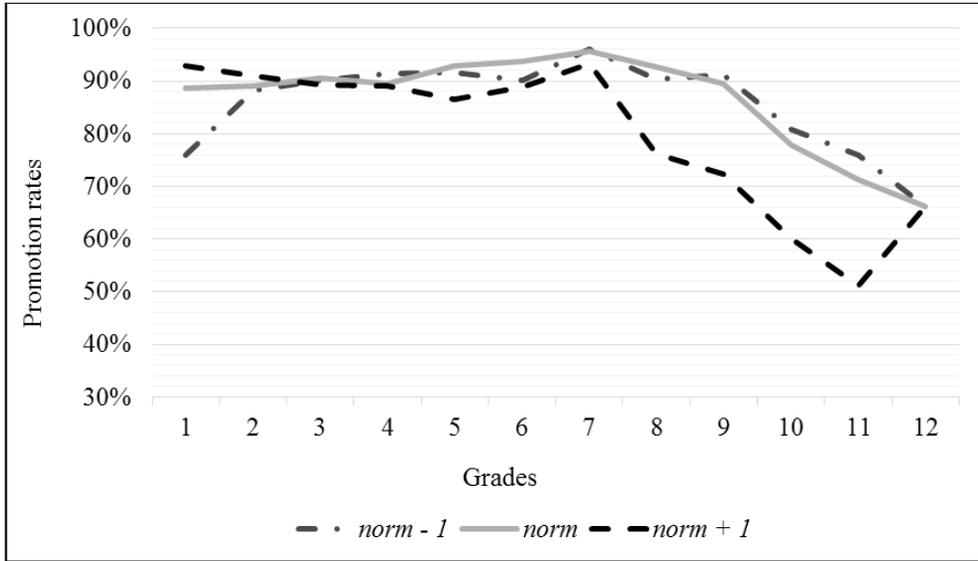| Age group	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
norm - 1	76	88	90	91	92	90	96	90	91	81	76	66
norm	89	89	91	90	93	94	96	93	90	78	71	66
norm + 1	93	91	89	89	86	89	93	76	72	60	51	66

Figure 7: Estimated promotion rates for the different age groups

Finally, Table 7 and Figure 8 show the promotion, repetition and dropout rates for Grade 1 to 12. It should be noted that these rates are given for all three age groups, by determining averages for each grade.

Table 7: Average repetition, dropout and promotion rates (in %) for Grades 1 to 12

Average rates	Grades											
	1	2	3	4	5	6	7	8	9	10	11	12
Repetition rate	6	8	8	8	7	6	5	7	11	19	17	11
Dropout rate	8	3	2	2	3	3	0	7	5	8	17	23
Promotion rate	86	89	90	90	90	91	95	86	84	73	66	66

In Figure 8, the estimated promotion rates are on the right-hand vertical axis, while the estimated repetition and dropout rates are on the left-hand vertical axis, for Grades 1 to 12 on the horizontal axis.
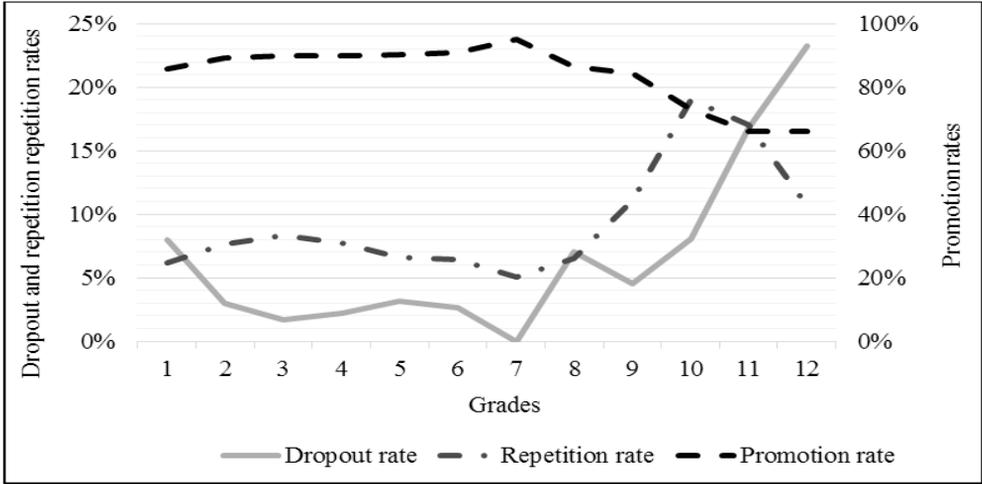
Figure 8: Average repetition, dropout and promotion rates for Grades 1 to 12

Figure 8 shows a steady increase in the dropout rates from Grades 9 to 12. The repetition rates reach a maximum at Grade 10, and decline then towards Grade 12. Reasonable stable promotion rates are observed from Grades 1 to 7, followed by a steady decline from Grade 8 towards Grade 12.

In an attempt to address the under- and over-reporting observed in the data between primary and secondary school, we smoothed the flow-through rates in Table 2. In particular, we fitted a linear trend line to the flow-through rates. The minimisation in (7) was repeated and the results are given in Figure 9.
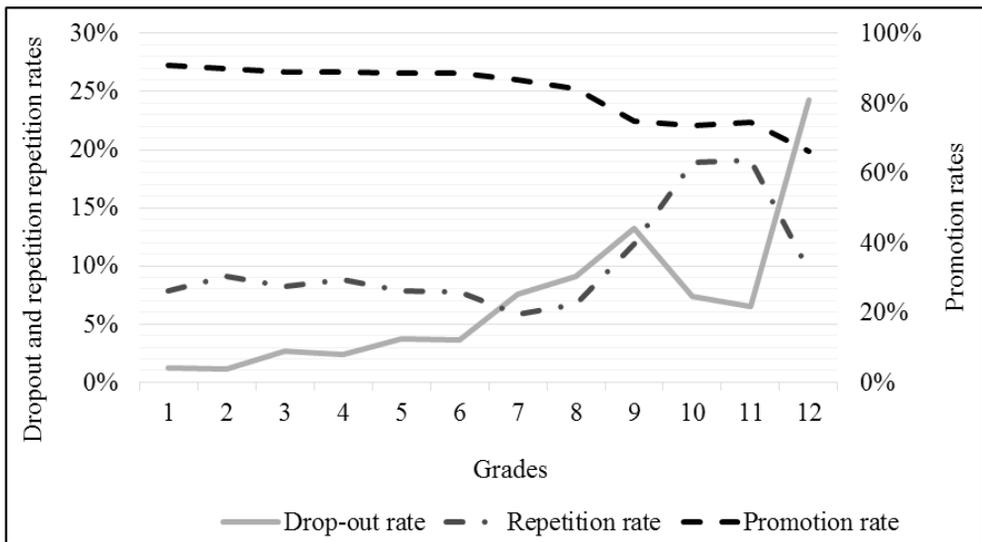

Figure 9: Average repetition, dropout and promotion rates for Grades 1 to 12 after smoothing of flow-through rates

Given the fixed repetition rates, the effect of the smoothing is evident in the dropout rates in Figure 9. This somewhat reduces the effect of the under- and over-reporting observed between primary and secondary schools. We are of the opinion that these results should be used in practice.

Final comments

In a DBE report, prepared for the Portfolio Committee on Education in 2011, on dropout and learner retention, the following dropout rates for 2007/2008, as given in Table 8, are reported.[8]

Table 8: Dropout rates (in %) for 2007/2008 for Grade 1 to 11

Grades	1	2	3	4	5	6	7	8	9	10	11
Dropout rate	1	0.5	1.2	0.3	2	1.5	2.7	2.8	6.5	11.5	11.8

The report intended to provide information on how successful the schooling system performed in South Africa, specifically with regard to retaining learners in certain grades. There are significant differences when comparing the dropout rates from the report with those estimated in this paper. Possible reasons for this are that repetition and promotion rates are calculated from data obtained from the National Income Dynamics Study Survey, and that a birth-year cohort approach is not followed in calculating dropout rates.[9] A cohort-based approach is, therefore, a necessity and even then, repeaters by age and grade still need to be taken into account. Ignoring these factors can lead to under- or over-estimation of the different rates. The report also mentions that approximately 39% of South Africa's youths obtain a National Senior Certificate. We performed further statistical analysis using the above results and found this value to be closer to 30%. The interested reader can contact the authors for further information regarding this analysis. Discrepancies can also be due to the fact that the results in this paper are based on EMIS data from the DBE extending over fourteen years. Differences can be expected when compared with results published for individual years.

When comparing our results with those in the DBE 2013 Macro indicators report,[10] we found that the repetition rates are similar, but that the dropout rates differ somewhat. This is particularly evident for the lower grades. The DBE publication reported low dropout rates for Grades 1 to 6, whereas our method gives steadily increasing dropout rates over these grades.

According to Murtin (2013), grade repetition is especially high in Grades 10 and 11. We found similar results. Murtin (2013) also states that 58% of the learners leave the schooling system without completing matriculation. We found that 70% of the learners who started in Grade 1 dropped out before completing Grade 12.

Pugatch's (2012) study concluded that nearly one third of learners are re-enrolled in school, after having been dis-enrolled for at least one year. It should be noted that the procedure in this paper takes the learners who re-enrolled into account, because such learners are included in the actual average flow-through rates, as calculated in Table 2.

We consider only learners in three different age groups for each grade, thus ignoring learners who are more than one year younger or older for a particular grade. Including such learners could lead to more accurate estimated dropout and promotion rates. Unfortunately, this will considerably complicate the estimation procedure. In addition, learner mortality is included in the dropout rates, which can be further refined if mortality is taken into account.

Repetition rates can be estimated from data sources other than the GHS. For example, the World Bank databank can be utilised for this purpose.[11] This will, of course, affect the dropout and promotion rates. In (4), average proportions calculated over all cohorts are used as starting proportions when the values in Table 3 are calculated. These starting proportions can be chosen differently. Say, for instance, that we are interested in the dropout and promotion rates of learners born in a specific year, the proportions of learners in that specific birth-year cohort could be used as starting proportions.

When the school grade progression of a learner has to be determined for the purpose of loss-of-income calculation, the repetition, dropout and promotion rates can be re-estimated for the specific case at hand. For example, data from the province in which the particular learner's school lies can be used in the estimation. The estimation can even be further refined, using only data from schools with the same socio-economic status as the particular learner's school.

The procedure outlined in this paper provides insight into the working of the basic education system in South Africa. The combination of GHS data, EMIS data, yearly reports of the DBE, and simulation, provide an innovative approach to assess the performance of learners in South African schools. As the DBE data improve and trends emerge, the procedure can be updated and improved. In this way, the effect of DBE initiatives can be assessed. The DBE has various goals (such as the "Action Plan to 2014") and the effect of the underlying actions related to these initiatives can be estimated using the proposed procedure.[12] Finally, applications of the proposed procedure are numerous and opportunities for future research are promising.

Endnotes

[i.] http://www.education.gov.za/

[ii.] http://www.statssa.gov.za/ghs/

[iii.] http://beta2.statssa.gov.za/

vi. http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx

v. http://www.info.gov.za/acts/1996/a84-96.pdf

vi. http://www.palisade.com/risk/

vii. http://www.education.gov.za

viii. http://www.education.gov.za/LinkClick.aspx?fileticket=jcSsY0rHcME%3d&tabid=358&mid=1261&forcedownload=true

ix. http://www.nids.uct.ac.za/

x. http://www.education.gov.za/LinkClick.aspx?fileticket=cjnR7hVwTVc%3d&tabid=838&mid=2824&forcedownload=true

xi. http://databank.worldbank.org/

xii. http://www.education.gov.za/Curriculum/ActionPlanto2014/tabid/418/Default.aspx

References

Anderson K, Case A & Lam D 2001. Causes and consequences of schooling in South Arica: Evidence from survey data. *Social Dynamics*, 27(1): 37-59.

Bot M 2011. An overview of selected schooling indicators 2001-2010. EduSource Data news, 1. Retrieved on 19 January 2015 from http://www.jet.org.za/publications/edusource-datanews-2011-issue1.pdf.

Burger R, Van der Berg S & Von Fintel D 2013. The unintended consequences of education policies on South African participation and unemployment. IZA discussion paper, 7450. Retrieved on 19 January 2015 from http://www.iza.org/conference_files/worldb2013/burger_r9046.pdf.

Chisholm L 2005. The quality of primary education in South Africa. Background paper prepared for the Education for All Global Monitoring Report: The quality imperative. Retrieved on 19 January 2015 from http://unesdoc.unesco.org/images/0014/001466/146636e.pdf.

Crouch L 2005. Disappearing schoolchildren or data misunderstanding? Dropout phenomena in South Africa. Retrieved on 19 January 2015 from http://www.rti.org/pubs/disappearing_school children.pdf.

Fleisch B & Shindler J 2007. School participation from birth to twenty; patterns of schooling in an urban child cohort study in South Africa. Retrieved on 19 January 2015 from http://www.create-rpc.org/pdf_documents/fleisch.pdf.

Grant MJ & Hallman KK 2008. Pregnancy-related school dropout and prior school performance in KwaZulu-Natal, South Africa. *Studies in Family Planning*, 39(4): 369-382.

Modisaotsile BM 2012. The failing standard of basic education in South Africa. Policy brief of the Africa Institute of South Africa, 72. Retrieved on 19 January 2015 from http://www.ai.org.za/wp-content/uploads/downloads/2012/03/No.-72.The-Failing-Standard-of-Basic-Education-in-South-Africa1.pdf.

Murtin F 2013. Improving education quality in South Africa. Organisation for economic co-operation and development. Economics Department working paper, 1056. Retrieved on 19 January 2015 from http://www.oecd-ilibrary.org/docserver/download/5k452klfn9ls.pdf?expires=143998616 0&id=id&accname=guest&checksum=4160319D215763C5E6D97874205B89C0.

Pugatch T 2012. Bumpy rides: Schoolwork transition in South Africa. IZA discussion paper, 6305. Retrieved on 19 January 2015 from http://ftp.iza.org/dp6305.pdf.

Sedibe M 2011. Inequality of access to resources in previously disadvantaged South African high schools. *Journal of Social Sciences*, 28(2): 129-135.

Shindler J & Fleisch B 2007. Schooling for all in South Africa: Closing the cap. *Review of Education*, 53: 135-157.