The gap between school enrolments and population in South Africa
Analysis of the possible explanations

MARTIN GUSTAFSSON

DECEMBER 2012

Stellenbosch Economic Working Papers: 23/12

MARTIN GUSTAFSSON
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: MGUSTAFSSON@SUN.AC.ZA

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

BER
BUREAU FOR ECONOMIC RESEARCH

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

The gap between school enrolments and population in South Africa
Analysis of the possible explanations

MARTIN GUSTAFSSON

ABSTRACT

In South Africa, like in many developing countries, the differences between enrolment totals, estimated by the education authorities, and the numbers of children in the country, estimated by demographers in the national statistical agency, defy easy explanations and suggest that one or both sets of statistics are inaccurate. In South Africa the gap between the two sets of estimates is substantially larger than one would expect. The typical reasons that have been found to underlie developing country data problems of this kind are discussed and their applicability to the South African data is investigated, using a variety of data sources. It is found that not clarifying the reasons behind the data discrepancies and not making necessary adjustments lead to distortions in commonly cited international development indicators that are not insignificant. It is demonstrated that analysing the various possible reasons for unexplained gaps between enrolment and population aggregates can reveal patterns that are in general useful for education planning. For instance, comparing the educational attainment of adults to enrolment patterns for children in the household data can help to gauge the extent to which the child enrolment responses are subject to typical upward biases. The analysis as a whole highlights the importance of collaboration between the education authorities and national statistical agencies to improve data collection and imputation techniques on both sides.

1   Introduction[1]

If one compares recent age-specific enrolment data collected from schools to the age-specific population estimates produced by Statistics South Africa (Stats SA), it is easy to conclude that around 14% of children aged 7 to 15, in other words children who by law should be in school, were not in school. This translates to around 1.4 million out-of-school children. However, if one examines what South African households report in various household surveys, conducted by Stats SA and other organisations, fewer than 2% of children aged 7 to 15 appear not to be attending school. The latter statistic is widely believed to be close to the truth. On the whole, this paper supports this belief. But the paper also explains how contradictions such as the one between the 14% and 2% values bedevil enrolment reporting, in particular formulations of the gross enrolment ratio, an indicator used widely in international reporting systems. The main aim of the paper, however, is to use a variety of data sources to shed light on why these contradictions exist and how to interpret them. In so doing, topics that are interesting and important in their own right, such as the extent to which less formal private schools are providing education outside the ambit of official school censuses and funding systems, are dealt with. Although this paper is about South Africa, part of its intention is to present ways of analysing enrolment versus population discrepancies in developing countries in general. The literature suggests that these discrepancies are common in these countries.

The paper approaches the problem mainly from an education planning angle, as opposed to a population demographics angle. It is not within the scope of the paper to examine the demographic estimation techniques employed by Stats SA, and other national statistical agencies. But one thing the paper seems to point towards is the need for closer collaboration between education ministries and national statistical agencies in interrogating important statistical discrepancies.

In order to present a more or less contained analysis, the paper focuses largely on the age range 7 to 15 and on data for 2009 or thereabouts. However, data for ages just below and above this range and for several years before and after 2009 would also be subject to the problems analysed in the paper, though to varying degrees. How the problem has varied over time receives some attention.

Section 2 describes the basic data contradiction. It also justifies why understanding the contradiction is important for planning, inter-governmental financing and the interpretation of commonly used international indicators. Section 2 also discusses literature that reminds us that the kinds of data problems discussed in the paper are common in developing countries. Some indication of the size of the contradiction in years before and after 2009 is given.

Section 3 presents an analysis of various household datasets which confirms that an out-of-school statistic of around 1% to 2% is plausible. At the same time measurement difficulties associated with the use of household surveys to gauge school participation are acknowledged and analytical approaches to verifying reported trends are presented.

Section 4 describes and analyses enrolment data obtained through school census collections, with a special focus on the need to take data on non-mainstream schooling into account when enrolment aggregates are calculated.

Section 5 discusses the problem, particularly common in developing countries, of the over-reporting of school enrolments. If enrolments are over-reported in South Africa, as one might

expect, then the puzzling 14% gap between enrolments and population referred to above becomes even larger. An analysis combining enrolment data from school censuses, Grade 12 examination data and household survey data is used to explore the possible extent of this problem.

Section 6 responds to existing analysis that argues that there is an enrolment *under*-reporting problem in South Africa due the exclusion of many private (or independent[2]) schools that operate on a more informal basis. Using household data, this paper argues that this phenomenon is unlikely to be as large as some have argued.

Section 7 brings together estimates discussed in the previous sections and argues that the possibility seems strong that commonly used population estimates for young South Africans are around 14% higher than they should be.

Finally, section 8 concludes the paper.

2   The puzzle described

To begin with, it is necessary to impute population figures for single ages from Stats SA's published five-year age cohort values. Stats SA publishes its mid-year population estimates on an annual basis, following projection methods promoted by international bodies such as the International Monetary Fund (IMF) and UNAIDS[3]. Mid-year estimates such as those for 2009 would to a large degree be projections flowing from the previous population census, in 2001. A census was again conducted in 2011. Statistics from the first reports emerging from the last census are considered below. Once microdata from the census become available, it would be possible to perform further analysis relating to the issues of this paper. Apart from a medium variant, Stats SA's mid-year estimates include low and high variants which, for the total national population, deviate from the medium variant by around 0.8%. Disaggregated figures by five-year age bins are available only for the medium variant.

The following graph illustrates the application of Sprague coefficients to convert five-year age bins to single ages, for 2009 but also 2010. This approach is commonly used, for instance in Ta-Ngoc Châu (2003: 106) and Stukel and Feroz-Zada (2008: 17). Stats SA's five-year age bins are 0 to 4 years, 5 to 9 years, and so on. The two continuous curves reflect population estimates for single ages for the two years after the Sprague coefficients have been applied.

---

[2] For the benefit of the non-South African reader the term 'private schools' is used instead of the official South African term 'independent schools'. Similarly, 'pupils' is used instead of 'learners'.
[3] Statistics South Africa (2009).

Figure 1: Population by age 2009-2010
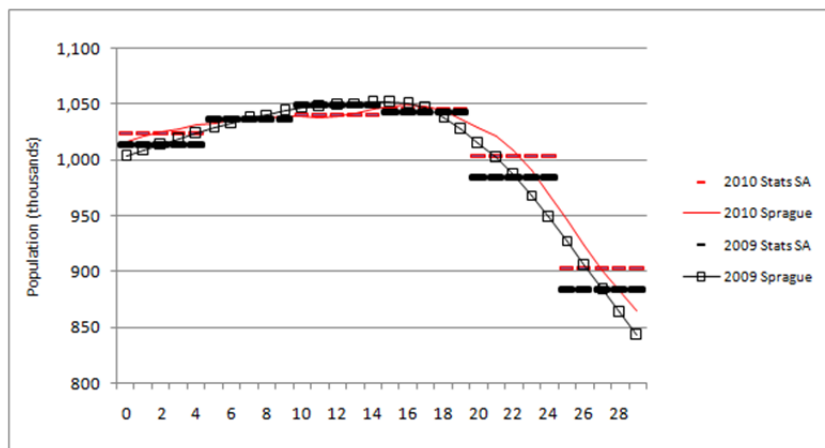
Sources: Mid-year population estimates of Statistics South Africa (Statistical release P0323).
Note: The horizontal Stats SA bars reflect the total for the five-year cohort divided by 5.

The population obtained for ages 7 to 15 is 9,421,627 and appears in the middle column of Table 1 below. That column also contains an estimate for the number of children aged 7 to 15 enrolled in education institutions, using official enrolment statistics. More specifically, age-specific data from the Annual Survey of Schools (ASS) are taken and adjusted upwards to deal with three things: the gap between the ASS totals and Snap Survey totals, exclusion of some pupils from the age tables in the ASS, though they were included in other ASS tables, and enrolment in special needs schools. The overall upward adjustment comes to just 1.7% and is discussed in more detail below. What will also be explained below is the important difference between enrolment and attendance and what this means for the current paper. What the middle column of Table 3 suggests is that only 86% of South Africa's children in this age range are enrolled and that approximately 1.4 million children were out of school in 2009. One reason why virtually no-one makes this calculation, and hence hardly anyone is worrying about the 86% ratio, is that official enrolment data by age are not widely available or analysed in South Africa. The numerator in the calculation is thus not widely known. What are widely used are enrolment aggregates just by grade (and not age). These statistics include many over-aged pupils who have repeated their grades and thus glaring differences between the enrolment and population values are not found. An example of a comparison where enrolment is not broken down by age is given in the last column of Table 1. The population value here of 12,556,855 represents the population aged 7 to 18, in other words twelve age cohorts, and is obtained using the Sprague coefficients approach. At first glance, this comparison does not raise any red flags. The 97.4 figure is in fact what would be referred to as the gross enrolment ratio for Grades 1 to 12. As implied by the discussion that follows, this 97.4 figure can be considered an under-estimate.

The first column in Table 1 uses only weighted observations from the 2009 General Household Survey (GHS) dataset of Stats SA. This survey is an annual survey involving a nationally representative sample of around 30,000 households. How the very small phenomenon of missing values was dealt with is explained below. However these issues are dealt with, the GHS for 2009 and adjacent years indicate an enrolment ratio for ages 7 to 15 that is close to 100% (it is 98.5% in Table 1). This extremely high enrolment ratio is generally

considered to reflect the situation with respect to children of compulsory school-going age. The situation in South Africa in this regard is thus considered excellent[4].

Table 1: A comparison of enrolment and population totals in 2009

	Ages 7 to 15: Enrolment and population from 2009 GHS	Ages 7 to 15: Enrolment from adjusted ASS 2009, population from Stats SA mid-year estimates	Enrolment in Grades 1 to 12 from official education reports, population from Stats SA mid-year estimates
Enrolments	9,178,880	8,068,327	12,227,963
Population	9,321,723	9,421,627	12,556,855
Enrolment ratio	98.5	85.6	97.4
Implied out-of-school	142,843	1,353,299	328,892

If one redoes the calculation in the middle column, but focussing on the age range 10 to 14, the only five-year year age bin falling entirely inside the 7 to 15 age range, and if one does this without the application of the Sprague coefficients (which are now not necessary), the resultant enrolment ratio is 85.5. This is very close to the 85.6 value in Table 1. Clearly it is not the imputations performed on the population data that are driving the low enrolment ratio.

The puzzle is why the gap between the enrolment and population values in the middle column of Table 1 should be so large. It should be emphasised that puzzles of this nature are common in developing countries, though they may not be widely acknowledged. Stukel and Feroz-Zada (2008: 7) find in a study of 16 developing countries (excluding South Africa) that in eight countries the enrolment ratio based on officially published enrolment and population statistics is higher, by at least 2 percentage points, relative to the enrolment ratio obtained from just household data. In five of these countries this gap is greater than 5 percentage points. The five countries are Bangladesh, Burkina Faso, Egypt, Guinea and Indonesia. In only two countries is the enrolment ratio based on just household data the higher one, by at least 2 percentage points. It is noteworthy that one of these two countries is Namibia, a country colonised by South Africa from 1920 to 1990 and on which South Africa imposed many of its own administrative systems, including statistical systems. In Namibia, the gap is 5 percentage points. The other country is Tanzania. In South Africa we have seen that the gap is a whole 12.9 percentage points (the difference between 85.6 and 98.5 in Table 1). The gap is thus particularly large in South Africa and runs in the opposite direction to those in most other developing countries, judging from the Stukel and Feroz-Zada (2008) study.

Stukel and Feroz-Zada (2008) identify a number of typical measurement problems underlying the discrepancies they find. Problems that would contribute to a situation such as the South African one include a tendency within household surveys for respondents to claim children attend school when they do not, the existence of unregistered schools which are not covered by the school censuses and problems in the demographic models used by the national statistical agency leading to an over-estimation of the young population. Problems that would contribute to the reverse situation found in, for instance, Bangladesh include a tendency for schools to inflate their enrolment figures in school censuses in order to attract more public funding and problems in the demographic models leading to an *under*-estimation of the young population.

Whatever the explanations, they have implications for various monitoring and even funding systems. The formula that distributes South Africa's nationally collected revenue to the nine provinces requires both enrolment data from the education authorities and population values

---

[4] See for instance South Africa: Department of Education (2008).

from Stats SA to be close to correct[5]. The last column of Table 2 indicates that, according to the household data, the picture of virtually complete enrolment for the ages 7 to 15 applies to all the nine provinces. However, the ratios between the separately collected enrolment and population values are clearly not uniform. They range from 78.1% in the case of Free State and Western Cape to 97.3% for Eastern Cape. One possible explanation is that the denominator, population, is under-estimated in the case of a province such as Eastern Cape, relative to other provinces. The Eastern Cape population value may still be an over-estimate, relative to the true population total, but this over-estimation may be less pronounced in Eastern Cape than in other provinces. This explanation would point towards a funding disadvantage in the case of Eastern Cape, and conversely, a funding advantage in the case of, for instance, Free State. A completely different explanation would be that the numerator, enrolments, in the case of Eastern Cape is inflated, to a greater degree than in other provinces, due to the kind of fraud alluded to above. This explanation implies that Eastern Cape is experiencing an unfair funding advantage flowing from corrupt practices in the collection of enrolment data from schools and that Free State and Western Cape are being penalised for their relative honesty.

Table 2: Enrolment ratios by province with respect to 7 to 15 year olds

Province	Enrolment from adjusted ASS 2009	Population from Stats SA mid-year estimates	Enrolment ratio using two previous columns	Enrolment ratio using just 2009 GHS
EC	1,345,187	1,381,845	97.3	97.9
FS	417,970	535,099	78.1	98.7
GP	1,354,462	1,632,668	83.0	98.9
KN	1,834,090	2,154,275	85.1	98.3
LP	1,023,914	1,182,691	86.6	99.2
MP	674,379	764,318	88.2	98.6
NC	185,616	227,049	81.8	98.6
NW	526,638	639,920	82.3	97.6
WC	706,071	903,643	78.1	98.3
SA	8,068,327	9,421,508	85.6	98.5

Gross enrolment ratios are widely used to gauge the degree to which countries enrol their children in the education system. UNESCO specifies that the gross enrolment ratio should be calculated by dividing official school enrolment by the official population of those ages one would expect in school if there were no grade repetition (and then multiplying by 100)[6]. The GER has been criticised for being insensitive to the presence of inappropriately aged pupils in the schooling system. To illustrate, it is possible for one country where only two-thirds of children of compulsory school-going age are in school to have the same GER as another country where all children of school-going age are in school, depending on the burden of grade repetition in the system[7]. Whilst this is true, the GER does provide an indication of whether a country has the capacity, in terms of for instance school buildings and teachers, to school all of those who should be in school. Specifically, if the GER is at least 100, then although some over-aged pupils may be taking the place of correctly aged pupils one knows that with some re-organisation, the schooling system would be able to accommodate all correctly aged children, assuming that there are no serious geographical distribution problems.

The following table provides, in the first column, the GER values obtained if one follows the UNESCO specifications[8]. The second column provides the GER values obtained if one uses

---

[5] South Africa: National Treasury, 2009: Website annexure W1.
[6] UNESCO: Institute for Statistics, 2003.
[7] Lewin, 2011.
[8] Here published and grade-specific enrolment values have been used, with special school enrolments spread across the grades in proportion to the distribution in ordinary schools.

only one data source, in this case the 2009 GHS data. In 2009, for the first time, the GHS dataset included a variable with the current grade of the pupil. This greatly facilitates the calculation of the GER as it becomes easy to separate out Grade R (the grade before Grade 1) and also to differentiate the primary and secondary levels (Grades 1 to 7 are considered primary schooling in South Africa). The differences between the two columns in Table 3 are considerable. The fact that the left-hand values are all below 100 implies that the South African schooling system was not large enough in 2009 to accommodate all children of the relevant ages. In contrast, the values derived from the GHS all indicate that there is sufficient capacity to accommodate these children. Given the arguments made in this paper, the figures in the right-hand column can be considered far closer to the truth.

Table 3: Versions of the gross enrolment ratio in 2009

	Official enrolments over population	*Using household data only*
Grades 1 to 12	93	108
Grades 1 to 7	99	114
Grades 8 to 12	85	100

The official *Education statistics in South Africa* report for 2009 provides a GER for Grades 1 to 12 of 92[9]. Clearly, this value must reflect the UNESCO methodology. The 2011 *Global monitoring report* for UNESCO's Education for All programme provides GER values of 105 and 95 for the primary and secondary levels in South Africa in 2008[10]. These values lie between the values of the two columns in Table 3. The explanation here is that the population statistics used by UNESCO are from the United Nations Population Division, which produces population estimates which are around 4% lower than those of Stats SA. The 2009 *Human Development Report* of the United Nations Development Programme (UNDP) uses, in the calculation of the 2007 Human Development Index (HDI) for South Africa, a GER that includes primary, secondary and tertiary education and equals 76.8[11]. The corresponding value if one uses 2007 GHS data is 85.6. Had the HDI used the latter and arguably more reliable figure, South Africa would have moved up a few places in the HDI ranking, from 129 to 125 (the weight attached to the GER in the index is only 0.11). In 2010 the UNDP changed the methodology for the HDI. The GER was removed and another indicator that is also sensitive to enrolment-population ratios, namely expected years of schooling, was introduced.

The following graph illustrates the very large increase in the estimated number of 5 to 14 year olds, according to Stats SA's mid-year population estimates (MYPE), between 2007 and 2008. Over these two years, the number increased by around 445,000 or 4.4%. Had the trend for 5 to 14 year olds more or less followed the trajectory of the 2005 to 2007 period, and assuming that the trend for 7 to 15 year olds, the ages focussed on in this paper, had been similar, the problem described in this paper existing in 2009 would have been smaller, though it would still have been noteworthy. Specifically, the implied out-of-school figure of 1.4 million from the middle column of Table 1 would have been just under one million. What the post-2009 trend for the 5 to 14 age population suggests is that high out-of-school estimates are also possible at least up to 2011. What is noteworthy is that the official 2011 census results suggest that the mid-year population estimates have been too high. Here the tables in the existing report (Statistics South Africa, 2012) only allow for a comparison of the 0 to 14 ages. For these ages, the Census 2011 estimate is 4.5% lower than the corresponding figure from the 2011 mid-year population estimates. If this reduction were applied to the 2009
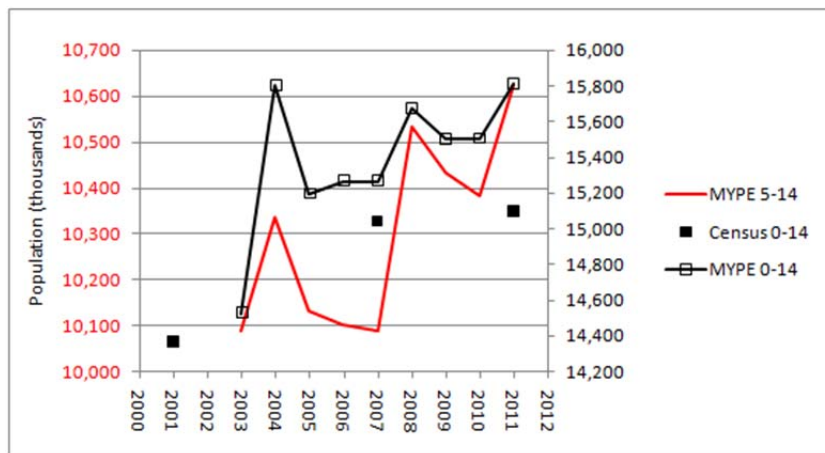
---

[9] South Africa: Department of Basic Education, 2010a: 8.
[10] UNESCO, 2011: 309, 325.
[11] UNDP, 2009: 173.

estimate for ages 7 to 15 then, once again, the problem of the implied out-of-school numbers would be reduced, by around one third, but it would still warrant concern.
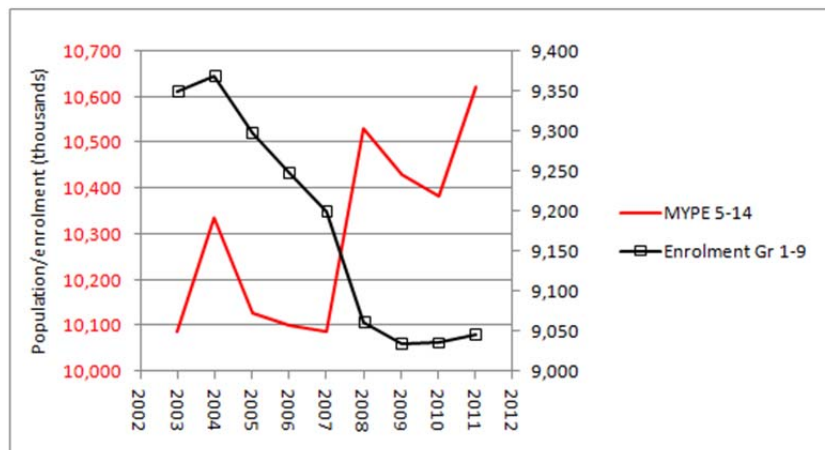
Figure 2: Estimates of young population 2001-2011

Sources: Mid-year population estimates of Statistics South Africa (Statistical release P0323). The three census points are based on statistics published in Statistics South Africa (2012).
Note: The MYPE 0-14 curve and the census points should be read against the right-hand vertical axis. The 2007 census point refers to the 2007 Community Survey, a source described in section 3.

The following graph describes the trend for enrolments beyond the 2008 to 2010 period dealt with in section 4. Here official enrolments in Grades 1 to 9, in public and independent ordinary schools, where most pupils aged 7 to 15 would be found, are illustrated. The downward trend would largely be the result of reductions in the number of grade repeaters. Analysis of the pre-2008 enrolment data broken down by age, which is not presented in this paper, indicates that there was no reduction in the enrolled numbers of children aged 7 to 15 that would influence the findings of this paper in any major way.

Figure 3: Population and enrolment 2003-2011

Note: The Enrolment Gr 1-9 curve should be read against the right-hand vertical axis.
Sources: Mid-year population estimates of Statistics South Africa (Statistical release P0323) and South Africa: Department of Education (2010).

3   A point of departure: Enrolment ratios from household data

The puzzle described above begs one crucial question: Can household responses regarding the enrolment of children be trusted? Children aged 7 to 15 must be enrolled by law in South Africa, so there would be a clear incentive for respondents whose children were not enrolled to lie and say they were enrolled. If, say, as many as one-tenth of responses indicating that children were enrolled were false, this would explain much of the puzzle. It would also completely discredit the basic assumptions underpinning education planning in South Africa. The conclusion in this section is that these assumptions are not fundamentally wrong.

The 98.5% value in the first column of Table 1 is derived from a question in the 2009 General Household Survey where the respondent is asked: 'Is ... currently attending any educational institution?'. This question is followed by examples of institutions, which include schools. Possible responses are yes, no and 'Do not know'. The last response was not provided for any child aged 7 to 15 in 2009. The 98.5% value is thus all yes responses divided by all yes or no responses, with Stats SA's household weights used. What was not considered was a code, separate from 'Do not know', described in the metadata and referring to 'Unspecified'. This code applied to only 0.3% of all household members aged 7 to 15, however.

A similar calculation to the one for 7 to 15 year olds was done for the population aged 7 to 24. That calculation resulted in an enrolment ratio that was identical to the one published in the official GHS publication for 2009[12].

Though the GHS questionnaire refers to 'attending', the yes-no nature of the question makes it clear that the focus is on enrolment. A child is enrolled if the child is officially registered with a school for the school year, whilst participation exists along a continuum between no attendance on any day in the school year and attendance every day. The concept of participation also encompasses important within-day dynamics. A child may arrive late in the school day, or depart early. As discussed by Stukel and Feroz-Zada (2008: 2), there is no clear and internationally agreed upon definition of school participation or even how to count enrolments, despite the fact that ambitious international reporting systems on this matter exist. Crucially, what is needed is better international guidance on how to reduce the impact of incentives that both households and schools may have for over-stating enrolments. In this paper, the focus is not on degrees of participation, but rather on whether a child is enrolled and attends for at least some of the school year.

The law in South Africa states that children should be enrolled up to the year in which they turn 15. If there was a large degree of enrolment inflation by households, one may expect a sharp decline in reported enrolment by households amongst children turning 16 in the current year, relative to those turning 15. The National Income Dynamics Study (NIDS) involved the collection of data from a nationally representative sample of around 7,300 households and included the collection of year of birth, something not included in the GHS dataset. The NIDS 2008 data[13] were used to check differences in enrolment ratios according to year of birth. There is indeed a marked decline, from 99.0% for those turning 15 in the current year, to 96.5% for those turning 16 (and for those turning 17 the figure is 91.5%). However, the decline from 99.0% to 96.5% is also likely to be real. There is certainly no indication of a large enrolment inflation phenomenon that gets close to being compatible with the 85.6% enrolment ratio in the middle column of Table 1.

---

[12] See Statistics South Africa (2010a: 10). Note that the national statistic at the top of this page should be ignored. It seems clear the provincial figures in the graph are the correct ones and these were used for the comparison referred to here.
[13] SALDRU, 2009.

Analysis of the NIDS 2008 data, which were collected not by Stats SA but by a unit at the University of Cape Town, points to an enrolment ratio of 99.0% for children aged 7 to 15. However, there is a no response rate of 3.2% for the enrolment question. Another useful data source is that of a 2007 nationally representative sample of around 4,400 households undertaken by Social Surveys Africa in collaboration with the Centre for Applied Legal Studies (CALS) of the University of the Witwatersrand. A report using that source[14] points to an enrolment ratio of 98.8% for 7 to 15 year olds. That these two non-Stats SA datasets should more or less confirm the findings of the 2009 GHS is significant.

Responses in the 2009 GHS on the educational attainment of adults, which would not be subject to the possible compliance biases of the child enrolment responses, confirm that enrolment ratios in South Africa have been high for many years. As seen in the following graph, amongst adults aged 20 to 24 in 2009, historical enrolment in Grades 1 to 4 had been at least 98%. Moreover, the curve for those aged 30 to 34 suggests that over time attainment of Grades 5 to 8 was improving by around 3 percentage points per decade. The two black curves labelled 'Best' reflect the future grade attainment scenario one would expect if one examines all household respondents aged 5 to 25. For 'Best 1', the percentage having attained a particular grade was broken down by single ages and then the maximum across the ages, per grade, was taken as 'best'. 'Best 2' is similar, except here if a child was enrolled in the next grade, it was assumed that the child would complete this next grade. For instance, in the case of a child whose response to the 'highest level of education ... successfully completed' question was Grade 4, but whose response to the 'which grade ... currently attending' question was Grade 5, would be considered someone who would complete at least Grade 5. Roughly speaking, the black curve suggests that by around 2019, 95% of adults aged 20 to 24 will have attained at least Grade 7. This is compatible to the trajectory of attainment improvements over time illustrated in Figure 4.

Figure 4: Educational attainment and education enrolment
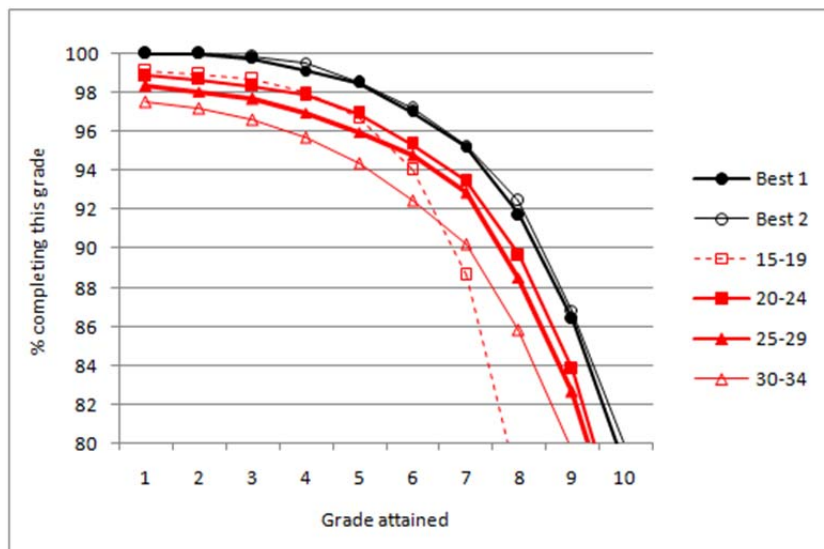
Source: 2009 General Household Survey dataset of Statistics South Africa.
Note: The curve for those aged 15 to 19 is low for Grade 6 and above because many who will attain this level have not attained it yet.

The 2009 GHS values in the highest grade successfully completed and current grade attended variables agree with each other to a high degree. For 98% of enrolled pupils, current grade is one grade up from highest grade successfully completed, which is what one would expect from good data. Nearly all of the remaining 2% is made up of pupils for whom the current

---

[14] Strassburg, 2010.

grade and the highest grade successfully completed is the same. This is probably the result of basic confusion between the two concepts during the household visit.

Figure 4 offers an important reminder regarding the level of enrolment in South Africa. Though enrolment ratios for those aged 7 to 15 may be very high, the fact that the ideal age for Grade 9 is 15 does not mean that the attainment of Grade 9, considered the final grade of basic education in South Africa, is close to 100%. The misunderstanding is reflected in, amongst other places, the pre-final version of South Africa's national development plan released in 2011 (the error was corrected in the 2012 version)[15]. The reality is that, in the context of grade repetition, many pupils stay on in school up to age 15 but do not complete Grade 9, or even Grades 8, 7 and 6. Whilst Figure 4 points to an improvement in grade attainment over time, enrolment patterns existing in 2009 were only good enough to support a future Grade 9 attainment ratio of around 87%.

The 2007 Community Survey of Stats SA warrants special attention as it has been a widely used source of enrolment ratios, yet it reveals ratios that are considerably lower than those discussed above. That the Community Survey should have been widely used is understandable. It covered around 250,000 households, against the approximately 30,000 of the GHS. A study by Fleisch, Schindler and Perry (2012) finds the enrolment ratio for 7 to 15 year olds reflected in the 2007 Community Survey to be 95.7%. The authors express reservations around the reliability of the data for whites and Indians. Analysis of the data done for this paper reveals enrolment ratios of 95.6% and 95.2% for 7 to 15 year olds in these two groups. However, the problem extends to all households with relatively educated adults, where one would expect the enrolment of children to be virtually 100%. In the following graph, the 2007 Community Survey is compared to the GHS of 2007. The gradient of the Community Survey (CS) curve seems more or less right, but across the entire range of maximum years of education in the household, the CS values seem too low by around 2 to 3 percentage points.

Figure 5: Parent education and enrolment of child
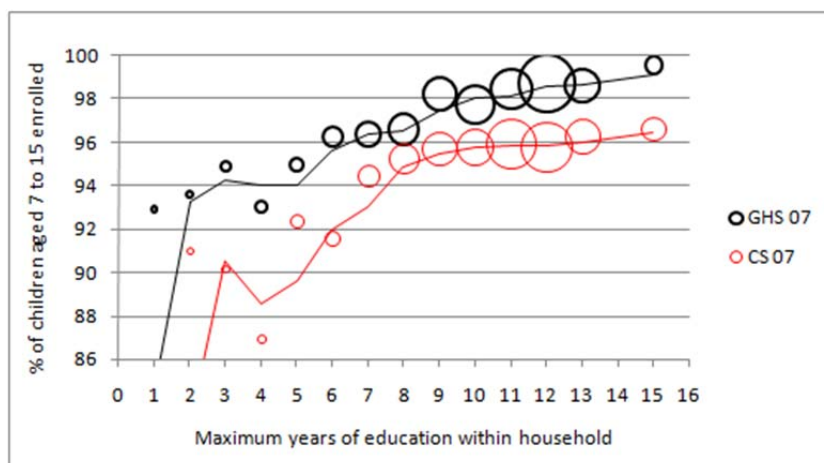
Sources: 2007 Community Survey and 2007 General Household Survey datasets of Statistics South Africa. The area of the bubbles is in proportion to the size of the age 7 to 15 population. The trend lines are moving average curves.

A logit model was run for each of the two survey datasets reflected in Figure 5 in order to see whether there was greater randomness in the case of the Community Survey in the relationship between enrolment and typical explanatory variables. Explanatory variables were highest education attained in the household, province, dwelling type and race. Each model

---

[15] South Africa: National Planning Commission, 2011: 268.

produced a pseudo R^2 value of around 0.30, suggesting there was not more randomness in the Community Survey. This confirms what was seen in the previous graph, namely that associations between variables are what one would expect, though the level of enrolment in the Community Survey seems to be consistently too low. Eyeballing of the Community Survey data revealed no patterns that might suggest what corrections should be performed. Spread fairly evenly across the dataset, even in households headed by university graduates, were children aged 7 to 15, who were not enrolled. It is worth noting that the Statistics Council expressed concern over a number of aspects of the Community Survey data, though not specifically the matter of enrolment responses[16].

4   The official enrolment data

The official source for school enrolment statistics in South Africa is the annually published *Education statistics in South Africa* of the Department of Basic Education (DBE), a series that started in 1999[17]. These reports are based on the Snap Survey, where school principals of both public and private schools complete enrolment data in a questionnaire during the first weeks of the school year[18]. In an increasing number of schools it has become common practice to enter the information onto a school computer with official software and transfer a copy of the data required for the Snap Survey from the school to the administration by means of a physical data transfer medium such as a compact disc. In around March of each year, a second and more detailed survey called the Annual Survey of Schools (ASS) is conducted. Whilst the Snap Survey disaggregates enrolments by grade but not age, the ASS involves a breakdown by both these categories simultaneously. The ASS, which has existed since the 1990s, has largely been used for analysis conducted internally by the education departments, though the DBE has recently begun using this survey for more in-depth public reports[19]. National Treasury uses the official enrolment statistics derived from the Snap Survey in the equitable share formula that determines each province's share of nationally raised revenue.

The correspondence between Snap Survey and ASS enrolment totals can be considered high. This is illustrated in Figure 6. Whilst the pattern of enrolments across Grades R to 12 fluctuates from year to year, within one year the two surveys appear to provide consistent results. In fact, the two curves for the same year are almost indistinguishable from each other in the graph. There are several reasons why patterns would change between years. One is that grade repetition patterns can change in response to policy and administrative pressures. Another is that over-reporting of enrolments by schools to attract more resources (discussed in more detail below) can change from one year to the next in response to, for instance, more stringent monitoring. Thirdly, the age at which pupils enter Grade 1 has not been entirely consistent, meaning that peaks and troughs in the per grade aggregates come about that persist within the schooling system, and move up the grades, for over a decade. These explanations are likely to lie behind certain trends seen in Figure 6, such as the decline in enrolments between Grades 2 and 5 between 2008 and 2009. In 2009, total Grades R to 12 enrolments according to the Snap Survey exceeded the ASS total by just 0.2%. In 2008, this figure was 0.3%. Comparison between the number of schools in the Snap and ASS data suggested that these differences mainly had to do with missing schools in the ASS data.
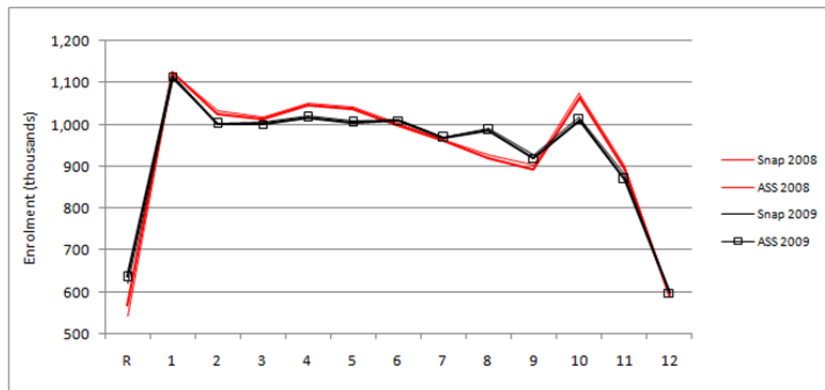
---

[16] Statistics South Africa, 2007.

[17] The Department of Basic Education (created in 2010 after the Department of Education was split) also publishes a shorter version of this publication titled *School Realities.*

[18] The school year begins in January.

[19] See *Report on the Annual Survey 2010*, published on the DBE website.

Figure 6: Enrolment by grade 2008-2009

Sources: For Snap, Annual publication Education statistics in South Africa *of the Department of Basic Education. For ASS, own analysis of the Annual Survey of Schools datasets.*

The two surveys exclude some enrolment that would be found in the ages 7 to 15 range. Firstly, they exclude enrolment in special schools, which was officially 100,717 in 2009[20], meaning 0.8% of overall school enrolments. However, this percentage may be lower for pupils aged 7 to 15 as historically the average age of pupils in special schools has tended to be higher than in ordinary schools. The percentage of 7 to 15 year old pupils in special schools could be around 0.7%[21]. Secondly, it is possible that children are enrolled in education institutions which are not considered 'schools'. Above all, policy suggests above all that one may find a small percentage of 15 year olds enrolled in Further Education and Training (FET) colleges.

FET colleges are, according to the GHS, not an enrolment location at all for children aged 7 to 15. Instead, as can been seen in Figure 7, 'pre-school' and 'other' are the predominant non-school enrolment locations. The pre-school values seem suspect if one considers the ages involved, suggesting that pre-school might appear where school should appear, or that age is incorrectly captured. Even 'other' is suspect insofar as institutional options outside schools for those aged 8 to 14 (after pre-school and before FET colleges) are difficult to imagine. Whilst home schooling exists and is legally possible, the general perception is that this is extremely limited, well below the 0.6% level for 'other' in the graph. Here again, it is possible that the enrolment is in fact in a school, for instance a special school. In the end it was decided, for the purposes of Table 1, not to inflate ASS enrolments according to assumed non-school enrolment.

---

[20] 2009 *Education statistics in South Africa.*
[21] Estimate based on special schools age distribution in Department of Education (2004, 54).

**Figure 7: Enrolment status by age according to 2009 GHS**
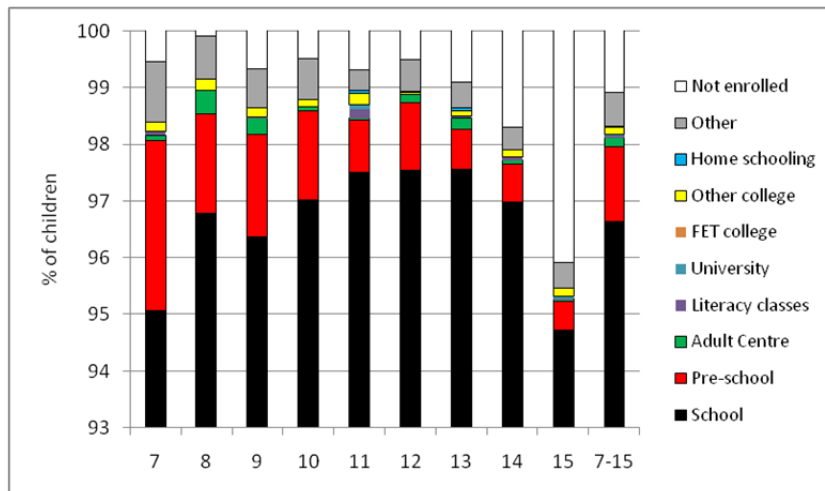
The next graph illustrates the distribution of 2008 and 2009 enrolment by age according to the ASS. The ASS patterns illustrated in Figure 6 are from a table in the survey where enrolment is broken down by just grade and gender. The Figure 8 curves are from different tables in the survey where enrolment is broken down by grade, age and gender. These tables, which involve more work for the school principal than the grade-gender table, yield lower totals. Specifically, the 2009 ASS enrolment total across the age categories in Figure 8 is 0.7% lower than the total across the grade categories in Figure 6. However, the gap is larger at the secondary (Grades 8 to 12) level than at the primary (Grades R to 7) level. The figures are 0.9% and 0.6% respectively. The gap is therefore relatively small, especially if one's focus is largely on the primary level. Nevertheless, this gap must be kept in mind when comparisons between enrolments and the population are made. The fact that the 2008 and 2009 curves in Figure 8 should be as similar as they are suggests that despite fluctuations in the distribution across grades from one year to the next, enrolment by age is stable.

Figure 8: Enrolment by age 2008-2009
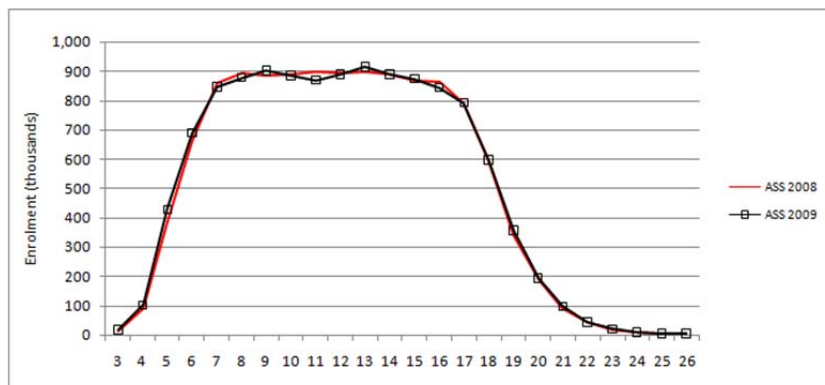
To obtain the 8,068,327 enrolment total in Table 1, 2009 ASS learners aged 7 to 15, who would be illustrated in the above graph, were adjusted upwards by 1.7%. This 1.7% was made up of 0.2% needed to deal with schools excluded in the ASS, 0.6% needed to deal with pupils

who were not counted in the ASS age tables (but were counted in other ASS tables) and 0.7% to deal with special school enrolments (the effects of compounding explain why the three do not add up to 1.7%).

5 Enrolment over-counts?

Education planners in South Africa consider it a given that fraudulent inflation of enrolment numbers by schools in order to attract more teaching posts and non-personnel funding exists. Audits to quantify the degree of inflation have occurred. However, as in other countries with similar audits, it has been difficult to obtain reliable estimates of this inflation due to matters such as pupil absenteeism on the day of audit and the enrolment of the same pupil in different schools, matters which complicate both documentary and physical headcount audits.

Probably the most reliable source of school enrolments is the Grade 12 examinations database. The population of this database occurs through processes that are separate from the ASS and Snap Survey. Grade 12 pupils must provide their 13-digit national identity numbers on registering for the examinations, which makes the duplication of pupils virtually impossible. There is moreover no incentive for duplication. The following table provides the official Grade 12 enrolments over three years as obtained through the Snap Survey as well as the number of candidates registered for the examinations. The correspondence between the two sets of statistics is consistent with a scenario of low to moderate enrolment inflation in the surveys, at least for Grade 12. Much of the gap between the two would be accounted for by the fact that around 10,000 Grade 12 pupils write the examinations of the Independent Examinations Board (IEB) and not the public examinations. Moreover, one would expect some attrition between enrolments at the start of the year (according to the Snap Survey) and registration on the examinations database as some pupils are likely to abandon their plans to write the examinations.

Table 4: Grade 12 enrolments 2008-2010

	Official Grade 12 enrolment	Number of full-time Grade 12 examination candidates	Deviation of examinations figure from official figure
2008	595,216	588,643	1.1%
2009	602,278	580,937	3.5%
2010	579,384	559,166	3.5%

Source: Annual publication Education statistics in South Africa of the
Department of Basic Education; Official 2010 examinations report
(South Africa: Department of Basic Education, 2010b).

Access to the 2010 examinations database with date of birth values allowed for an age-specific comparison between the 2010 ASS and 2010 full-time examinations candidates. The results are illustrated in Figure 9. Between 2009 and 2010, the methodology for collecting age in the ASS improved markedly. Instead of asking the school principal for a count of pupils by age, in 2010 the survey asked for a count of pupils by year of birth. This removed problems relating to the point in the year in which the survey is run. The comparison presented below uses year of birth in both sources and reflects age as it would have been on 31 December 2010. Some birth year data were missing in both sources: in the ASS 1.2% of Grade 12 pupils had no birth year against 0.6% in the examinations database. The examinations curve in Figure 9 represents 2.7% fewer pupils overall than the 2010 ASS curve. When the missing age data are taken into account, the ASS total exceeds the examinations total by 3.4%.

Figure 9: Grade 12 enrolment by age 2010
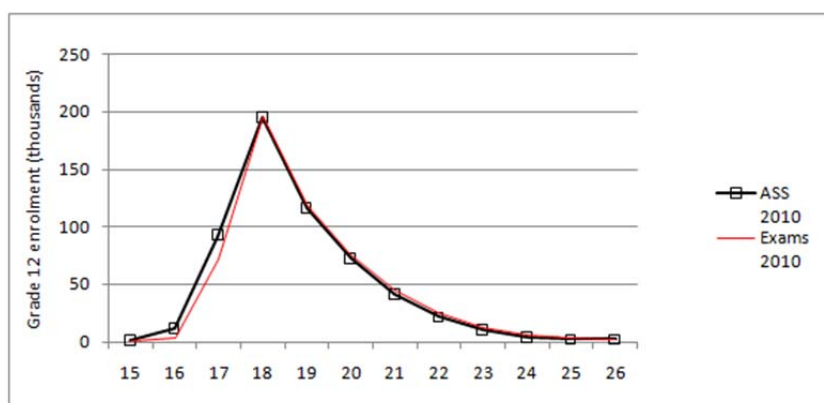
*Sources: Own analysis of the Annual Survey of Schools dataset and Grade 12
examinations database.*

When comparing two enrolment data sources, it is useful to perform the comparison at the level of the institution to ensure that differences in the totals are not simply the result of excluded institutions. When the comparison in Figure 9 was run for just those schools with data in both sources, the ASS total exceeded the examinations total by 3.0%, a figure close to the 3.4% mentioned above. The previous calculation is thus not unduly influenced by missing schools in either data source.

The ASS data, at least at the Grade 12 level, seem to reflect reality fairly well and are not distorted, at least not beyond a 3.0% margin, by fraud or poor data collection practices. As mentioned above, there are other possible explanations for the 3.0% discrepancy. This is not to say that serious over-reporting problems are not widespread at the lower grades, where the main focus of this paper lies. It is possible that school principals pay special attention to submitting correct Grade 12 enrolment figures as they know that the examinations system could be used for verification purposes. Data from the Department of Basic Education's new LURITS[22] system, when available, will allow for similar verifications in all the school grades. LURITS is a computerised system that by 2012 had been introduced in most public schools and that will eventually keep unit records of all pupils, each of whom will have a permanently assigned unique number. Moreover, the recently introduced Annual National Assessment programme, which includes controls in Grades 1 to 6 and 9 which are similar to those of the Grade 12 examinations, is likely to reduce fraudulent over-reporting.

If there was large-scale over-reporting in the grades below Grade 12, but not in Grade 12, one might expect this to be reflected in the relative sizes of the different age cohorts. Such a picture does not emerge in the following graph, where the average cohort size for the three ages 17 to 19 was made to equal 100 and other ages were adjusted to reflect relative cohort sizes within each dataset. The GHS curve reflects only those whose enrolment is in a school or pre-school. The general shapes of the two curves are fairly similar, though the curve is less smooth in the case of the GHS, something one would expect from a sample of this size. Importantly, the fact that the ASS curve is below and not above the GHS curve for the 7 to 15 age range suggests that there is not a large over-reporting of enrolments in the ASS in these ages. Households and not individuals are weighted in the GHS data, meaning that the distribution across ages in the GHS seen below would reflect distributions actually seen in surveyed households. The weights within the GHS data would thus not directly inflate the number of younger children relative to older children.

---

[22] Learner Unit Record Information Tracking System.

Figure 10: Enrolment by age in school survey and household data
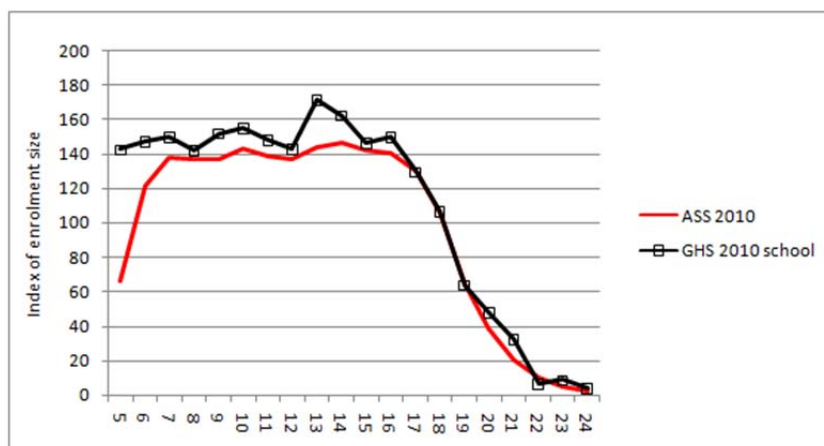
Sources: Annual Survey of Schools and General Household Survey datasets.

If we assume the population figure in the middle column of Table 1 to be correct, then the percentage, whatever this may be, of official enrolments attributable to over-reporting in the schooling system would indicate the degree to which the true enrolment ratio was *worse* than the 85.6% seen in Table 1.

6 Enrolment under-counts?

It has been argued by, amongst others, Schirmer (2010) that a substantial number of private schools serving poorer communities in South Africa are not officially registered with the education authorities and that this contributes towards an under-count in the official enrolment statistics. Schirmer (2010), in a study that focussed on six localities in three provinces, half of the localities being rural, found that 30% of schools were private schools, and this figure was then compared to the corresponding 4.5% appearing in official reports (Schirmer, 2010: 28). The conclusion was drawn that this was indicative of a large under-count of private schools in the official reports. However, elsewhere in the analysis the point was made that the sample of six localities cannot be considered nationally representative. Given this limitation, the 30% figure would not be indicative of the situation in the country. The six localities could have exceptionally high proportions of private schools, a distinct possibility given that the aim of the study was to examine what was happening in private schools. Unfortunately, the six localities were not sufficiently contained geographically and schools participated in the study on a voluntary basis, meaning a comparison of the patterns found in the study against what appears in the official master list of schools would not have pointed to anything conclusive. What is more informative for the purposes of the current paper than the 30% statistic is the finding by Schirmer (2010: 6) that a quarter of private schools were found to be unregistered. Official statistics rest on the assumption that all private schools are registered with the authorities.

Why would some private schools not be registered with the authorities? One clear incentive for being registered is that there is a pro-poor funding formula whereby private schools serving the poorest communities can receive funding as high as 60% of the average per capita public expenditure, including personnel, on public schools. Only the 20% least poor private schools would not receive any public funding, according to the formula. One can assume that private schools serving the rich would virtually all be registered as it is unlikely that the socio-economically advantaged would place their children in schools operating outside the relevant legal parameters. As pointed out by Schirmer (2010), non-registered private schools tend to be found in less advantaged areas. One likely reason behind non-registration would be the fact that registration involves complying with relatively stringent conditions relating to things

18

such as the school building and the professional registration of teachers. For some private schools, and their parents, the costs associated with compliance may be too high.

The obvious source for clarifying the true proportion of enrolment occurring in private schools is household survey data. The way this issue is dealt with in household surveys appears to have improved. In 2009 the GHS included for the first time a distinction between 'Public (Government)' and 'Private (Independent)' schools, 'Do not know' being a third option. Key statistics from the household and enrolment surveys are provided in Table 5. The fact that the proportion of schooling occurring in private schools according to the household data, across many years, is around double what one finds in the official enrolment data, suggests that the latter do indeed display a serious under-count. However, people who have worked with these data tend to believe another factor would largely explain the difference, namely confusion amongst respondents regarding what schools are private.

Table 5: Percentage of school pupils in private schools

	All ages	Just age 7 to 15
Snap 2009	3.2	
ASS 2009	3.1	3.1
GHS 2009	6.6	7.2
GHS 2010	5.8	6.1
GHS 2011	6.0	6.2

Sources: Annual Survey of Schools and General Household Survey datasets. For Snap, annual publication Education statistics in South Africa *of the Department of Basic Education.*

It is generally believed that many black households consider formerly white schools which are public but charge relatively high fees, as private schools, although they are public schools. These formerly white schools accommodate around 10% of all public school pupils and since the de-racialisation of the education system in the early 1990s, the proportion of this enrolment that is black has increased. A further institutional matter to keep in mind is that it is unlikely that there would be many non-registered schools with Grade 12 given that it is virtually impossible for a private school to operate as an unregistered institution and yet to offer the national Grade 12 examinations. What seems needed is a breakdown of private school enrolments by race and grade. The next two graphs provide this breakdown. In both graphs enrolments by grade are relative to total public and private enrolments per grade in Grades 2 to 7. The average enrolment per grade across these grades is set at 100, so for instance the 3.1 value for Grade 8, across all races, in Figure 11 is 3.1% of this figure of 100. Note that the first graph, which is based on ASS data, has a vertical axis which is less compact than the vertical axis in the second graph.

Figure 11: Private school enrolment according to the ASS 2009
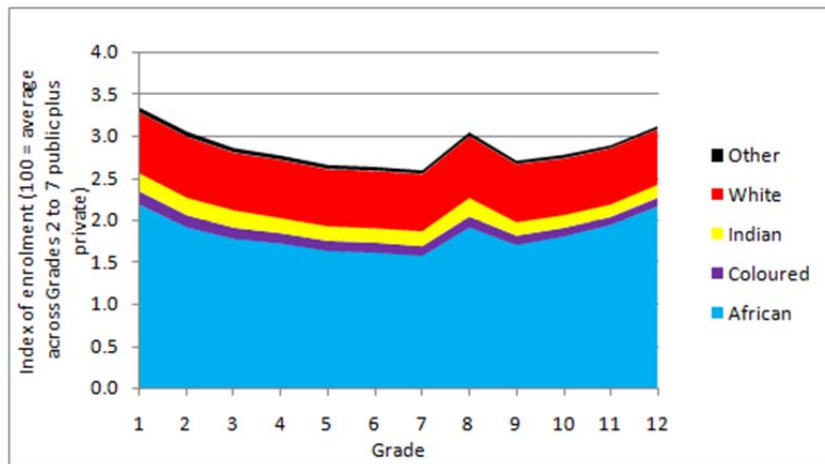
Source: Annual Survey of Schools dataset.

Figure 12: Private school enrolment according to the GHS 2009-2011
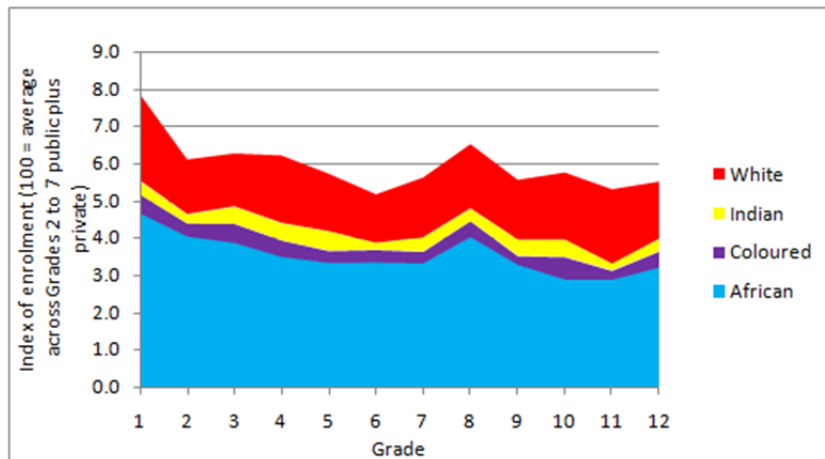
Source: General Household Survey dataset.
Note: Data from three years were pooled in order to smoothen the curves. If only the GHS 2009 were used, the general pattern would not be substantially different to what is seen above. The Grade 8 peak, which is prominent here and in the previous graph, is barely visible in the public school data. That this peak in the private system should emerge so clearly in both data sources seems to confirm that it is real.

The patterns in the previous two graphs are instructive, but they are not able to resolve conclusively the question of the degree of under-estimation of private schooling in the official statistics, at least not if one considers the institutional and definitional caveats. What is significant is that private school enrolment in the household survey data appears to be twice that in the school census data for all races. There is thus no indication in the data that African households in particular do not distinguish between private schools, on the one hand, and public schools charging high fees, on the other. If this is happening, even whites are not able to make the distinction, something many would argue is unlikely for historical reasons. What is also mysterious is that even at the Grade 12 level, where one would expect the official records not to miss any schools, the difference between the household values and the school census values should be almost as great as at other grades. The 2010 examinations dataset, which has its own separately generated variable for whether a school is public or private, was found to have a proportion of Grade 12 pupils in private schools that was similar to what was found in the ASS data.

How can the previous two graphs be reconciled? The most plausible explanation seems to be that all races are confusing private schools with public schools that charge high fees and are responding that children are enrolled in private schools, when in fact they are enrolled in public schools. This is of course very speculative. However, given that the household data yield a percentage of children in private schools that is almost twice that of the official school data even in Grade 12, it seems unlikely that the household data are accurate when they essentially say that around half of private school enrolment is not counted in South Africa. Yet, according to Schirmer's (2010) data, non-registered schools do exist, implying there is at least some degree of under-counting of private school enrolments in the official records. For the purposes of the next section, the loose assumption will be made that a quarter of private schools are unregistered across the country and that official private school enrolment figures should be inflated by 33% to produce figures that are closer to the true picture. Using the assumption that the localities selected for the Schirmer (2010) study were localities with particularly interesting private school dynamics as far as the research project was concerned, in other words places where the likelihood of finding non-registered private schools was relatively high, the 33% adjustment can be considered a maximum. There is a high probability that the upward adjustment should be smaller.

What indications are there that these private school enrolment anomalies will be resolved in the near future? Clearly, if a special effort was made to clarify the public-private distinction in the household surveys, for instance by making explicit reference to 'ex-Model C' schools as being public schools, then one might expect the picture to become clearer. The NIDS collection has asked households the names and physical addresses of schools attended by children. Whilst these details cannot be made widely available to researchers for ethical reasons, the producers of the NIDS data are attempting to link households, through the official master list of schools, to school characteristics, such as whether the school is public or private. These characteristics would become available to researchers. Yet the NIDS data are not without their own problems. In particular, it appears as if the omission of school details in the data collection process is not random, so any public-private distinction drawing from the NIDS data may be biased.

7 The residual gap: A population over-estimate?

The next graph summarises a set of explanations for the puzzle that an informed analyst using plausible values may construct. It is not the only plausible set of explanations, and a few alternatives are explained below. Note that the vertical axis of the graph does start at zero.

Figure 13: A piecing together of the 2009 evidence

21
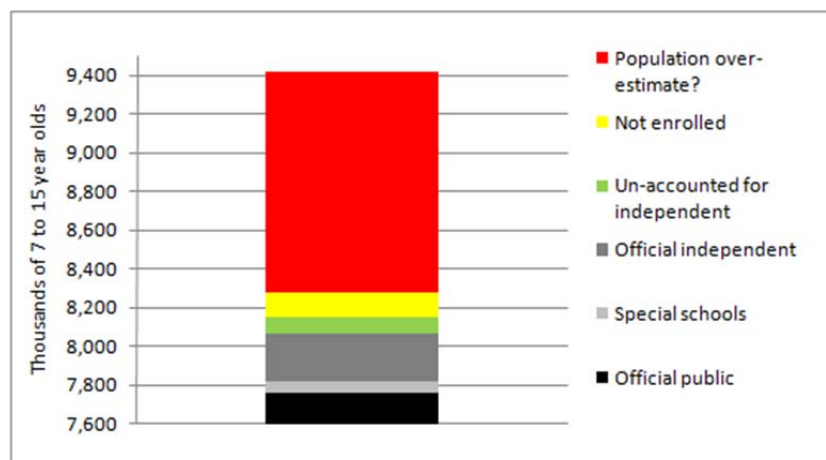
Official education sector reports, plus some analysis of the school census microdata, point to a total enrolment figure for 7 to 15 year olds in 2009 of just under 8.1 million. To this has been added just under 100,000 pupils attending non-registered private schools, meaning schools whose enrolments would not be included in the official data. As discussed in the previous section, this upward adjustment can probably be considered a plausible maximum. The adjustment should probably be smaller. Using the assumption, derived from several different household survey sources, that around 1.5% of 7 to 15 year olds are not enrolled in any education institution, the graph includes a segment for around 125,000 non-enrolled children. The official population figures in the mid-year population estimates of Statistics South Africa point to there being just over 9.4 million children aged 7 to 14. When this figure is compared to the enrolled and non-enrolled figures, is it reasonable to draw the conclusion that the 9.4 million population count represents an over-estimate of around 1.1 million children, or 14% over the implied actual figure, of around 8.3 million. Figures here are deliberately rounded in this discussion (but not in the underlying analysis) to underline the fact that one is dealing with estimates, which are subject to assumptions and adjustments, on both the enrolment but especially the population sides.

How might different estimates affect the picture provided in Figure 13? The most influential variables are the population estimate used and the possible over-reporting of enrolments in the school censuses. As discussed in section 2, two things in the Statistics South Africa reports suggest that the mid-year population estimates used for Figure 13 should be adjusted downwards. These two things are the preliminary results of the 2011 national population census and the fact that the official mid-year estimates experienced a very sharp increase between 2007 and 2008, meaning the trajectory before 2008 was more compatible with the enrolment and non-enrolment figures represented in Figure 13. Of the two, it is the 2011 census results that point to the largest downward adjustment. Specifically, they point to the need for a 4.5% downward adjustment, which would take the 9.4 million total in Figure 13 to around 9.0 million. This 9.0 million total would reduce the population over-estimate, and thus the implied over-estimate of non-enrolled children, by around 425,000 children. However, most of the problem would remain as the implied number of non-enrolled children would still be high, at around 850,000, well above what this paper has argued is a more plausible figure, of around 125,000.

As discussed earlier, some over-reporting of enrolments by schools can be considered a given. The question is how large this over-reporting is. The analysis presented in section 4 suggests that a degree of over-reporting exceeding around 3% seems unlikely. It is not concluded that there is a 3% over-reporting phenomenon, but rather that this can be considered a ceiling. If over-reporting were at this ceiling level, the picture presented in Figure 13 would change and the population over-estimate segment would grow from 1.1 million to around 1.4 million.

8   Conclusion

This paper has approached discrepancies between South Africa's enrolment and population aggregates largely from an education planning perspective, with a view to warning against simplistic comparisons of the two types of aggregates amongst those who monitor education enrolments and to checking whether key assumptions held by education planners might be incorrect.

Key conclusions that can be drawn from the analysis are the following: published gross enrolment ratios for the schooling system in South Africa, but even elsewhere, tend to be substantially incorrect because simplistic comparisons have been made. The available evidence points strongly towards a percentage of children aged 7 to 15 who were not enrolled in education institutions in 2009 that is very low, around 1.5%. Evidence, including simplistic comparisons between enrolment and population aggregates, which point to a substantially higher percentage can safely be disbelieved. Whilst evidence that might reveal the extent of

fraudulent over-reporting of enrolments by schools is hard to come by, over-reporting in excess of around 3% seems unlikely. Evidence on the exclusion of non-registered private schools from the official enrolment statistics is also hard to come by, but an assessment of the evidence there is suggests that not more than a quarter of private school enrolments are excluded. The indication provided by household data that the degree of exclusion is much larger than this appears to be the result of confusion amongst respondents as to when a school is public and when it is private. One solution here would be to tighten up the definitions in the household surveys. Lastly, it was concluded that official population estimates for 7 to 15 year olds in South Africa are around 14% to high, though this over-estimate becomes around 9% if population levels put forward by the recent 2011 national population census reports are considered.

References

Fleisch, B., Schindler, J. & Perry, H. (2012). Who is out of school? Evidence from the Statistics South Africa Community Survey. *International Journal of Educational Development*, 32(4): 529-536.

SALDRU (2009). National Income Dynamics Study: Wave 1 of 2008 (dataset). Cape Town Available from: <http://www.nids.uct.ac.za> [Accessed January 2010].

Schirmer, S. (2010). *Hidden assets: South Africa's low-fee private schools*. Johannesburg: Centre for Development and Enterprise. Available from: <http://www.cde.org.za/article.php?a_id=371> [Accessed September 2011].

South Africa: Department of Basic Education (2010a). *Education statistics in South Africa 2009*. Pretoria. Available from: <http://www.education.gov.za> [Accessed January 2011. Later and earlier issues also consulted.]

South Africa: Department of Basic Education (2010b). *Report on the National Senior Certificate examination results*. Pretoria. Available from: <http://www.education.gov.za/> [Accessed March 2010].

South Africa: Department of Education (2004). *The national audit of special education*. Pretoria.

South Africa: Department of Education (2008). *Ministerial Committee on learner retention in the South African schooling system*. Pretoria. Available from: <http://www.education.gov.za> [Accessed April 2008].

South Africa: National Planning Commission (2011). *National development plan: Vision for 2030*. Pretoria. Available from: <http://npconline.co.za> [Accessed March 2012].

South Africa: National Treasury (2009). *2009 budget review*. Pretoria. Available from: <http://www.treasury.gov.za> [Accessed April 2009].

Statistics South Africa (2007). *South African Statistics Council: Statement on the results of the Community Survey (CS)*. Pretoria.

Statistics South Africa (2009). *Mid-year population estimates of Statistics South Africa 2009. Statistical release P0323*. Pretoria. [Later and earlier issues also consulted.]

Statistics South Africa (2010a). *General Household Survey 2009*. Statistical release P0318. Pretoria.

Statistics South Africa (2010b). General Household Survey 2009 (dataset). Pretoria.

Statistics South Africa (2012). *Census 2011 (Statistical release P0301.4)*. Pretoria. Available from: <http://www.statssa.gov.za> [Accessed November 2012].

Strassburg, S. (2010). *Access to education in South Africa: Submission to the Portfolio Committee on Basic Education*. Johannesburg: Social Surveys Africa. Available from: <http://edulibpretoria.files.wordpress.com/2008/01/access-to-education-in-south-africa-social-surveys-feb-2010.pdf> [Accessed November 2012].

Stukel, D.M. & Feroz-Zada, Y. (2008). *Ten data quality studies on the impact of using household surveys versus administrative data combined with population census data for monitoring educational participation*. Montreal: UNESCO Institute for Statistics. Available from: <http://unesdoc.unesco.org/images/0018/001875/187548e.pdf> [Accessed July 2011].

Ta-Ngoc Châu (2003). *Demographic aspects of educational planning*. Paris: UNESCO: IIEP. Available from: <http://www.unesco.org/iiep> [Accessed October 2005].

UNDP (2009). *Human Development Report 2009: Overcoming barriers: Human development and mobility*. New York.

UNESCO: Institute for Statistics (2003). *Education indicators: Technical guidelines*. Paris. Available from: <http://www.uis.unesco.org> [Accessed January 2005].

UNESCO (2011). *Education for All global monitoring report 2011: The hidden crisis: Armed conflict and education*. Paris.