Disappearing schoolchildren or data misunderstandings?
Dropout phenomena in South Africa

Luis Crouch[1]
December 2005

1. Introduction and summary

Towards the end of 2005 a controversy erupted in the South African popular and technical media. "734 000 children disappear between Grades 1 and 6 and we do not know where they are", "half of pupils do not reach Grade 12", "there are more than one million children of school-going age out of school", "only 65% of children reach grade five", "there is a huge dropout problem", "all this means that the education system is contributing to perpetuating social inequality" were (with our paraphrasing) some of the head-wagging and tongue-clucking assertions. Whilst statements based on these sorts of factoids make good press, and serve to draw the sorts of attention-to-self on which politicians and academics thrive (and which is an essential feature of a lustily robust democracy), it is important to take a careful and dispassionate view of this controversy using the best data possible. It is important, in Edward Tufte's words, to elevate the ratio of facts to ink. Which of the factoids mentioned above might be close to true, and which of them are largely simple mistakes in understanding education systems and education data? And if South Africa has *some* dropout problem, how bad is it compared to that of other countries? It is unproductive to reason in the abstract and to berate oneself for not reaching an unreasonable degree of perfection, because this distracts the system's attention from the real problems.

To summarize and anticipate our analysis, and to address the factoids above, we can note the following.

1. Nowhere close to 700 000 children disappear from the system between Grades 1 and 6. On the contrary, if the number of children in grade 6 in, say, 2003, had been the same as the number of children in Grade 1 six years prior, or 1 600 000, then this fact,

[1] Senior Economist and Research Vice President, RTI, North Carolina, USA. This research was made possible through services provided by USAID to the Department of Education through the Integrated Education Project (IEP), and through the collaboration of numerous colleagues at the Department of Education.

rather than the fact that there were only 1 000 000 or so children in grade 6 in 2003, would have been the real disaster, not its opposite. This is because there were only about 1 000 000 children in the population of 7-year-olds, and thus an enrollment of 1 600 000 in, say, 1996, represented a huge over-enrollment, and carrying that over-enrolment through to grade 6 would be a terrible inefficiency.

2. Nor is it even close to true that only 65% of children reach grade five. This belief, as well as the one above, is based on a classical and simple misunderstanding regarding education data or, indeed, unfortunately, on data from the apartheid and early post-apartheid period that were particularly poor (though poor in a manner very typical of middle-income countries) and cannot be taken uncritically. Again, the problem arises in looking at an enrolment of 1 600 000 in Grade 1, which contains a lot of repetition due to under-age intake, and assuming that, if there are only 1 000 000 in, say, Grade 2, the year after, then 600 000 must have dropped out. It simply does not work that way. The difference is mostly due to a large number of repeaters in Grade 1, not to dropouts between Grades 1 and 2. In spite of the age-grade norms introduced in the late 1990s, South Africa still suffers from a lot of repetition in the early grades, and certainly did in the mid 1990s (see below). Much of this repetition is not reported as repetition (even with an improved definition in the questionnaires sent to schools) because it is not based on academic failure. Enrolment is artificially increased in Grade 1 for several reasons. Parents in many areas have no access to Early Childhood Development (ECD) opportunities and principals face moral pressure to accommodate these children in an informal ECD which consists of early admission to Grade 1, but with the expectation that the child will sit in Grade 1 two years. Furthermore, the system contains incentives to boost enrolment, and it is easier to enrol a child early in Grade 1, and make him or her repeat the grade (if the child does not repeat, then the enrolment-boosting does not take place, except temporarily, and principals know this), than to entirely make up a non-existent child in later grades. These children were (and are) typically not reported as repeaters. They appear, instead, as new intake, but twice. Whilst much of this sort of practice has been cleaned up from the system since the mid 1990s, some of it persists, and, in any case, much of the argument is based on Grade 1 data from the mid 1990s.

3. The statement that about half of all pupils do not reach Grade 12 is, of all the various assertions, closest to the truth, but the real number appears to be closer to 40% than to 50%. Hence, some 60% do reach 12th grade or its equivalent in Further Education and Training (FET). Some error in these assertions is due to the failure to take into account the fact that some youth do not want to go to Grade 12, but go on to FET instead, and FET is often not reported in the same data sources as general education.

4. Whether there are more than a million (or anywhere close) children of school-going age out of school depends entirely on what one means by “school-going age” and on whether one takes a rather generous (for a middle-income country) definition of this age bracket. If one takes the “legal” definition of school-going age, then there are closer to 220 000 children out of school.

5. Next, compared to other middle-income countries, it would be entirely unreasonable to classify South Africa's dropout problem as "huge," though obviously there are some dropouts. South Africa does a little better than other middle income countries, and certainly better than its SADC neighbours.

6. And finally, it is somewhat tendentious to claim that South Africa's education system is merely reproducing social inequalities. The system is not fighting inequalities as hard as it could be (partly because of distracting debates such as this one?), but it is hardly contributing to the reproduction of social inequalities. The inequality in years of achievement, or in learning results, is much lower than the inequality of parental income, for example, and has been decreasing markedly over the past decade.

So it is important to lay these matters to rest, hopefully once and for all, so as to give space to more useful debates.

If dropping out as a whole is not a problem, except perhaps in the final few grades of the R-12 system, then what is? What is a more useful debate? The main problem is the extremely unequal distribution of learning achievement *of those enrolled*, as well low learning achievement on average. Of all countries for which more or less rigorous measurements could be made, South Africa has the greatest *inequality* of learning achievement, by far, particularly in mathematics and science, as well as having rather low achievement even amongst those who achieve fairly well. One of the likely reasons why serious dropping out does not show up until the very last few grades of the R-12 system is, most likely, because the only meaningful measurement and accountability in the system is faced at Grade 12, and never earlier. This, ironically, increases the inequality of achievement in the earlier grades, and reduces achievement overall. Lack of rigorous standards, and of assessment against standards, affects the poor far more than it affects the well-off, and creates inequality. The rich find myriad ways to self-protect against the lack of public standards; the poor cannot. As has been shown in countries that recently introduced standards and manage to get teachers to teach to the standards, such as Uruguay (by using assessment and feedback techniques), the poor benefit much more than the rich from the use of such standards, because they are further from the standards than the rich to start with, and because teachers in poor areas are in greater need of assessment- and standards-based guidance and higher expectations.

The rest of this paper is divided as follows. First, we lay out the basic facts on the size of the dropout problem, using a variety of sources of data that are hard to fudge and mis-interpret. Second, we note some of the reasons why children do drop out. Third, we make some international comparisons, to see whether the problem that does remain is truly large, from a realistic rather than abstract perspective. Finally, we note that the education system's results are much more equal than the society itself, and thus education is most likely contributing to equality, though it could be contributing more.

2. Will the dropouts please raise their hands?

The main problem with counting dropouts is, of course, that they are not "there" to be counted. Thus, invariably, indirect techniques are needed to assess their numbers. The fact that indirect techniques are needed is what leads to much confusion. One indirect but naïve technique is to compare Grade 1 enrolment at some point, with enrolment in Grade x, but x years later. But this is treacherous because of the repetition problem already discussed, and because of the fact that even if one tried to take into account repetition, repetition is poorly reported and thus poorly estimated. We have already seen what sorts of mistakes this can lead to. (Note that in the international indicators to which South Africa contributes, such as the data on flow-through to Grade 5, and which have been the source of some of the confusion, the data on repetition are taken at face value. That is, repetition is under-reported and under-estimated. This then makes the net or new ernolment in Grade 1 seem very large, which leads to the conclusion that many children are dropping out. But the repetition data cannot be taken at face value. This is a well-known problem in middle-income and low-income countries with poor record-keeping and much use of Grade 1 as informal ECD.)

A better indirect technique is just to compare enrollment to population numbers *at any given point in time*. This is possible in two ways. First, using official enrollment counts, from administrative data (the data from the "EMIS"—the Education Management Information System), in the numerator, and population data, based on analysis and projections of the census, in the denominator. The problem with this method is that the sources of data are quite different, that there are incentives to misrepresent the data (e.g., to exaggerate enrollment, since funding and other advantages flow from exaggerating enrollments), and that the population data are not true counts but are based on analyses and projections. Second, one can rely on household surveys that take counts of the children and simply ask whether the children are enrolled or not. Fortunately South Africa has a good household survey with sufficient questions on enrollment and dropouts. And, even more fortunately, the two sources of data (official and survey) more or less coincide to make the same points.

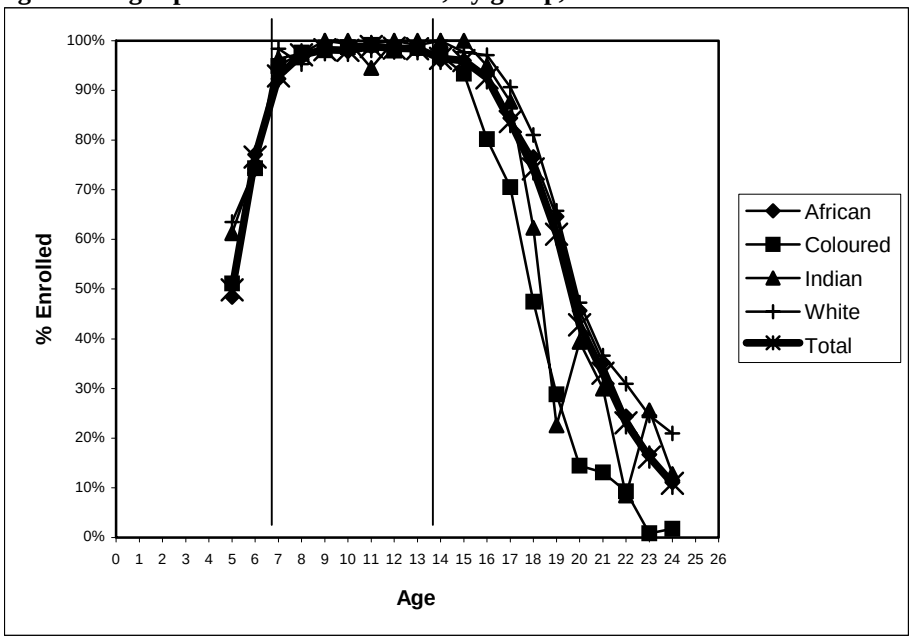
The simplest exercise is to take the numbers of enrolled children, by age, and compare this to the total number of children. Whilst using ratios such as gross and net enrollment ratios is of some use, they tend to hide more than they reveal. A more intuitive and at the same time correct approach is to simply look at the age-specific or grade-specific enrollment ratios. This can be done with South Africa's General Household Survey. Though a 2004 version is available, we have chosen to use 2003 so as to make it more compatible with other sources of data. As Figure 1 shows, in the compulsory age range (shown by the vertical lines at ages 7 and 14) **enrollment is nearly complete: 97%**. Not only that, but there are also two fairly large "tails" to the right and left of the compulsory range which essentially add 100% to the enrollment found in the compulsory range. There are interesting facts in the pattern after about age 15: Coloured and Indian youth do drop out earlier than White and African youth. (The spikes in the lines for the Indian population are an artifact of the sample size and should be ignored. Just the general trend

is important.) This probably has to do with sociological reasons as well as with school-flow practices still left over from the old HoR and (perhaps) HoD systems, but more research would be needed to know exactly why this happens.

If one simply sums the height of the curves at each age, one gets what educational statisticians call the "school-years life expectancy," namely, the total number of years of education the system is offering children at present. In South Africa this is a rather-high 14.3 years of education, and this does not differ much by race. In fact, Africans achieve 14.5 years of education, and Whites 15.2, with the Coloured and Indian populations achieving "only" 12.7 and 14.1, respectively.

The percentages enrolled can be used to calculate the percentages non-enrolled. This then can be applied to the total population to get the numbers of children not in school. The calculation is, as one would expect from glancing at Figure 1, extremely sensitive to the age-range one takes. If one takes only the compulsory age range (7 to 14) then there are some 220 000 children out of school. If one takes the 12 nominal years encompassing the 12 years of the 1-12 school system, namely 7 through 18, then there are some 725 000 children out of school. And, finally, if one takes the very generous range of 5 to 18, then there are some 1 500 000 children out of school, but this is because in the early-childhood development zone, namely 5 and 6, there are some 750 000 children out of school. It is indeed the policy of the government to reduce these numbers by offering a full grade R, and poverty-targeted earlier care, but it is important to note that in spite of these numbers, some 50% and 77% of 5- and 6-year-olds, respectively, are already claimed by their parents to attend some form of pre-school or day-care.

Figure 1. Age-specific enrolment ratios, by group, 2003



Source: General Household Survey 2003

The data presented in Figure 1 are also shown below, in Table 1.

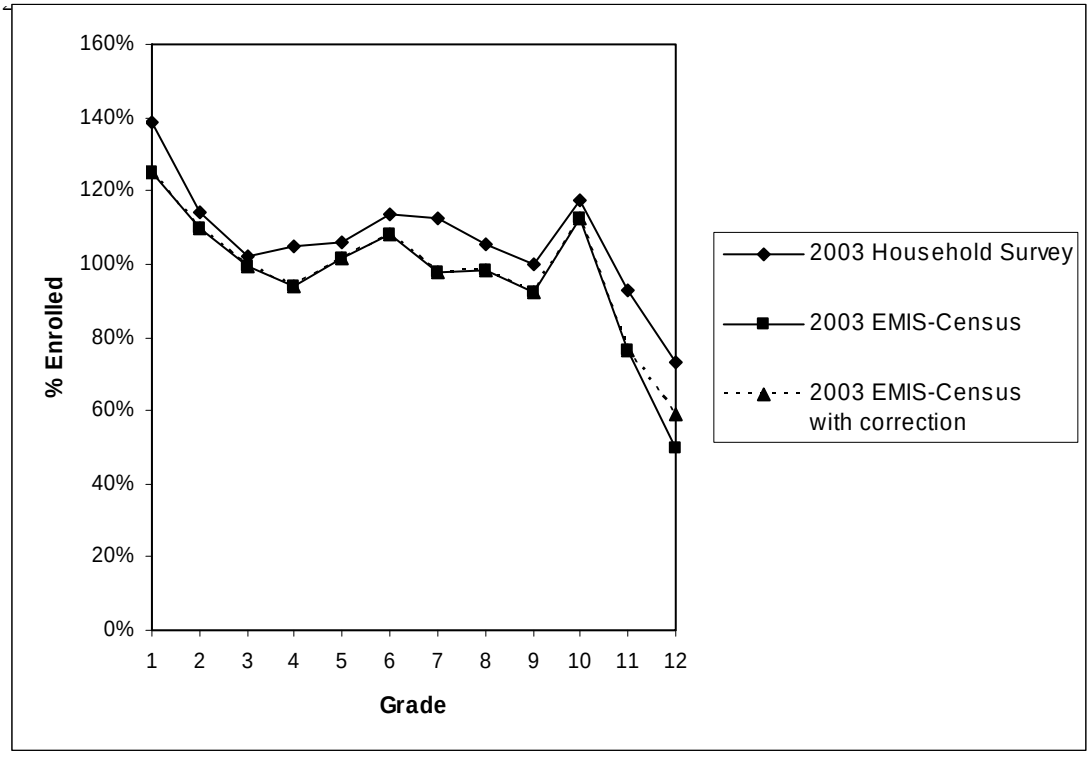
Table 1. Age-specific enrolment ratios, by group, 2003

Age	African	Coloured	Indian	White	Total
5	48%	51%	61%	63%	50%
6	77%	74%	75%	74%	77%
7	92%	95%	96%	98%	93%
8	97%	98%	97%	95%	97%
9	98%	98%	100%	100%	98%
10	98%	99%	100%	99%	98%
11	99%	99%	94%	99%	99%
12	98%	98%	100%	100%	99%
13	98%	99%	100%	99%	98%
14	96%	97%	100%	100%	96%
15	96%	93%	100%	98%	96%
16	93%	80%	95%	97%	93%
17	84%	71%	88%	91%	84%
18	77%	47%	62%	81%	74%
19	65%	29%	23%	66%	61%
20	46%	14%	39%	47%	43%
21	35%	13%	30%	37%	33%
22	24%	9%	8%	31%	23%
23	17%	1%	26%	25%	16%
24	11%	2%	13%	21%	11%
School-years life expectancy	14.5	12.7	14.1	15.2	14.4

Source: Calculated by the author using the 2003 General Household Survey

Whilst children do persist in the system through to a fairly advanced age, they do not necessarily complete as many grades as their persistence would imply, because there is still a considerable amount of repetition. This can be shown by looking at enrollment proportions by grade rather than by age, and is done in Figure 2.

Figure 2. Grade-specific enrolment ratios, 2003, with source comparison



Source: see Table 2

As can be seen, the data sources mostly agree. The survey data put enrollment a little higher, but the patterns track each other almost perfectly—the correlation between the two sources is 96%. (And this is, in itself, a key source of mutual validation of both data sources.) The key difference, at least for understanding cumulative dropping out through the whole R-12 cycle, is in the last grade. Here the difference arises, most likely, because the survey data show all forms of enrolment in Grade 12-equivalent situations including FET and part-time forms of involvement in the education sector, whereas the EMIS data do not. Note that, as explained above, in spite of various types of clean-up in the system, the over-enrollment in Grade 1 persists, most likely due to the lack of ECD facilities. There is also a backup-up of enrolment due to repetition in Grade 10. This is most likely caused by a flow-through control imposed by schools as the pressure of performance on the Senior Certificate Examination has increased. The data used in Figure 2 are also shown below, in Table 2. The total enrollment in Grade 12 or Grade 12-equivalent as a proportion of the population thus seems to be about 75%.

[2] The reasoning is thus. There are about 1 million 18-year olds in 2003. EMIS reports an enrolment in Grade 12 of some 500 000, for a ratio of about 50%. But the survey data show a ratio closer to 75% in Grade 12 or Grade 12-equivalent. Now, there are some 143 000 Full-Time Equivalent students enrolled in FET. If one apportions this approximately as 70 000 to Grade 12-equivalent, then this explains about 7 percentage points of the 25-point difference. (Unfortunately there is no rigorous way to carry out this apportionment. We have assumed about half are doing N3 which is taken to be equivalent to Grade 12.) From Senior Certificate data we also know that there are some 20 000 students who write the SC exam on a part time basis in the maxium-enrollment age cohort. This accounts for perhaps 2 percentage points. These corrections form the basis of the third line in Figure 2.

Table 2. Grade-specific enrolment ratios, 2003, with source comparison

Grade	2003 Household Survey	2003 EMIS-Census
1	139%	125%
2	114%	109%
3	102%	99%
4	105%	94%
5	106%	102%
6	113%	108%
7	113%	98%
8	105%	98%
9	100%	92%
10	117%	113%
11	89%	76%
12	72%	50%

Sources: calculated by the author from the 2003 General Household Survey, 2003 Education Statistics at a Glance, and population data "back-cast" two years using simple ratios from population projections of StatsSA.

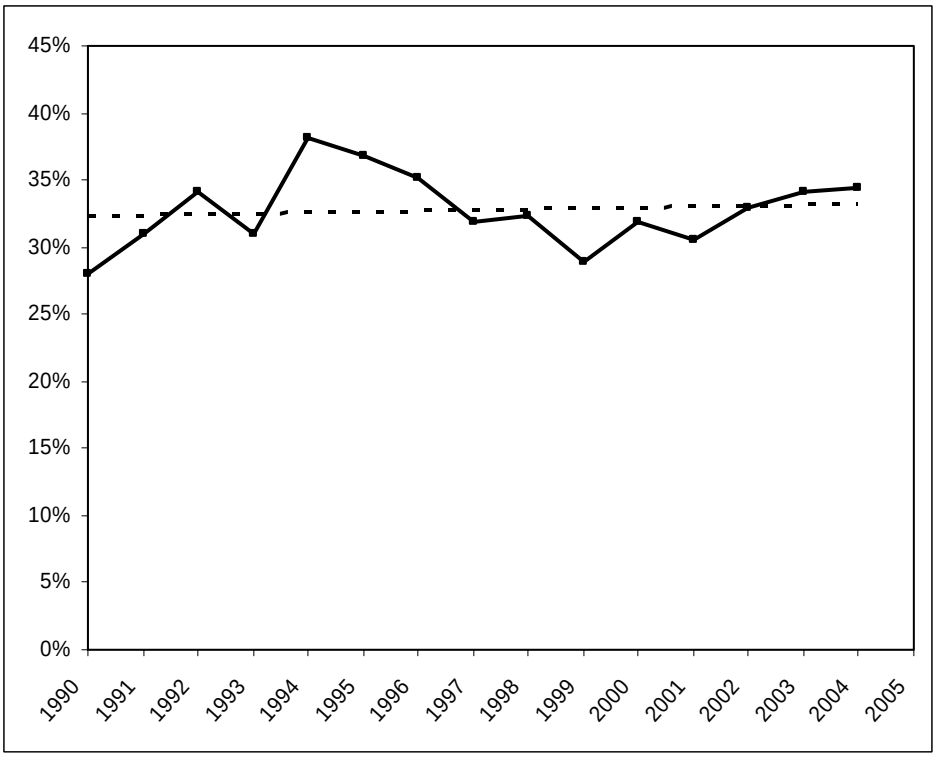
Whilst it is safe to conclude from the Household Survey data that enrolment in Grade 12 or its FET equivalent is 75% of the size of the cohort of 18-year-olds (rather than the 50% from the EMIS-Census data), it is inappropriate to conclude that 75% of the cohort "makes it" to Grade 12, because of repetition. Repeaters show up twice, as it were, in the enrolment data, so one has to subtract them in order to get a sense of the net numbers of learners actually making it up to Grade 12. If one adjusts for fairly high repetition in Grade 12 using historical and calculated data (not having exact repetition data for Grade 12 in 2003 or good data in any other year), the number that "makes it" to Grade 12 is between 60% and 65%. This is lower than the numbers in OECD countries, but is not far distant from what one finds in other middle-income areas such as Latin America. (Unfortunately there are no systematic sources of international data on this parameter. From country-specific research we know that such numbers for Latin America are typically in the same range as South Africa's. For example, Perú's rate is 65%, and Mexico's is 70%.)

3. Is there a time trend? Is the country doing better or worse over time?

There appears to be basically no trend in most enrolment ratio data, except perhaps a trend for there to be less enrolment, as a proportion of population, in the older (post secondary?) age groups.

But perhaps the most robust indicator of total system throughput is to take the ratio of Senior Certificate passes (an essentially one-off event, therefore containing very little repetition) to population of 18-year-olds. This could be called the pass ratio. This has shown a slight improvement over the long term, as can be seen in Figure 3.

Figure 3. Long-term Trend in Ratio of Senior Certificate Passes to Age-18 Population ("Pass Ratio")



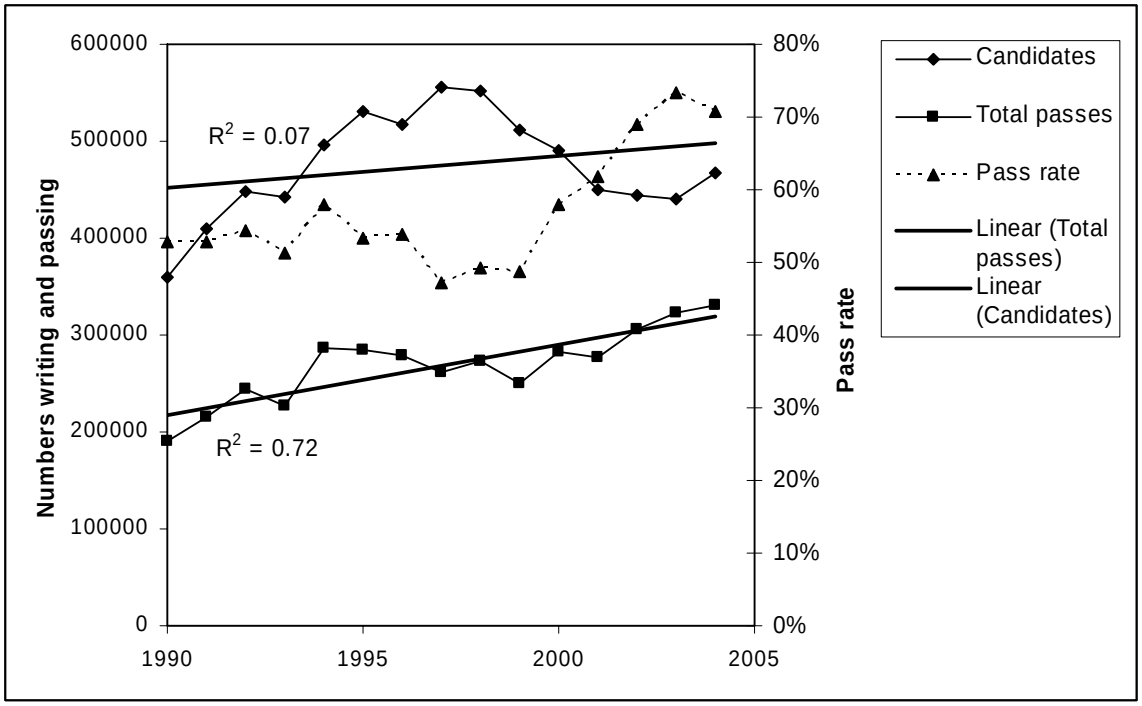
Sources: SC Passes are sourced from EduSource Datanews, varios issues since 1990. Population is sourced from the ASSA demographic model given that this source contains smoothed population data starting in 1985. Single-year estimates were produced using Beers' coefficients.

There was a decline from the mid-1990s to the late 1990s. This was due to a decline in the pass *rate* (the classical notion of passes over writers) which led to pressure to increase the pass rate, which resulted in a decline in the number of writers. The decline in the pass rate at first, and then a decline in the number of writers in spite of an increase in the pass

rate, led to a decline in the pass *ratio* that lasted until about 1999. Since then, things have been improving quite consistently. Note that even though the pass rate declined in 2004, the ratio did not, because there were so many more writers. For purposes of assessing total system throughput the ratio is almost certainly a better indicator. The dotted line shows that the long-term trend is slightly up, but, in terms of statistical significance, essentially flat.

Figure 4, though a bit complicated, throws a lot of light on what has been happening with the Senior Certificate Pass rate over the past decade or so. Studying it will pay off in insight. First, note that the numbers writing, or the numbers of candidates, are highly variable around the trend, as noted by the fact that its R^2 index (a measure of how much an indicator deviates from its trend) is only 0.07. There seems to have been little control or planning of this variable. The numbers of passes, on the other hand, have an R^2 of 0.71, or ten times the predictability that the numbers writing have. It is as if the number of passes were driven by more fundamental social or bureaucratic (e.g., quality control) forces, whereas the numbers writing were subject to short-term trends and are not very easily controlled (or may be over-controlled, which is another way of saying it is not easily controlled). The pass rate then is simply a residual, which is why the pass rate is, at least for the last ten years, almost the exact mirror (opposite) image of the numbers *writing*, not of a tracking image the numbers passing. When pressure is put on the pass rate, though, the system adjusts by altering the numbers writing, because the numbers that pass are too unchanging since they are driven, as they seem to be, by a social reality or a quality control system that is rather inflexible, in spite of some tweaking and leniency. This reaction sets in the middle 1990s and particularly in 1999. This then improves the pass rates, as those who are unlikely to have passed are kept from writing. The number of passes itself continues to improve, but essentially at its long trend.

Figure 4. Relationship Between Writing, Passing, and Pass Rate

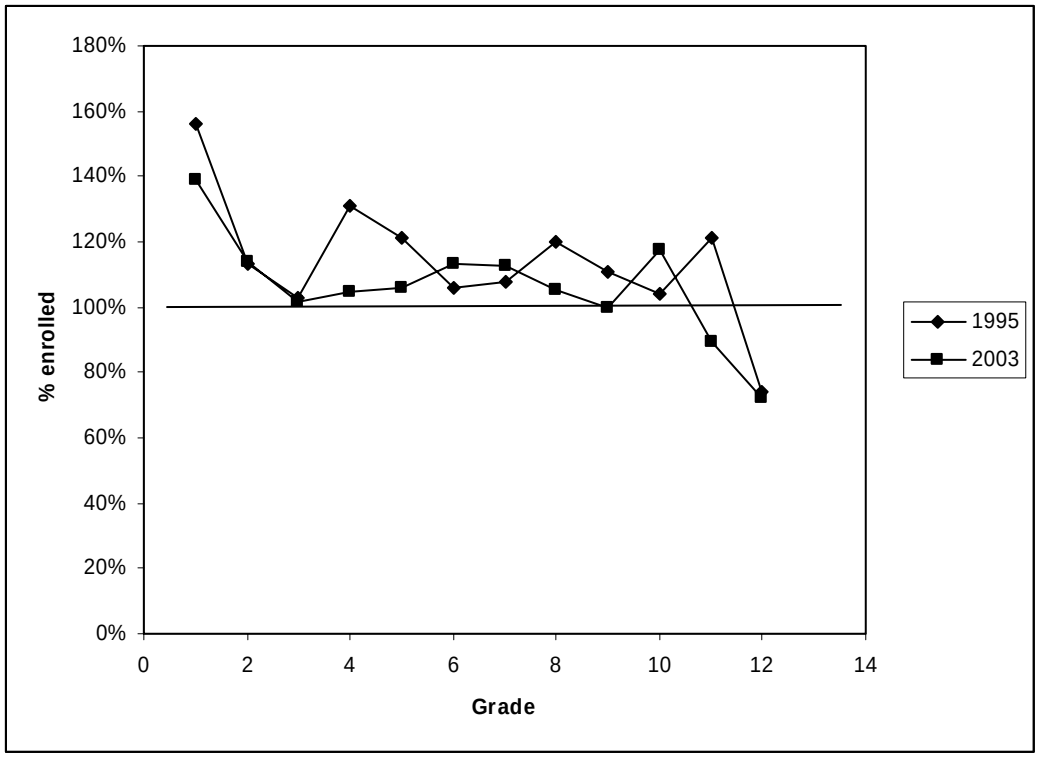


Source: compiled by Jennifer Schindler from historical EduSource sources, personal communication 14 December 2005

That said, it is important to note two other things. First, over the long term, the numbers passing have a steeper trend than the numbers writing, so the pass rate tends to go up. Second, this illustrates the peculiar phenomenon of 2004, where the pass ratio (passes over 18-year-olds) has gone up in spite of the pass rate going down: the numbers writing, as a proportion of the population, went up more slowly than the numbers of passes as a proportion of the writers went down. All this suggests that as an indicator of system "culmination" the pass ratio is a better tracking indicator than the pass rate.

Grade-specific survey-based data can be used to compare the mid-1990s to 2003, as is done in Figure 5. The diagram shows that over-enrollment in Grade 1 has come down, perhaps as a result of the age-grade norms, but is still quite high. The final throughput of the system, as measured by the final (Grade 12) ratio is essentially the same. The peak in Grade 11 in 1995 is now a peak in Grade 10 in 2003, probably as a result of the pressure on the Senior Certificate pass rate. Children appear to be held back just as much as in 1995, but now one grade earlier.

Figure 5. Grade-specific Enrollment Ratios, 1995 and 2003



Source: 2005 General Household Survey and 1995 October Household Survey

It should be noted that an age-specific pattern shows lower enrolment ratios in 2003 than in 1995 in the older age groups, starting at about 19. This may be due to a "clean up" of perennial repeaters in Grade 12, which took place in the late 1990s, but would require further research to establish.

4. Why do children drop out?

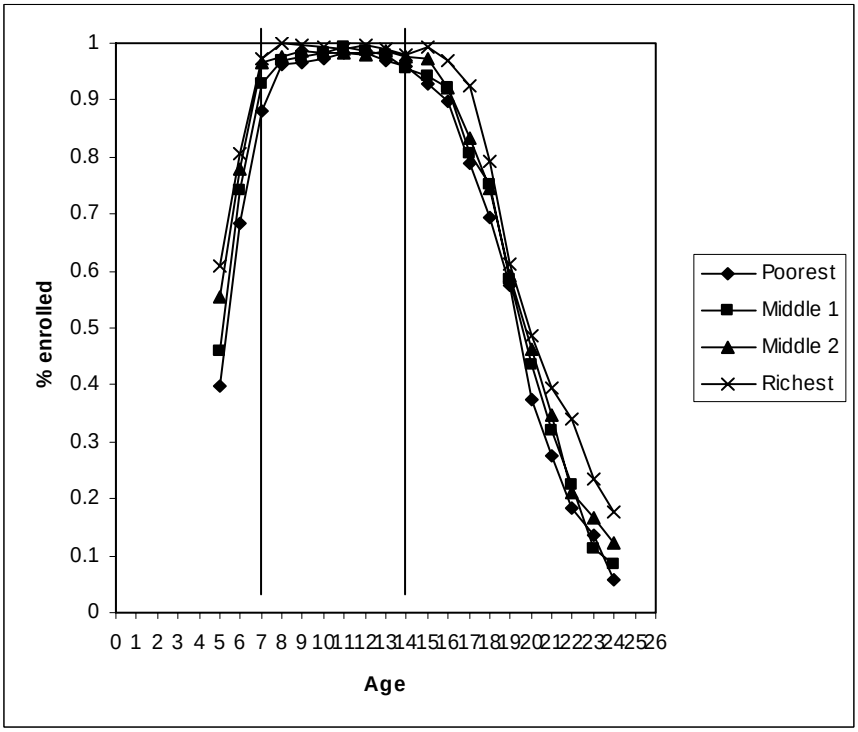
As noted, there is relatively limited dropping out. Nonetheless, there is *some* dropping out, particularly in the last few grades, and some non-attendance before Grade 1. Why does it take place? The most systematic and large-scale source on reasons for dropping out is the 2003 General Household Survey, which directly asks those not attending school why they no longer attend. Whilst this approach does not allow for a profound understanding of the dropout problem, since it assumes a simple reason and a simple "decision" model (rather than a complex gradual outcome), the approach does have the virtue of using large scale data separable by gender, population group, and income (or total family expenditure—a good proxy for income).

The following two graphs make it clear that gender makes some difference to attendance and dropout behavior, but a relatively minor difference. (The compulsory age range is indicated, as usual, with two vertical lines.) Income also makes some difference, but only

in the earliest and older ages.[3] The nearness of the curves in the extremes is visually deceptive. There is a difference of 20 percentage points, for example, between the enrolment of the richest and the poorest at age 5, for example. But, in any case, if one compares these two figures to Figure 1, particularly for the older ages, it is apparent that population group makes a bigger difference than gender or income. We must note that with regard to the latter we were rather surprised. The expectation was that in today's South Africa income would be a more important driver of dropout behaviour than race. But it is interesting to note that the two race groups that seem to drop out earlier, namely the Coloured and Indian groups, are not the poorest. Thus, some suspicion arises that what seems to be race-based is actually based on administrative practices of traditional HoR and perhaps (but less likely) HoD schools. There may also be sociological factors unrelated to income that account for these patterns. All this merits further investigation.

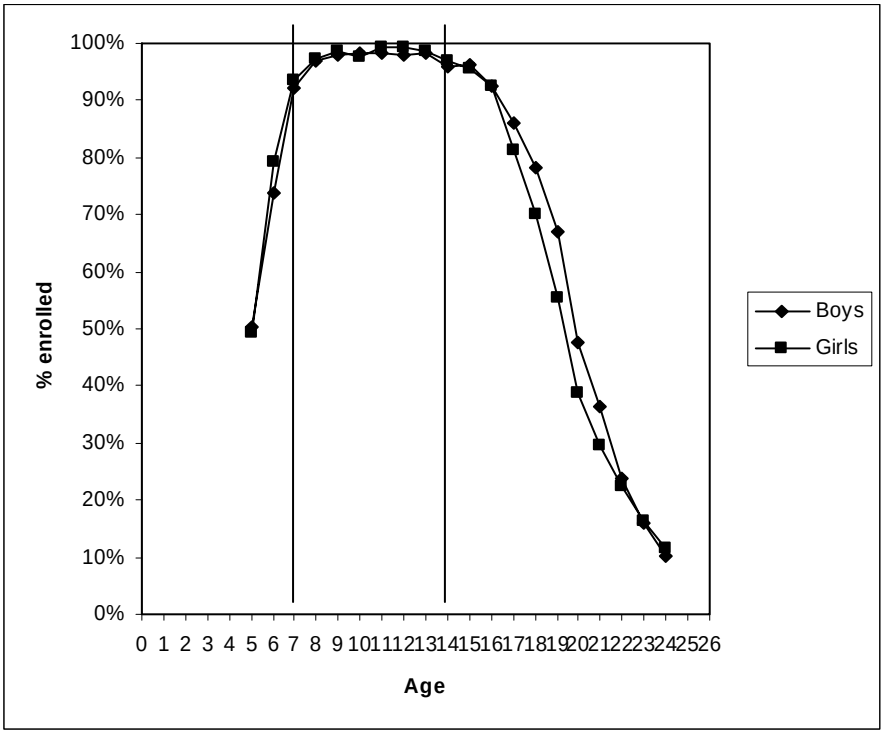
[3] An explanation on income data is in order. First, the General Household Survey asks about total expenditure, not income, so it is actually expenditure that is used. But expenditure is a good proxy for income and in many countries is estimated to be more accurately reported, in any case. Second, because of the way data are reported, it is impossible to create expenditure quintiles, unfortunately, or even quartiles. The data are reported in expenditure ranges, rather than as an actual amount. The data then do not naturally fall into quintiles or quartiles. The percentile groupings of the data that are reported in the four groups that appear in the figures are 24, 33, 24 and 18 respectively. That is the best we could do. This is a relatively minor point, though.

Figure 6. Enrollment ratios by income, 2003



Source: calculated by the author from 2003 General Household Survey

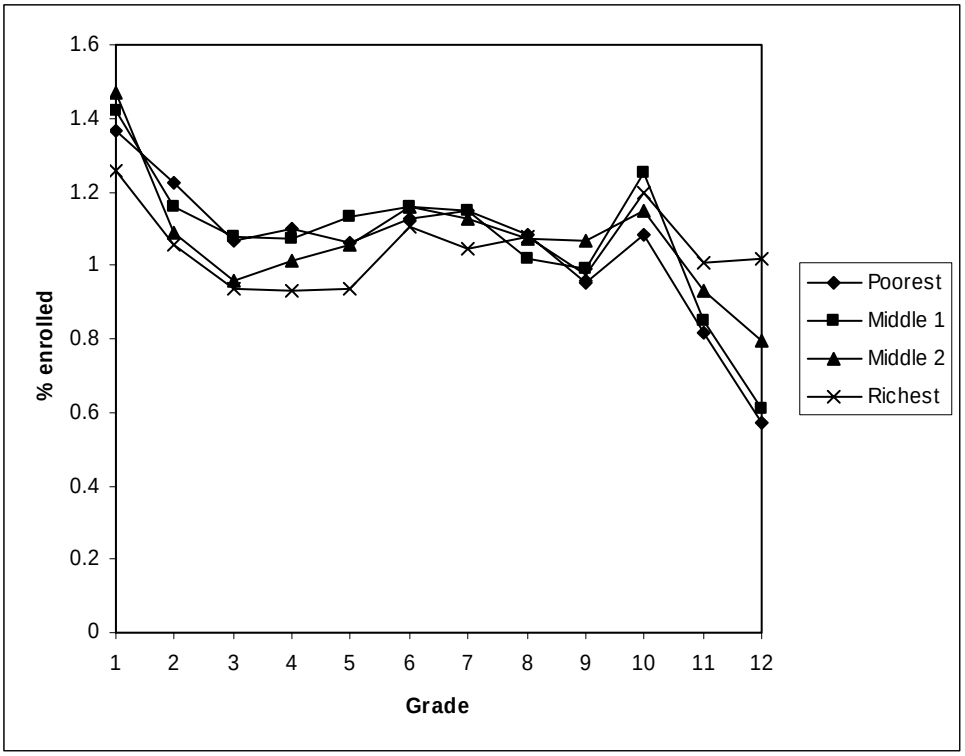
Figure 7. Enrollment ratios by gender, 2003



Source: calculated by the author from 2003 General Household Survey 2003

However, if one studies enrolment ratios by grade rather than age, then some interesting income-based patterns do emerge. The most interesting patterns are three. First, there is the systematic over-enrolment of the poorest groups in the early grades. This represents the formal and informal repetition phenomena already discussed above, related to the lack of ECD as well as, most likely, more lax enforcement of age norms in the schools frequented by the poor. Second, the situation is notably reversed in the latter grades, and here there is definitely a pattern for the richer to persist through the latter grades. Thirdly, the tendency for a bottleneck in Grade 10 to develop, due most likely to repetition in attempting to control flow to Grade 12 (and thus improve performance on the only significant accountability test the system possesses), is common to all income groups. Even schools serving the richest segments of the population appear to be holding children back in Grade 10.

Figure 8. Grade-specific Enrollment Ratios, by Income Group



Source: calculated by the author from 2003 General Household Survey

Finally, those who have dropped out have signaled why they dropped out. The reasons are shown in Table 3. Since the group of most interest is the 16 to 18 year-olds, the table selects that group for analysis. In reading the data, it is important to keep in perspective that these figures do not affect the whole age group, but only those not attending. Thus, the last row of the table shows, as a reminder item, the percentage of the age group that is not attending. To show the percentage of the total population that is affected by the dropout phenomena, the percentage not attending is multiplied by each specific cause. This is shown in the last two columns. Finally, cases of important contrast between males and females are shown. The following facts stand out. Some 6% to 8% of the age group is affected by fees. No other factor really stands out as importantly affecting the age group as a whole, and even this one factor (fees) is relatively small in its effect on the *whole* age group. When viewed as affecting the dropouts themselves, fees stand out. As a factor affecting the dropouts, education being uninteresting affects a rather large proportion of the boys (17%), and pregnancy, marriage, or family concerns affect a very large percentage of the girls (13% + 3% + 10% = 26%). The difference between male and female dropouts finding education useless or uninteresting is notable.

Table 3. Self-reported Reasons for Dropping Out, 16-18 Age Group, 2003

				Percentage of total age group affected	
Reason	Male dropouts	Female dropouts	Male-Female contrast 10% or larger	Male	Female
Too old/young	1%	0%		0%	0%
Has completed education	7%	5%		1%	1%
School too far away	4%	3%		1%	0%
No money for fees	42%	41%		6%	8%
Has to work (at home or job)	9%	4%		1%	1%
Education is useless or uninteresting	17%	7%	10%	2%	1%
Illness	8%	5%		1%	1%
Pregnancy	0%	13%	-12%	0%	2%
Failed exams	7%	4%		1%	1%
Got married	1%	3%		0%	1%
Family commitments	0%	10%	-10%	0%	2%
Other	3%	5%		0%	1%
Total	100%	100%		14%	19%
Percentage not attending	14%	19%			

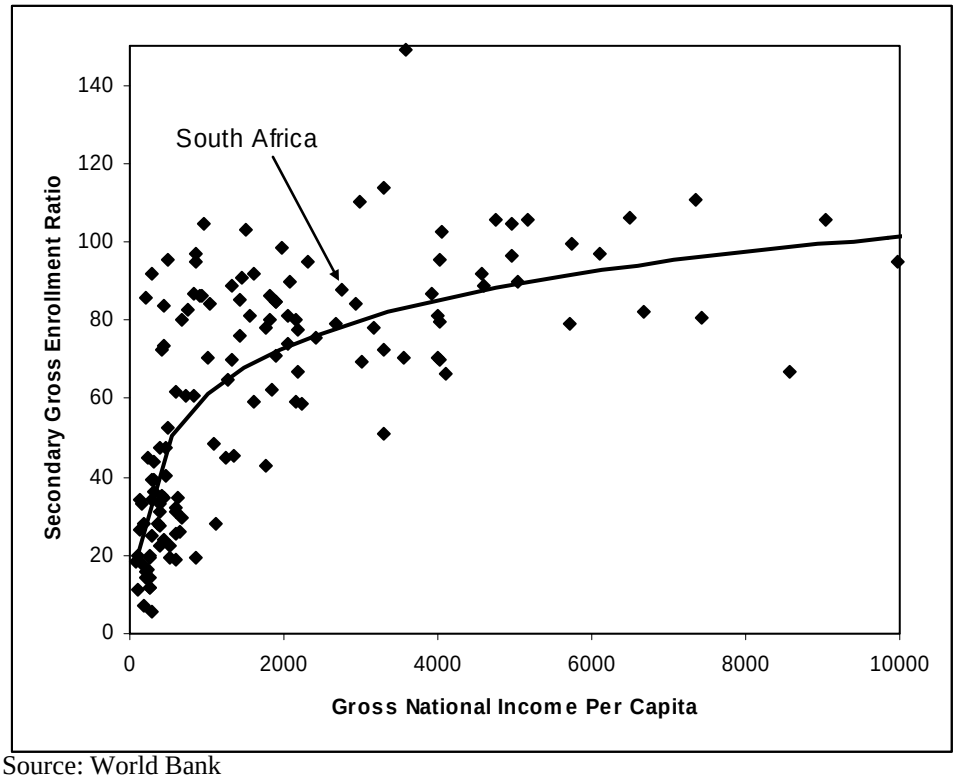
Source: calculated by the author from 2003 General Household Survey

5. Where is South Africa in comparison to other countries?

Whilst it might be ideal not to have any dropouts even through secondary, this is an ideal that hardly any countries realize. (And this may not be that much of an ideal in any

case—to our knowledge, no one has demonstrated that having 100% secondary school completion, as opposed to, say, 80%, is key to any worthWhilst social goal, whereas the case is well-made for primary education.) Countries such as Japan come close, but other than some of the Asian countries, even fully developed countries typically have some dropouts. Comparing South Africa's case against some abstraction of zero dropouts is therefore unrealistic. Comparing against countries at a similar level of development might be more interesting. Unfortunately there is very little international evidence on the secondary completion rate, which would be the most useful synthetic indicator. However, the secondary gross enrollment ratio is a fairly good indicator, because it captures the completion of primary, the progression to secondary, and some of the dropout behavior *during* secondary, though this is conflated with repetition. Figure 9 shows where South Africa lies in comparison with other countries, and shows her to be slightly better than what one might expect given the trend line shown. Table 4 shows some detailed comparisons, making essentially the same point.

Figure 9. Secondary School Gross Enrolment Ratios, by Income



Source: World Bank

Table 4. Secondary gross enrollment ratio, international comparisons

Income comparators	Gross National Income Per Capita	GER Secondary
Dominican Republic	2223	59
Russian Federation	2304	95
Turkey	2678	79
South Africa	*2750*	*88*
Jamaica	2946	84
Brazil	2998	110
Botswana	3308	73
Average of comparators	*2743*	*83*

Neighbourhood comparators	Gross National Income Per Capita	GER Secondary
South Africa	*2750*	*88*
Botswana	3308	73
Zimbawe	460	40
Namibia	1832	62

Source: World Bank

Thus, South Africa does about as well as one might expect, or a little better, in terms of income comparisons, and much better in terms of neighbourhood comparisons.

6. So, then, what *are* the problems?

It seems fairly clear that Whilst there are, of course, some dropouts in South Africa as elsewhere, the problem is not nearly as big ("huge") as has been reported in the media, nor is it larger than what one would expect for a country at South Africa's level of per capita income.

There is little doubt that the biggest two problems South Africa faces are the extreme inequality in actual learning achievement, and the relatively low level in this achievement across all groups. There is also a problem in South Africa in, precisely, the gap between very large access or enrolment proportions, and the low level of achievement. The numbers of those who reach the last year of secondary school in South Africa are equivalent to those from any other middle-country, but they are far worse equipped for the labour force. This must be leading to, or has already created, a crisis of expectations. This final section documents all these problems.

First, the achievement inequality in South Africa is a) very large, and b) to a remarkable extent unexplained and unexplainable via reference to economic inequality. This is particularly true in mathematics achievement, so we will focus on this issue. The roots of this problem are sociological, curricular, linguistic, more than income-related, and also have to do with the lack of use of standards and high expectations for all.

The basic problems can be seen in Table 5.[4] This shows South Africa and Perú to be far and away the most unequal countries. Two measures of inequality are considered: the coefficient of variation (standard deviation over the mean) and the ratio of performance at the 95th percentile to performance at the 5th percentile. On both scores South Africa and Perú (the two most unequal countries of those with rigorous measurement) are some 30% more unequal that the next most unequal countries (Ghana, Argentina, Israel, and Brazil).

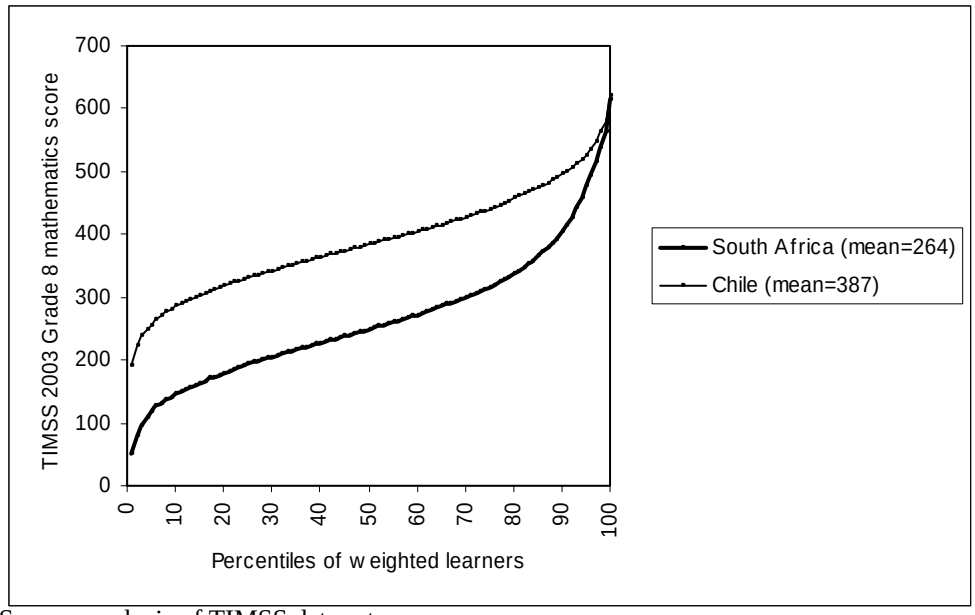
Most importantly, the next set of countries shows that there are countries with considerable social inequality and ethnic divisions that nonetheless manage to have much less (about half as much) inequality and much better averages in educational results than South Africa. The two exemplars are Malaysia and Mexico. It is important, thought, to note how they achieve less inequality and better averages: it is by bringing up the bottom. The last two columns in the table show the difference of performance between these countries and South Africa both at the bottom (5th percentile) and at the top (95th percentile) of the performance distribution. The difference (especially the percentage difference) at the bottom is much greater than the difference at the top.

Whilst improving scores at the poorly performing end of the spectrum must clearly be a priority, Figure 10 illustrates that under-performance relative to a middle income country such as Chile (chosen because this is a middle-income country often cited noted for its policies) occurs essentially across the entire spectrum. It is only above more or less the 80th percentile that the gap in absolute terms between the two countries appears to narrow. What needs to happen if the South African mean is to be raised to that of Chile, is that performance of essentially *all* learners needs to improve, though the gap that needs narrowing at the best-performing end is of a smaller magnitude.[5]

[4] It may be thought, particularly by testing purists, somewhat illegitimate to put both PISA and TIMSS results into the same table, as they test different populations and follow different methodologies. However, the following should be noted. First, the mean PISA and TIMSS scores across countries are quite similar—different by only about 10 points out of 500, or so (2%). Second, the average within-country standard deviation of PISA scores is about the same as that for TIMSS—around 60. Finally, because there are many countries that participate in both TIMSS and PISA, one can draw a correlation between them, and the correlation is 0.73. So they are not measuring quite the same thing, but close enough, in our judgment, to allow us the liberty of making some comparisons across tests, in comparisons that are as gross as noting that South Africa has nearly fifty percent more inequality in TIMSS as Mexico has in PISA.

[5] Thanks to Martin Gustafsson who contributed some of the analysis and points in this paragraph in particular and in this section more generally, including Table 6 and Figure 10.

Figure 10. Performance by percentile in Chile (top line)) and South Africa (bottom line)



Source: analysis of TIMSS data sets

The careful reader will have noticed that much of the difference in the inequality coefficients is due to the fact that the denominator (either the mean, in the case of the coefficient of variation, or the performance at the 5th percentile in the case of the 95/5 ratio) in the case of South Africa is so low, namely, average achievement is so low, as shown in Figure 10 and Table 5. This is to some degree true: absolute performance in South Africa is generally low, except at the top end of the scale. This is worth considering in some detail, but it is also worth noting that this is less important to a discussion of inequality (though important to a discussion of quality) than might seem at first blush. First, it is true that even South Africa's elite performs only at about the average of the top-performing countries, or a little worse. (The performance of South Africa at the 95th percentile is at about the mean performance in a country such as the Netherlands—actually a little below the Netherlands average as shown in Table 5.) It can also be noted that the "pure" variability, namely the standard deviation, really does not vary too much between countries, and what reduces the inequality in a country such as Mexico is that the average is higher. But the average is higher, we have seen, partly because the bottom performers perform so much better in Mexico or Chile than in South Africa. It is also important to remind oneself that this is precisely what measuring inequality is all about. In a country where half of the population has an income of 100, and half has an income of 200, the pure variability is the same as in a country where half the population has an income of 1100 and the other half an income of 1200. But a moment's intuitive reflection will show that the situation of inequality is clearly much worse in the first than in the second case, because in the first case the rich are twice as rich as the poor, but in the second case the rich are only 9% richer than the poor. Thus, Whilst it is true that the standard deviation does not vary all that much between, say,

South Africa and Mexico, the reality is that it is the variation relative to the mean that creates a sense of inequality, and this is the way to measure inequality.

Table 5. Comparative Performance on International Assessments

Country	Mean	Std. Dev.	Perform. at the 5th Percentile	Perform. at the 95th Percentile	CV	95/5	Test	Difference of lowest with respect to SA's lowest	Difference of highest with respect to SA's highest
Countries with most inequality									
South Africa	264	107	117	484	0.41	4.1	TIMSS 2003		
Peru	292	108	116	470	0.37	4.1	PISA 2000		
Ghana	276	91	130	430	0.33	3.3	TIMSS 2003		
Argentina	388	120	180	574	0.31	3.2	PISA 2000		
Israel	433	131	206	637	0.30	3.1	PISA 2000		
Brazil	334	97	179	499	0.29	2.8	PISA 2000		
Developing countries with much less inequality than South Africa									
Mexico	400	95	238	552	0.24	2.3	PISA 2003	104%	14%
Chile	387	83	258	531	0.21	2.1	TIMSS 2003	121%	10%
Malaysia	508	74	388	630	0.15	1.6	TIMSS 2003	232%	30%
Thailand	420	78	293	550	0.19	1.9	PISA 2003	151%	14%
Some top performers									
New Zealand	522	105	338	682	0.20	2.0	PISA 2003	189%	41%
Finland	543	81	400	666	0.15	1.7	PISA 2003	242%	38%
Netherlands	536	69	417	644	0.13	1.5	TIMSS 2003	256%	33%
Australia	505	82	368	634	0.16	1.7	TIMSS 2003	215%	31%
Flemish Belgium	537	73	398	643	0.14	1.6	TIMSS 2003	240%	33%
Countries whose *mean* performance is equivalent to South Africa's *elite* performance									
Hungary	488	98	327	648	0.20	2.0	PISA 2000	179%	34%
Armenia	478	84	330	605	0.18	1.8	TIMSS 2003	182%	25%
Serbia	477	89	326	618	0.19	1.9	TIMSS 2003	179%	28%
Spain	476	91	323	621	0.19	1.9	PISA 2000	176%	28%

Source: calculated by the author from data in PISA and TIMSS reports.

An analysis of the relationships between socio-economic advantage and performance, as opposed to an analysis of performance alone, provides a somewhat stronger indication that the educational elite in South Africa does not perform as well as its peers in other countries. The following table, based on TIMSS 2003 data, reveals that learners in South Africa who have at least one parent with a university education perform substantially worse than their peers in the other selected countries except for Ghana. These learners also perform worse that the national average in the other countries, except for Ghana. The last column indicates the correlation between the level of education of parents (where the lowest value represents the highest level, so the correlations are negative) and the mathematics performance of learners. The statistic for South Africa is not outside the range of the other countries, confirming that South Africa is no exception to the general

rule: learners with more educated parents are at a distinct advantage at school—they can be expected to perform better than most.

The 2003 General Household Survey throws some light on the racial breakdown of the educational elite being referred to in **Error! Reference source not found.**. Given the apartheid legacies, this breakdown can be informative. Firstly, the percentage of Grades 7 to 10 learners with at least one person with a university education present in the same household is 5 per cent, in other words substantially lower than the 8 per cent suggested by TIMSS. There are many possible explanations for this discrepancy. According to the GHS, the proportion of Africans, whites, coloureds and Indians amongst these learners is 59, 27, 8 and 4 per cent respectively. It is possible that the performance problem in the educational elite is applicable to all race groups.

Table 6. Characteristics of learners in various countries, TIMSS 2003

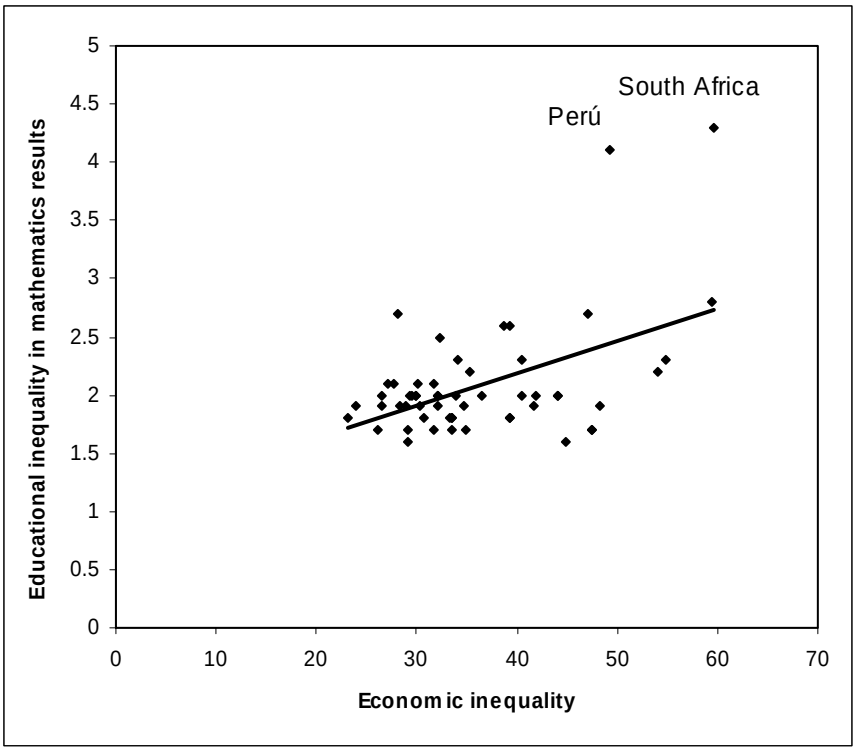
Country	Country mean score for mathematics	Standard error of this (95% confidence interval)	% of learners with at least one parent possessing a university education (according to TIMSS)	Mean mathematics score of these learners	Standard error of this (95% confidence interval)	Correlation parent education to learner performance
South Africa	264	5.5	8%	353	5.5	-0.34
Botswana	366	2.6	9%	414	3.8	-0.21
Ghana	276	4.7	8%	323	4.3	-0.24
Egypt	406	3.5	21%	463	1.8	-0.33
Chile	387	3.3	15%	467	2.1	-0.51
Malaysia	508	4.1	9%	542	3.2	-0.23
New Zealand	494	5.3	15%	534	3.3	-0.24
Netherlands	536	3.8	16%	568	2.9	-0.29

Source: analysis of TIMSS data sets by Martin Gustafsson

There are several other interesting facts about inequality of educational results in South Africa.

The first is that educational inequality in South Africa is much greater than would be predicted by its income inequality, and its income inequality is high. Using the same ratio of achievement at the 95th percentile to achievement at the 5th percentile as an index of educational inequality, and the Gini coefficient of income inequality as an index of economic inequality, we can see in Figure 11 that in general income inequality and educational inequality are fairly well correlated, but we can also see how South Africa (and Perú) are the two countries that truly diverge from the main trend and beat the correlation.

Figure 11. Educational and economic inequality

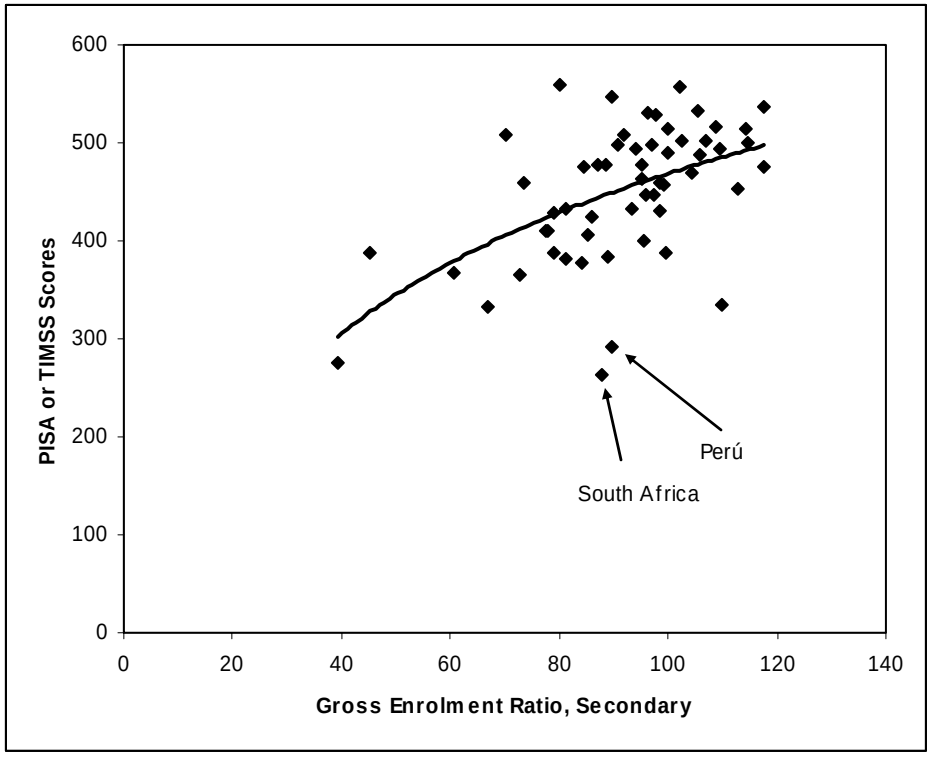


Sources: calculated by the author from PISA and TIMSS reports
and the World Income Inequality Database at
http://www.wider.unu.edu/wiid/wiid.htm, downloaded on 14 Dec 2005

Thus, educational inequality in South Africa cannot be blamed entirely on economic inequality: educational inequality is much greater than can be predicted via economic inequality. (But note: educational inequality is emphatically *not* greater than economic inequality, and this is an important issue. To be utterly clear: educational inequality in South Africa is greater than that which can be predicted by economic inequality, but it is not greater than economic inequality.) Thus, in South Africa (as in Perú) the determinants of educational inequality are other: sociological inequality, and, most likely, the fact that poor standards and low expectations for the poor characterize the system, and this is also embodied in the lack of proven and effective pedagogical models that can reach the poor, particularly with good methods (and social marketing) for starting children in home language but transitioning children from mother tongue to the *lingua franca* (if that is what is desired, but in any case starting with mother tongue). Exactly the same problem exists in South Africa as in Perú, the former being characterized by formal *apartheid* and the latter by an informal, but just as noxious, *apartheid*, and both characterised by a romantic patronising of the poor by those who decry rigourous standards as unequalising or retrograde. The systems are characterized also by laxity of effort and low time-on-task.

Similarly worrisome is the gap between enrolment and achievement in South Africa. Most countries that achieve the same high levels of secondary enrollment as South Africa also achieve fairly good results. In South Africa the results are way out of line with the enrolment levels. This is an imbalance that cannot help but lead to increased social tension, as those who exit from secondary school expect the standard of life that secondary school graduates in competing countries can have. It is interesting that, again, Perú, one of the bastions of informal *apartheid*, accompanies South Africa on this front as well. In this respect it is important to note that one is talking about higher-order cognitive skills here. The mismatch between expectations and labour force reality that we are addressing here is not likely to be remedied through technical skilling in vocational tasks, but by thorough-going reform that pushes up the analytical and more serious cognitive skills of South Africa's poor. Neither vocational remediation nor patronising the poor (in the name of some supposed progressivism) via the promotion of soft skills will do the trick.

Figure 12. Enrolment in secondary school and learning achievement



Sources: PISA and TIMSS databases and World Bank

Before ending this discussion of what the real problems are, it is important to touch upon the issue of whether education is really just reproducing the traditional inequalities of South Africa. Every indication we have is that this is very far from the truth. On the other hand, we also know that educational inequality in South Africa is objectively very high—we have just got done demonstrating this. Thus, Whilst education is almost certainly contributing to improving social equality, it could be contributing much more,

by improving the cognitive performance of the poor, though the use of better, more rigorous pedagogical models, more and better standards of performance, more accountability, better management of mother tongue initial education and transition to the *lingua franca*, and more time on task.

The results in terms of inequality (using now a measure of inequality based on the Gini coefficient of concentration) show the following stylized facts:

Table 7. Inequality of various factors

Variable	"Stylized" Gini or concentration index
Parental income	60
Educational resources approximately fifteen years ago	25
Educational resources today	0 to 10, depending on the input in question
Grade attainment approximately twenty years ago	30
Grade attainment today	15
Distribution of learning results today	15 to 30

Sources are as follows. Parental income coefficient is from Statistics South Africa, 2002, "Earning and spending in South Africa: Selected findings and comparisons from the income and expenditure surveys of October 1995 and October 2000." Available from: <http://www.statssa.gov.za> [Accessed August 2005]. Other sources give indices between 55 and 60. Indices on educational resources and distribution of learning results are from Crouch, L., M. Gustafsson and P. Lavado, "Measuring Educational Inequality in South Africa and Perú," in D. B. Holsinger & W. J. Jacob (Eds.) **International Handbook on the Inequality of Education**, Dordrecht, The Netherlands: Springer. Data on grade attainment and distribution of learning results are from Crouch, L., "South Africa Equity and Quality Reforms," **Journal of Education for International Development**, 1(1). 2005.

In interpreting this table, it is well to remember that the Gini index of concentration has a theoretical maximum of 100, and can take on a negative value if resources are spent in a pro-poor fashion. The following can be discerned from these results. First, the inequality of results, either using grade attainment or learning achievement, is much lower than the inequality of parental income distribution. This bears repeating, since there seems to be so much misunderstanding around this issue: the educational results of the children are distributed much more equally than the income of the parents. This bodes well for the future, presumably, and means that education is probably contributing to equalization. Second, educational results are distributed more unequally than educational inputs. This does not mean that one should not engage in pro-poor spending, but it does strongly suggest that further equalization of spending is not the main constraint to further equalization in learning achievement. Third, educational equality has been improving markedly over the past twenty years, at least in terms of grade attainment. It is difficult to say whether equality in actual learning results has increased. But learning results today are better-distributed than mere grade attainment twenty years ago, so one could hope actual learning will be more equally distributed in future, though this will not happen automatically and will require instead policies that insist on good standards of learning for all, especially in the early grades. Finally, there has been a radical improvement in the equality of distribution of educational resources in the last fifteen years or so (but really since 1995—the problem is that the only semi-rigorous measure of inequality prior to 1995 or so is from 1991). In short, Whilst the system can and ought to

do much more, much has been and is being done to improve equality along many dimensions.

One cannot emphasize too much: more can indeed be done to increase equality of results. This will require a get-tough attitude on issues such as the specification of learning objectives (already much improved, but can be improved vastly still), the training of teachers to deliver on specific objectives according to simple and planned yearly instructional sequences, and the development of much better accountability and incentives systems for putting the training to work. It will also require starting on the basics: literacy and numeracy in the earliest grades. At present an unacceptably large proportion of children simply cannot read, for example, even beyond Grade 2. All children should become fluent readers in at least one language by the end of Grade 2, regardless of social class or linguistic origin. There are plenty of examples from within South Africa itself of pedagogy that is effective even in very poor situations. Sadly, there are, as yet, few effective managerial mechanisms for replicating this knowledge and taking it to scale country-wide. Perhaps ensuring that we have less fruitless debates, such as arguing over how many dropouts there are, will let us concentrate attention on the problems that are really plaguing the system.